Bayesian Weighted Inference from Surveys

David Gunawan
University of New South Wales

Anastasios Panagiotelis
Monash University

William Griffiths
University of Melbourne

Duangkamon Chotikapanich
Monash University

23 March 2017

Abstract

Data from large surveys are often supplemented with sampling weights that are designed to reflect unequal probabilities of response and selection inherent in complex survey sampling methods. We propose two methods for Bayesian estimation of parametric models in a setting where the survey data and the weights are available, but where information on how the weights were constructed is unavailable. The first approach is to simply replace the likelihood with the pseudo likelihood in the formulation of Bayes theorem. This is proven to lead to a consistent estimator but also leads to credible intervals that suffer from systematic undercoverage. Our second approach involves using the weights to generate a representative sample which is embedded within a Markov chain Monte Carlo (MCMC) or other simulation algorithm designed to estimate the parameters of the model. In extensive simulation studies, the latter methodology is shown to achieve performance comparable to the standard frequentist solution of pseudo maximum likelihood, with the added advantage of being applicable to models that require inference via MCMC. The methodology is demonstrated further by fitting a mixture of gamma densities to a sample of Australian household income.

Keywords: Sampling weights; Latent representative sample; Markov chain Monte Carlo; Gamma mixture; Pseudo maximum likelihood.

1. Introduction

Raw data from surveys seldom come from a simple random sample where selection of each individual is equiprobable but from complex survey sampling methods such as stratification and multistage sampling that exhibit unequal probabilities of selection and non-response. Examples of large surveys with these characteristics are the Panel Study of Income Dynamics (PSID), the British Household Panel Survey (BHPS), and the Household Income and Labour Dynamics in Australia (HILDA) survey, all of which are increasingly used in applied statistical research. For samples that are non-representative in the sense that individuals with different characteristics have different probabilities of selection, the standard methods of inference and estimation may be biased or inconsistent; this issue is discussed in detail by Korinek et al. (2007), Pfeffermann (1996), Breunig (2001), and Wooldridge (1999, 2001, 2007). A common way to address the problem is to use sampling weights provided with the survey datasets. Sampling weights act as "expansion factors" that scale and correct the representativeness of the sample to the population. They accommodate complex sampling designs and may be modified to ensure demographics such as sex, race, and age from the weighted sample match known census figures. If a survey respondent comes from a demographic group that has a low probability of selection or response, they are allocated a higher weight. Because sampling weights must take into account a large number of factors, their computation is often complicated (see Gelman (2007), Korinek et al. (2007), and references therein), and detailed information on how they were constructed may not be available to researchers. We are concerned with a situation typical in much applied work where the only available information is the dataset and the sampling weights for each unit in the sample, with little or no information regarding the complex sampling design or how the weights were computed. In line with this information set, we treat the sampling weights as given and do not focus on their estimation and construction.

A number of methodologies that exploit survey weights to obtain unbiased and consistent estimation and inference have been proposed. One of the earliest approaches for estimating the population mean of a random variable is the classical weighted ratio estimator (see Horvitz and Thompson 1952). 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 distributions. They provide Bayesian interval estimates for the population mean that are asymptotically valid in a frequentist framework. Although such nonparametric estimators can be used for estimating the population mean, parametric statistical models require a more general estimation framework. The most popular framework for taking sampling weights into account when estimating parametric models is pseudo maximum likelihood; see, for example, Godambe and Thompson (1986), Molina and Skinner (1992), Hesketh and Skrondal (2006), Skinner and Mason (2012), and references therein. To obtain the pseudo maximum likelihood estimator (PMLE), the usual log-likelihood is replaced with an objective function that is the sum of each sample weight multiplied by the contribution of its corresponding observation to the log likelihood. The resulting estimator is a special case of a general inverse probability weighted M-estimator (Wooldridge 1999, 2001, 2007).

Although the PMLE provides a way of accounting for sampling weights when estimating parametric models within a frequentist framework, to the best of our knowledge no general alternative method exists for estimation within a Bayesian framework, despite the increasing popularity of Bayesian inference in statistics. Such an alternative is useful if the aim is the fundamental one of wanting to base inferences on the posterior distribution, or if it is simply to exploit numerical methods such as Markov chain Monte Carlo (MCMC) for estimating complex statistical models such as mixtures, or multinomial and multivariate probit models.

We consider two approaches for incorporating the information from sampling weights into Bayesian inference. The first, which we call the Bayesian Pseudo Posterior Estimator (BPPE) simply replaces the likelihood with the pseudo-likelihood in the usual formulation of Bayes' theorem. The second which we call the Bayesian Weighted Estimator (BWE) is a data-augmentation approach where a "pseudo representative sample" is treated as missing data and generated by resampling with replacement from the observed data, using the normalized sampling weights. This step is combined with a standard MCMC sampler, with inference about the unknown parameters based on the pseudo representative sample. Since the early work of Tanner and Wong (1987), data augmentation has been used extensively for Bayesian estimation of a variety of statistical models. See, for example, Chib (1992), Albert and Chib (1993), Geweke and Keane (2007) and Geweke and Amisano (2011).

Replacing the likelihood with some other function of the parameters and data is an idea that goes at least as far back as the notion of 'proper likelihoods' introduced by Monahan and Boos (1992) and has received significant treatment in the case of Bayesian empirical likelihood (see Lazar (2003), Schennach (2005) and Rao and Wu (2010)). We evaluate the asymptotic behavior of our two proposed approaches under an assumption of non-informative priors. For the BPPE we are able to derive theoretical results that suggest consistency, but an asymptotic variance that leads to undercoverage of credible intervals in repeated sampling. These theoretical results are validated in a simulation study. In the case of the BWE, the likelihood is replaced with a Monte Carlo estimate of a density that is a discrete mixture over all possible pseudo-samples. Although this mixture is difficult to work with theoretically, we provide a sound intuitive justification for its use, and show through extensive simulations that the BWE achieves accurate empirical coverage.

We begin Section 2 with a brief description of the PMLE and its sandwich covariance matrix estimator, followed by a discussion of the problems that arise if this approach is

adopted within a Bayesian framework. The details of our proposal for an alternative Bayesian weighted estimator that utilizes generation of a representative sample are presented in Section 2.3. In Section 3 we use two simulation studies to illustrate application of the proposed estimator and to compare its repeated sampling properties to those of alternative estimators. Two quite different models are chosen for these illustrations – estimation of the mean and variance of a Gaussian distribution, and estimation of the parameters of a two-component mixture of gamma densities. In Section 4 Bayesian weighted and unweighted estimates of an Australian income distribution, modelled as a three component mixture of gamma densities, are presented. A conclusion is provided in Section 5.

2. Methodologies

Assume we have a random variable Y whose population can be described by the density function $p(y|\theta)$, θ being an unknown vector of parameters we wish to estimate. We are supplied with a non-representative sample $\boldsymbol{y}' = (y_1, y_2, \ldots, y_n)$ that is based on a complex survey design, typically involving several demographic factors. Corresponding to each sample observation, we are also supplied with sampling weights $\boldsymbol{w}' = (w_1, w_2, \ldots, w_n)$, $0 < w_i < \infty$, but the details of the survey design and how the weights are calculated are not available to the investigator. It is assumed that the weights have been constructed such that a weight w_i is inversely proportional to the probability that the survey design selected an observation with the demographic characteristics of observation y_i. For estimation, observations whose probability of being selected is less than it would be under simple random sampling are weighted more heavily than they would be under simple random sampling, and vice versa. We assume that the w_i have been scaled such that $\sum_{i=1}^{n} w_i = n$. A normalized weight $\tilde{w}_i = w_i / n$ corresponds to the proportion of the population with the

demographic characteristics of observation y_i. In what follows we first briefly describe the pseudo maximum likelihood estimator for θ (Section 2.1), followed by a Bayesian estimator that uses the pseudo likelihood function (Section 2.2). Our proposal for a Bayesian weighted estimator designed to overcome problems with using the pseudo likelihood within a Bayesian framework is described in Section 2.3.

2.1 Pseudo Maximum Likelihood Estimator

A pseudo log likelihood is defined as $L_P(\theta ; \boldsymbol{y}) = \sum_{i=1}^{n} w_i \log p(y_i \mid \theta)$. The PMLE $\hat{\theta}_{PML}$ satisfies the first-order conditions

$$\frac{\partial L_P(\theta ; \boldsymbol{y})}{\partial \theta} = \sum_{i=1}^{n} w_i \frac{\partial \log p(y_i \mid \theta)}{\partial \theta} = \mathbf{0}$$

This estimator is consistent but not efficient (Wooldridge 1999, 2001, 2007). Under some regularity conditions $\sqrt{n}\left(\hat{\theta}_{PML} - \theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}_w^{-1}\boldsymbol{V}_w\boldsymbol{H}_w^{-1}\right)$ where θ_0 is the true value for θ and $\boldsymbol{H}_w$ and $\boldsymbol{V}_w$ are consistently estimated using

$$\hat{\boldsymbol{H}}_w = \frac{1}{n}\sum_{i=1}^{n} w_i \frac{\partial^2 \log p\left(y_i \mid \hat{\theta}_{PML}\right)}{\partial \theta \partial \theta'} \tag{1}$$

and

$$\hat{\boldsymbol{V}}_w = \frac{1}{n}\sum_{i=1}^{n} w_i^2 \frac{\partial \log p\left(y_i \mid \hat{\theta}_{PML}\right)}{\partial \theta} \frac{\partial \log p\left(y_i \mid \hat{\theta}_{PML}\right)}{\partial \theta'} \tag{2}$$

respectively. For making inferences about θ the standard errors are obtained from the observed sandwich covariance matrix estimator $n^{-1}\hat{\boldsymbol{H}}_w^{-1}\hat{\boldsymbol{V}}_w\hat{\boldsymbol{H}}_w^{-1}$ (White 1980, 1982).

2.2 Bayesian Pseudo Posterior Estimator

Given the successful development of the pseudo likelihood sampling theory approach to estimating θ, a natural question to ask is whether a Bayesian approach with the usual

likelihood function replaced by the pseudo likelihood would be suitable. For a given prior distribution $p(\boldsymbol{\theta})$, the posterior density obtained using this approach is given by

$$\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{w}) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} p(y_i \mid \boldsymbol{\theta})^{w_i} \tag{3}$$

Theorem: The pseudo posterior $\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{w})$ converges to a normal distribution with mean $\hat{\boldsymbol{\theta}}$ and covariance matrix $-n^{-1}\hat{\boldsymbol{H}}_w^{-1}$ where $\hat{\boldsymbol{\theta}}$ is the posterior mode and $\hat{\boldsymbol{H}}_w = n^{-1}\sum_{i=1}^{n} w_i \, \partial^2 \log p\left(y_i \mid \hat{\boldsymbol{\theta}}\right) \big/ \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is the weighted Hessian.

Corollary: The posterior mode $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is a unique solution to the population maximization problem $\boldsymbol{\theta}_0 = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} E_y \left[\log\left(p(y \mid \boldsymbol{\theta}) \right) \right]$.

Proof: See appendix A

Since the pseudo posterior distribution converges to a normal distribution with a covariance matrix which differs from that of the PMLE, interval estimates derived from it will not have the correct frequentist coverage, a property usually regarded as desirable, even for Bayesian estimators. This is apparent in our Monte Carlo simulations where these intervals suffer from undercoverage of the true parameter. Another disadvantage of this approach is that simple algorithms based on conjugate, or at least conditionally conjugate priors may not be applicable to the pseudo likelihood necessitating the development of entirely new sampling schemes.

2.3 A Bayesian Weighted Estimator

The incorrect frequentist coverage of interval estimates for the BPPE, and difficulties extending it to complex models, led us to explore the development of an alternative "Bayesian weighted estimator". The BWE that we propose is based on a data augmentation approach where a pseudo representative sample is treated as missing data. A representative sample $\boldsymbol{z}' = \left(z_1, z_2, \ldots, z_n\right)$ is obtained by sampling with replacement from the observed

sample $\boldsymbol{y}$ according to the probabilities implied by the weights $\tilde{\boldsymbol{w}}' = (\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n)$. This step is easily embedded into a MCMC sampling scheme where other steps are used to draw posterior observations on θ for each repeated draw of $\boldsymbol{z}$.

To describe our proposed scheme, we begin by noting that, for each sampled observation z_i, there are n possible outcomes $y_1, y_2, \ldots, y_n$ to be selected with respective probabilities $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n$. That is, $\Pr(z_i = y_j \mid \tilde{\boldsymbol{w}}) = w_j$. In any given iteration some elements in $\boldsymbol{y}$ will be included in $\boldsymbol{z}$, possibly more than once, and some will be excluded. It is convenient to introduce a variable c_j that gives the number of times y_j appears in the pseudo random sample $\boldsymbol{z}$. The vector $\boldsymbol{c}' = (c_1, c_2, \ldots, c_n)$, describing the number of times each y_j is selected into $\boldsymbol{z}$, is distributed $\boldsymbol{c} \sim$ multinomial $(n, \tilde{\boldsymbol{w}})$. That is,

$$p(\boldsymbol{c} \mid \tilde{\boldsymbol{w}}) \propto \prod_{j=1}^{n} \tilde{w}_j^{c_j} \tag{4}$$

We assume the likelihood function for our pseudo representative sample is given by

$$p(\boldsymbol{z} \mid \theta) = \prod_{i=1}^{n} p(z_i \mid \theta) = \prod_{j=1}^{n} p(y_j \mid \theta)^{c_j} \tag{5}$$

Thus, we can set up an iterative process where $\boldsymbol{c}$ is drawn from a multinomial distribution, and then, using conventional sampling methods such as MCMC, θ is drawn from the posterior density $p(\theta \mid \boldsymbol{z}) \propto p(\boldsymbol{z} \mid \theta)\, p(\theta)$, using the values $\boldsymbol{z}$ implied by $\boldsymbol{c}$. This process is repeated for a large number of draws of $\boldsymbol{c}$, $\boldsymbol{z}$ and θ.

For expressing this iterative process in terms of conditional density functions, we first note that our objective is to find, or obtains draws from, the posterior density $p(\theta \mid \boldsymbol{y}, \tilde{\boldsymbol{w}})$. Our iterative scheme augments this posterior with draws on $\boldsymbol{c}$ and $\boldsymbol{z}$, so that we have

$$p(\theta, \boldsymbol{z}, \boldsymbol{c} \mid \boldsymbol{y}, \tilde{\boldsymbol{w}}) = p(\boldsymbol{c} \mid \boldsymbol{y}, \tilde{\boldsymbol{w}})\, p(\boldsymbol{z} \mid \boldsymbol{c}, \boldsymbol{y}, \tilde{\boldsymbol{w}})\, p(\theta \mid \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{y}, \tilde{\boldsymbol{w}}) \tag{6}$$

Now, $p(\boldsymbol{c}\,|\,\boldsymbol{y},\tilde{\boldsymbol{w}}) = p(\boldsymbol{c}\,|\,\tilde{\boldsymbol{w}})$ is multinomial $(n,\tilde{\boldsymbol{w}})$. The expression $p(\boldsymbol{z}\,|\,\boldsymbol{c},\boldsymbol{y},\tilde{\boldsymbol{w}})$ will be equal to 1 as long as y_j appears in $\boldsymbol{z}$ exactly c_j times and 0 otherwise. The last density in (6), $p(\theta\,|\,\boldsymbol{z},\boldsymbol{c},\boldsymbol{y},\tilde{\boldsymbol{w}})$, is not clearly defined, and so we approximate it with $p(\theta\,|\,\boldsymbol{z})$, relying on the repeated sampling of $\boldsymbol{z}$ from $\boldsymbol{y}$ to make it a good approximation in large samples. The results of our Monte Carlo experiments that examine the frequency properties of the posterior mean as an estimator support this conjecture.

Subject to the approximation, we can then set up the following general algorithm for drawing observations $\theta^{(1)},\theta^{(2)},\ldots,\theta^{(M)}$ from the posterior $p(\theta\,|\,\boldsymbol{y},\tilde{\boldsymbol{w}})$.

Algorithm 1 Bayesian Weighted Sampling Scheme

For $s = 1:M$

1. Generate $\left(\boldsymbol{c}^{(s)}\,|\,\tilde{\boldsymbol{w}}\right)$ from multinomial$(n,\tilde{\boldsymbol{w}})$.
2. Use $\boldsymbol{c}^{(s)}$ and $\boldsymbol{y}$ to identify the composition of the pseudo sample $\boldsymbol{z}^{(s)}$.
3. Generate $\left(\theta^{(s)}\,|\,\boldsymbol{z}^{(s)}\right)$ from $p\left(\theta^{(s)}\,|\,\boldsymbol{z}^{(s)}\right) \propto \prod_{i=1}^{n} p\left(z_i^{(s)}\,|\,\theta^{(s)}\right) p\left(\theta^{(s)}\right)$.

End

In the above analysis we have attempted to ease the notational burden by considering only a single variable *y*. In practice large surveys consider many variables and it is likely that more than one will be of interest. The above analysis is readily extended to multivariate distributions for a vector of variables.

3. Simulation Studies

In this section we describe two simulation studies that serve a dual purpose – to illustrate how the Bayesian weighted estimator is implemented in two specific cases, and to compare the sampling-theory performance of a variety of weighted and unweighted Bayesian and sampling theory estimators. In the first experiment the response variable Y is assumed to

follow a normal distribution, while in the second experiment Y is assumed to follow a mixture of gamma distributions. To obtain weights we introduce a normally distributed selection variable X, where dependence between X and Y is induced via a Gaussian copula. The probability that a value of the response variable is observed depends on the selection variable via a probit link function. In both cases we assume that the weights derived from probabilities computed using the probit function are observed, but realizations of X that are used to compute the probabilities and weights are not observed.

3.1 Simulation 1: Normal Response

When both Y and X are marginally Gaussian and bound by a Gaussian copula the values (Y, X) have a bivariate normal distribution

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim BVN\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix}\right) \tag{7}$$

The variable Y is a response variable; we are interested in estimating its mean μ_y and variance σ_y^2. The variable X is a selection variable. When a sample is taken from the population, the X-value for a member of the population determines the probability of selecting that member of the population into the sample. Specifically, we assume that Y_i is selected into the sample if and only if $I_i = 1$ where

$$\Pr\left(I_i = 1 \mid Y_i, X_i\right) = \Pr\left(I_i \mid X_i\right) = \pi_i = \Phi\left(\beta_0 + \beta_1 X_i\right) \tag{8}$$

with $\Phi(\cdot)$ denoting the cumulative distribution function of a standard normal distribution. When a member of the population is selected into the sample, we observe Y_i and a weight w_i assumed to be such that $w_i \propto 1/\pi_i$, but we do not observe X_i. The selected sample is denoted as $(\boldsymbol{y}, \boldsymbol{w})$. Scaling the weights so that they sum to the sample size, we have

$w_i = n\pi_i^{-1} \Big/ \sum_{j=1}^{n} \pi_j^{-1}$. The normalized sampling weights are given by $\tilde{w}_i = w_i \Big/ \sum_{j=1}^{n} w_j$. The objective is to use $(\boldsymbol{y}, \tilde{\boldsymbol{w}})$ to estimate μ_y and σ_y^2.

The simulation setup we used is as follows: $N = 100,000$ values of (Y_i, X_i) are generated as a finite population, with $\mu_x = 0$, $\sigma_x^2 = 9$, $\mu_y = 10$, $\sigma_y^2 = \{4, 16, 100\}$ and $\rho = \{0, 0.2, 0.8\}$. A sample drawn from this population will be representative, in the sense that each population value of *Y* has an equal chance of being selected, if $\rho = 0$ or $\beta_1 = 0$. Thus, for $\beta_1 \neq 0$, the three values of ρ control the representativeness of the sample. Three different variances are used because the impact of an unrepresentative sample is potentially worse for larger variances. With larger variances, extreme values of *Y* will be systematically omitted from the sample. To obtain an observed sample, each population pair (Y_i, X_i) is assigned a probability π_i from the probit function and selected with probability π_i. The probit function parameters used for this exercise were $\beta_0 = \{-1.8, -2.7\}$ and $\beta_1 = 0.1$. For a given β_1, the setting for β_0 controls the sample size; $\beta_0 = -1.8$ leads to a sample of approximately 4000, and, for $\beta_0 = -2.7$, $n \approx 500$.

In Figure 1 we plot histograms for examples of samples of *Y* generated with $\beta_0 = -2.7$, $\beta_1 = 0.1$, $\sigma_y^2 = 16$ and the three values $\rho = \{0, 0.2, 0.8\}$. When $\rho = 0$, the sample is "representative" and the histogram is centered close to the true value $\mu_y = 10$. Increasing ρ to 0.2 moves the distribution slightly to the right centering it at $\bar{y} = 10.81$. A further increase in ρ to 0.8 leads to a substantial shift, centering the distribution at $\bar{y} = 12.81$.

We use 10,000 Monte Carlo replications to examine the performance of three Bayesian and one sampling theory estimators for μ_y. For each estimator results are reported for:

1. The average of estimates for μ_y.

2. The average of the variance estimates for each estimator for μ_y – either the relevant sampling theory estimator or the posterior variance for μ_y.
3. The coverage of 95% interval estimates for μ_y constructed using the estimates from (1) and (2).

Details of the estimators follow. Derivations are provided in the web-based supplementary material, Appendices B (the PMLE) and C (the PPBE).

1. **Pseudo MLE** (PMLE): The closed form solutions are $\hat{\mu}_{y,PMLE} = (1/n)\sum_{i=1}^{n} w_i y_i$ and $\hat{\sigma}^2_{y,PMLE} = (1/n)\sum_{i=1}^{n} w_i \left(y_i - \hat{\mu}_{y,PMLE}\right)^2$. The variance estimator for $\hat{\mu}_{y,PMLE}$ is given by the "sandwich" estimator $\hat{\sigma}^2_{\mu,PMLE} = \left(1/n^2\right)\sum_{i=1}^{n} w_i^2 \left(y_i - \hat{\mu}_{y,PMLE}\right)^2$.
2. **Unweighted Bayesian** (UBE): Using the non-informative joint prior distribution $p\left(\mu_y, \sigma_y^2\right) = 1/\sigma_y^2$, we obtain the marginal posterior densities $p\left(\sigma_y^2 \mid \boldsymbol{y}\right) = IG\left(\frac{v}{2}, \frac{v}{2}s^2\right)$, and $p\left(\mu_y \mid \boldsymbol{y}\right) = t\left(\bar{y}, \frac{v}{v-2}\frac{s^2}{n}\right)$ where $v = n-1$ and $s^2 = \frac{1}{v}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$. The posterior mean $\bar{y}$ is used as a point estimate for μ_y, and the posterior variance for μ_y is used as the variance estimate for $\bar{y}$. Except for a degrees of freedom correction which is inconsequential for the sample sizes considered here, the posterior mean and variance are identical to the mean and variance for an unweighted MLE. Thus, the results for the UBE are also indicative of those for unweighted MLE.
3. **Bayesian Pseudo Posterior** (BPPE): From the joint pseudo posterior density $\tilde{p}\left(\mu_y, \sigma_y^2 \mid \boldsymbol{y}, \boldsymbol{w}\right) \propto \sigma_y^{-2}\prod_{i=1}^{n}\left(N\left(y_i \mid \mu_y, \sigma_y^2\right)\right)^{w_i}$ we obtain the marginal distributions $\tilde{p}\left(\sigma_y^2 \mid \boldsymbol{y}, \boldsymbol{w}\right) = IG\left(\frac{v}{2}, \frac{v}{2}s^{*2}\right)$ and $\tilde{p}\left(\mu_y \mid \boldsymbol{y}, \boldsymbol{w}\right) = t\left(\bar{y}^*, \frac{v}{v-2}\frac{s^{*2}}{n}\right)$ where $\bar{y}^* = n^{-1}\sum_{i=1}^{n} w_i y_i$

and $s^{*2} = \frac{1}{v}\sum_{i=1}^{n} w_i \left(y_i - \bar{y}^*\right)^2$. The posterior mean $\bar{y}^*$ is used as a point estimate for μ_y, and the posterior variance of μ_y is used as the variance of this estimate.

4. **Bayesian Weighted** (BWE): Adapting Algorithm 1 to this setup, the augmented posterior density for the BWE can be written as

$$p\left(\mu_y, \sigma_y^2, \boldsymbol{z}, \boldsymbol{c} \mid \boldsymbol{y}, \tilde{\boldsymbol{w}}\right) = p\left(\mu_y \mid \sigma_y^2, \boldsymbol{z}\right) p\left(\sigma_y^2 \mid \boldsymbol{z}\right) p\left(\boldsymbol{z} \mid \boldsymbol{c}, \boldsymbol{y}\right) p\left(\boldsymbol{c} \mid \tilde{\boldsymbol{w}}\right) \tag{9}$$

The sampling scheme for drawing observations from this density is described in Algorithm 2. A total of $M = 1{,}000$ posterior draws were generated. The posterior draws were used to estimate posterior mean and variances for $\left(\mu_y, \sigma_y^2\right)$.

Algorithm 2 Bayesian Weighted Algorithm for Simple Gaussian Model

For $t = 1:M$

1. Generate $\left(\boldsymbol{c}^{(t)} \mid \tilde{\boldsymbol{w}}\right)$ from multinomial$\left(n, \tilde{\boldsymbol{w}}\right)$.
2. Use $\boldsymbol{c}^{(t)}$ and $\boldsymbol{y}$ to identify the composition of the pseudo sample $\boldsymbol{z}^{(t)}$.
3. Compute $\bar{z}^{(t)} = n^{-1}\sum_{i=1}^{n} z_i^{(t)}$ and $s^{2(t)} = \left(n-1\right)^{-1}\sum_{i=1}^{n}\left(z_i^{(t)} - \bar{z}^{(t)}\right)^2$ and draw $\sigma_y^{2(t)}$ from $p\left(\sigma_y^{2(t)} \mid \boldsymbol{z}^{(t)}\right) = IG\left(v/2, \left(v/2\right)s^{2(t)}\right)$, where $v = n-1$.
4. Draw $\mu_y^{(t)}$ from $p\left(\mu_y^{(t)} \mid \sigma_y^{2(t)}, \boldsymbol{z}^{(t)}\right) = N\left(\bar{z}^{(t)}, \sigma_y^{2(t)}/n\right)$.

End

The $\left(\mu_y, \sigma_y^2\right)^{(1)}, \ldots, \left(\mu_y, \sigma_y^2\right)^{(M)}$ approximate draws from the posterior distribution $p\left(\mu_y, \sigma_y^2 \mid \boldsymbol{y}, \tilde{\boldsymbol{w}}\right)$. For estimates of the posterior mean and variance of μ_y, we can use $\hat{\mu}_y = M^{-1}\sum_{t=1}^{M}\bar{z}^{(t)}$ and $\hat{\sigma}_{\mu}^2 = M^{-1}\sum_{t=1}^{M}\hat{\sigma}_y^{2(t)}/n + M^{-1}\sum_{t=1}^{M}\left(\bar{z}^{(t)} - \hat{\mu}_y\right)^2$.

The means of the point estimates for μ_y and its variance σ_{μ}^2 were calculated over $R = 10{,}000$ replications for each method. We use $\bar{\mu}_y = \left(1/R\right)\sum_{r=1}^{R}\hat{\mu}_{y,r}$ to denote the average

of the estimates of μ_y and $\bar{\sigma}^2_{\mu} = (1/R)\sum_{r=1}^{R} \hat{\sigma}^2_{\mu,r}$ to denote the average of the estimates of the variance of $\hat{\mu}_y$ where $\hat{\sigma}^2_{\mu,r}$ is the posterior variance of μ_y under Bayesian frameworks and the variance of $\hat{\mu}_y$ under a frequentist framework. In Tables 1 to 5, we report the results for $\bar{\mu}_y$ and $\bar{\sigma}^2_{\mu}$ from the various estimators, together with the coverage of 95% Bayesian credible intervals and 95% frequentist confidence intervals. A coverage less than 95% suggests that the variance of an estimate for μ_y is biased downwards and a coverage greater than 95% suggests the variance estimate is biased upwards. Table 1 contains results for the case where Y and X are uncorrelated ($\rho = 0$). Tables 2 and 3 contain results for a large observed sample size, high and low correlation ($\rho = 0.8, 0.2$) and different values for the variance of Y ($\sigma^2_y = 4, 16,$ and 100). Tables 4 and 5 contain the corresponding results for a small observed sample size. We observe that:

1. The estimates for μ_y from PMLE, BPPE, and BWE, the estimators which utilize the weights, are close to the true value $\mu_y = 10$, even when the observed sample size is only approximately 500, suggesting that any bias in these estimators in negligible. The unweighted estimator is biased, however. The amount of bias depends on three things: the true variance of Y, the degree of correlation between Y and X, and the sample size. The higher the degree of correlation ρ, the larger the true variance of Y, or the smaller the observed sample size, the larger the bias of the unweighted estimator.
2. From Table 1 where $\rho = 0,$ the mean of the unweighted estimates for the parameter μ_y is close to the true value suggesting that when Y is not correlated with X, the unweighted estimator is unbiased. Also, with the exception of BPPE, whose variance estimate consistently understates the true variance, the interval estimates have the

correct coverage. In this case the sampling design is ignorable. The PMLE and BWE have higher variance estimates on average compared to UBE, reflecting the effect of unnecessary complexity.

3. The average of the variance estimates over the replications, $\bar{\sigma}^2_{\mu}$, is always smaller for BPPE compared to PMLE and BWE (Tables 2 to 5). These smaller variance estimates for BPPE lead to interval estimate coverage that is less than the nominal 95%. Using BPPE, the variance of the estimates is underestimated since the wrong variance matrix is employed. PMLE uses the robust "sandwich estimator" to correctly estimate the variance matrix; BWE integrates out across latent variables $\boldsymbol{c}$ and $\boldsymbol{z}$ with respect to their posterior distributions.
4. Tables 4 and 5 present the results for the case of a small observed sample size. Compared to BWE, the PMLE has a higher average variance $\bar{\sigma}^2_{\mu}$, particularly when $\rho = 0.8$. This higher average is attributable to a few very extreme values for PMLE, which in turn have led to a higher coverage compared to the BWE. The BWE seems to be more robust than the PMLE although it has a slightly low coverage.
5. In most cases, the coverage of BWE is comparable in magnitude to the 95% confidence intervals of PMLE; an exception is the case of a small sample size and large ρ. This suggests that the BWE achieves the correct coverage asymptotically. The averages of the variances of the estimates of μ_y are also quite comparable for PMLE and BWE. Thus, the BWE's posterior variance can be thought of as a Bayesian way of correcting the posterior variance when the sampling weights are taken into account.
6. Increasing the variance σ^2_y increases the variance $\bar{\sigma}^2_{\mu}$, but it does not change the coverage.

3.2 Simulation 2: Finite Gamma Mixture

In this section we illustrate how to embed the Bayesian weighted estimator within an MCMC algorithm for estimation of the parameters of a more complex model. We consider a finite mixture of gamma densities with two components. The procedure can be readily extended to the case of K components. We assume that the population distribution for a response variable Y can be described by the density

$$p\left(y \mid \xi, \boldsymbol{\mu}, \boldsymbol{\nu}\right) = \xi G\left(y \mid \nu_1, \mu_1\right) + \left(1-\xi\right) G\left(y \mid \nu_2, \mu_2\right) \tag{10}$$

where ν_k is the shape parameter and μ_k is the mean of the gamma density

$$G\left(y \mid \nu_k, \mu_k\right) = \frac{\left(\nu_k / \mu_k\right)^{\nu_k}}{\Gamma\left(\nu_k\right)} y^{\nu_k - 1} \exp\left(-\frac{\nu_k y}{\mu_k}\right) \tag{11}$$

The marginal distribution of the selection variable X is assumed to be $N\left(\mu_X, \sigma_X^2\right)$, as in the first simulation and a bivariate Gaussian copula is employed to construct a joint distribution between X and Y. Steps to generate a population for (Y, X) are given in Appendix D of the supplementary material. A similar set up to simulation 1 is used to select the sample and to compute the sampling weights. For the estimation of $\left(\xi, \mu_1, \mu_2, \nu_1, \nu_2\right)'$, we assume that only the sampling weights and the sample observations $\boldsymbol{y}$ are observed.

The true parameters for the mixture of gamma densities were set as follows: $\xi = 0.6$, $\mu_1 = 208$, $\mu_2 = 700$, $\nu_1 = 3$, and $\nu_2 = 2$. Those for X were $\mu_X = 0$ and $\sigma_X^2 = 9$. The correlation ρ was set to be 0.8. The parameters $\beta_0 = -1.8$ and $\beta_1 = 0.1$ were set so that the selected sample is approximately 4% of the whole finite population distribution. The total number of Monte Carlo replications R was set at 1000.

The MCMC algorithm used to estimate the model combines that suggested by Wiper et al (2001) with our proposal for including the weights. We describe it in terms of a general

model with K components. The priors employed by Wiper et al (2001) are a Dirichlet prior for $\boldsymbol{\xi}$

$$p(\boldsymbol{\xi}) \propto \xi_1^{\phi_1-1}\xi_2^{\phi_2-1}\ldots\xi_K^{\phi_K-1}$$

an inverted gamma prior for μ_k

$$p(\mu_k) = IG(\alpha_k,\beta_k) \propto (\mu_k)^{-(\alpha_k+1)}\exp\left\{-\frac{\beta_k}{\mu_k}\right\}$$

and an exponential prior for v_k

$$p(v_k) \propto \exp(-\lambda v_k).$$

The MCMC steps are summarized as Algorithm 3.

Algorithm 3: Bayesian Weighted Algorithm for Gamma Mixture Model

Set starting values for $\boldsymbol{\xi}^{(0)} = \left(\xi_1^{(0)},\xi_2^{(0)},\ldots,\xi_K^{(0)}\right)'$, $\boldsymbol{\mu}^{(0)} = \left(\mu_1^{(0)},\mu_2^{(0)},\ldots,\mu_K^{(0)}\right)'$, and $\boldsymbol{v}^{(0)} = \left(v_1^{(0)},v_2^{(0)},\ldots,v_K^{(0)}\right)'$.

For $t = 1:M$

1. Generate $\left(\boldsymbol{c}^{(t)} \mid \tilde{\boldsymbol{w}}\right)$ from multinomial$\left(n,\tilde{\boldsymbol{w}}\right)$.
2. Use $\boldsymbol{c}^{(t)}$ and $\boldsymbol{y}$ to identify the composition of the pseudo sample $\boldsymbol{z}^{(t)}$.
3. Generate $\left(\boldsymbol{d}_i^{(t)} \mid \boldsymbol{\xi}^{(t-1)},\boldsymbol{\mu}^{(t-1)},\boldsymbol{v}^{(t-1)},\boldsymbol{z}^{(t)}\right)$ for $i = 1,2,\ldots,n$ where $\boldsymbol{d}_i = (d_{i1},d_{i2},\ldots,d_{iK})$, and d_{ik} is an indicator variable equal to 1 if the i th observation is identified as coming from the k th component of the mixture according to the probability

$$P(d_{ik} = 1 \mid \boldsymbol{z},\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{v}) = \frac{p_{ik}}{p_{i1}+p_{i2}+\ldots+p_{iK}}$$

where $$p_{ik} = \xi_k \frac{(v_k/\mu_k)^{v_k}}{\Gamma(v_k)} z_i^{v_k-1}\exp\left\{-\frac{v_k}{\mu_k}z_i\right\}.$$

Let $\boldsymbol{D}$ be the $(n\times K)$ matrix of components d_{ik} and $n_k = \sum_{i=1}^{n} d_{ik}$.

4. Generate $\left(\boldsymbol{\xi}^{(t)} \mid \boldsymbol{D}^{(t)}, \boldsymbol{\mu}^{(t-1)}, \boldsymbol{v}^{(t-1)}, \boldsymbol{z}^{(t)}\right)$ from the Dirichlet distribution

   $p\left(\boldsymbol{\xi} \mid \boldsymbol{z}, \boldsymbol{D}, \boldsymbol{\mu}, \boldsymbol{v}\right) = D\left(\boldsymbol{\phi} + \boldsymbol{n}\right)$ where $\boldsymbol{n}' = \left(n_1, n_2, \ldots, n_K\right)$ and $\boldsymbol{\phi}' = \left(\phi_1, \phi_2 \ldots, \phi_K\right)$.

5. Generate $\left(\mu_k^{(t)} \mid \boldsymbol{D}^{(t)}, \boldsymbol{\xi}^{(t)}, \boldsymbol{v}^{(t-1)}, \boldsymbol{z}^{(t)}\right)$, for $k = 1, 2, \ldots, K$ from the inverted gamma density $p\left(\mu_k \mid \boldsymbol{z}, \boldsymbol{D}, \boldsymbol{v}, \boldsymbol{\xi}\right) = IG\left(\alpha_k + n_k v_k, \beta_k + S_k v_k\right)$ where $S_k = \sum_{i=1}^{n} d_{ik} z_i$.

6. Generate $\left(v_k^{(t)} \mid \boldsymbol{D}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\xi}^{(t)}, \boldsymbol{z}^{(t)}\right)$, for $k = 1, 2, \ldots, K$ from

$$p\left(v_k \mid \boldsymbol{z}, \boldsymbol{D}, \boldsymbol{\mu}, \boldsymbol{\xi}\right) \propto \frac{v_k^{n_k v_k}}{\left[\Gamma\left(v_k\right)\right]^{n_k}} \exp\left\{-v_k\left(\lambda + \frac{S_k}{\mu_k} + n_k \log \mu_k - P_k\right)\right\}$$

   where $P_k = \sum_{i=1}^{n} d_{ik} \log z_i$. Values are drawn from this density using a Metropolis step with a gamma candidate generating function $v_k^{*(t)} \sim G\left(r_k, r_k / v_k^{(t-1)}\right)$ with r_k chosen by experimentation to obtain a reasonable acceptance rate.

7. For identification, order the elements according to $\mu_1 < \mu_2 < \cdots < \mu_K$.

End

In the Monte Carlo study, a total of 11,000 observations on $\boldsymbol{\theta} = \left(\boldsymbol{\mu}, \boldsymbol{v}, \xi\right)$ were generated for each replication, with the first 1,000 draws discarded as a burn in. A total of 1,000 replications were taken and, for a sample of the replications, the observations were plotted to confirm convergence of the Markov chains. Following Wiper et al (2001), relatively noninformative priors were used with the parameter settings $\phi_1 = \phi_2 = 1$, $\alpha = 1$, $\beta = 1$ and $\lambda = 0.01$. We also impose the a priori restriction $\mu_1 < \mu_2$ for identification of the mixture components. If the objective is estimation of the overall mixture distribution and not the individual parameters, as is the case for our empirical example in the next section, then the identification restriction is unnecessary (Geweke, 2007). In Table 6 we report the averages of

the posterior means $\bar{\theta}$, coverage of the 95% Bayesian credible intervals, and the averages of the posterior variances $\bar{\sigma}_0^2$. We observe the following.

1. The components of $\bar{\theta}$ are close to their true counterparts suggesting that any bias in the BWE is negligible. As shown in Figure 2, plots of the true density and the density using the values $\bar{\theta}$ are indistinguishable.
2. With the exception of μ_2, the averages of the posterior variances are relatively small, implying estimation is relatively precise. The coverage of the 95% credible interval for μ_1 is 0.95, and that for the other parameters is slightly less than 0.95.

Thus, we conclude the BWE algorithm works not only for the simple model described in the first simulation, but also for estimating unknown parameters of gamma mixture models. It is a very general algorithm that can be easily extended to include the usual MCMC algorithms, such as the Metropolis-Hastings and Gibbs sampling schemes. One of the primary advantages of BWE over the Bayesian pseudo posterior estimator is that, given a sampling scheme for a particular model of interest, we only need to add two additional steps to the existing sampling scheme, whereas BPPE may require a new sampling scheme.

4. Application to an Australian Income Distribution

In this section we illustrate our BWE methodology by fitting a mixture of gamma densities with 3 components to 2009 household disposable income from the HILDA survey.[1] This survey is a national longitudinal survey, which began in Australia in 2001 and is conducted annually (Watson and Wooden 2002). It was initiated and funded by the Australian Government through the Department of Families, Housing, Community Services, and Indigenous Affairs, and is designed, managed, and maintained by the Melbourne Institute of Applied Economic and Social Research, University of Melbourne. The survey is a broad

[1] An unweighted gamma mixture for a Canadian income distribution and its corresponding Lorenz curve have been estimated by Chotikapanich and Griffiths (2008).

economic and social survey that collects key variables concerning family and household structure, as well as data on education, income, health, life satisfaction and other measures of economic and subjective wellbeing. The households are sampled using a multistage sampling design; the sampling weights are provided.

Results for both the UBE and the BWE were obtained using a MCMC sample of 11,000 of which 1,000 were discarded as burn in. All the UBE and BWE parameters showed evidence of convergence. The posterior means and standard deviations are reported in Table 7. The posterior means from UBE and BWE are similar in magnitude with the exception of μ_1 where there is a marked difference. The posterior standard deviations for BWE are larger, in line with the results of our Monte Carlo experiment. In Figure 3 we plot the weighted histogram, and the density estimates at the posterior means of UBE and BWE. One major difference between the two density estimates is in their ability to capture the first mode. The weighted gamma mixture fits the first mode well, but the unweighted gamma mixture overestimates the height of the density at the mode. More generally, relative to the weighted estimates, the unweighted estimates overstate the proportion of the population in the lower portion of the distribution, and understate the proportion of the population in the upper portion of the distribution.

The different estimates of the distribution have implications for three important summary statistics that are often of interest when estimating income distributions, namely mean income μ, the Gini coefficient as a measure of inequality, *G*, and the proportion of the population below a poverty line (the headcount ratio *H*). Draws from the posterior distributions of these quantities can be obtained from the following equations.[2]

$$\mu = \sum_{k=1}^{3} \xi_k \mu_k$$

[2] The expression for the Gini coefficient for a mixture of gamma densities has been derived by Griffiths and Hajargasht (2012).

$$G = -1 + \frac{2}{\mu}\sum_{k=1}^{3}\sum_{j=1}^{3}\xi_k \xi_j \mu_k F_B\left(x_{k,j}; v_j, v_{k+1}\right)$$

$$H = F_G(y_p)$$

where $F_B\left(x_{k,j}; v_j, v_{k+1}\right)$ is the distribution function for a standard beta random variable with parameters v_j and v_{k+1} evaluated at $x_{k,j} = \left(\mu_k / v_k\right) \Big/ \left(\left(\mu_k / v_k\right) + \left(\mu_j / v_j\right)\right)$ and $F_G(y_p)$ is the distribution function for the gamma mixture evaluated at a poverty line of $y_p = \$20{,}000$. The posterior means and 95% credible intervals for μ, *G* and *H* are reported in Table 8. Because the distribution that ignores the weights has led to a larger estimate for the proportion of the population in the lower portion of the distribution, the unweighted estimate for μ is smaller and that for *H* is larger than their respective estimates from the weighted distribution. Moreover, the interval estimates for μ and *H* do not overlap, implying quite distinct estimates for these quantities. The difference in estimates for the Gini coefficient is less pronounced, with the unweighted estimate suggesting greater inequality.

5. Conclusion

Empirical work in model-based inference often ignores sampling weights or makes use of the classical pseudo maximum likelihood estimator. In this paper we propose two Bayesian alternatives. Both theoretical and empirical results support the use of the BWE which is based on the generation of a representative sample as a latent variable that can be embedded within an MCMC or other simulation algorithm. We compare methods using two Monte Carlo simulations, one using a simple Gaussian model and one with a more complex mixture of gamma densities. These simulations show that the Bayesian weighted estimator has a posterior variance that is comparable to that of the sandwich covariance matrix of the pseudo maximum likelihood estimator. Also, using the pseudo likelihood within a Bayesian framework can lead to a posterior variance that understates the repeated sampling variation of

the posterior mean, a result in line with the asymptotic theory that we have derived. An additional advantage of the Bayesian weighted estimator over the pseudo maximum likelihood estimator is that it can easily be applied to a general set of possibly complex models that can be estimated by MCMC. In an application to estimation of an Australian income distribution, we illustrate how to estimate the parameters of a three component gamma mixture model, and how to obtain posterior densities for economic quantities of interest that depend on those parameters. We find that inference about the quantities of interest – mean income, the Gini coefficient and the headcount ratio – can be sensitive to exclusion or inclusion of the weights in the analysis.

Appendix A: Asymptotic Properties of the BPPE

A.1 Consistency

Under some regularity conditions, Walker (1969) derived the asymptotic behavior of proper posterior distributions under unweighted, independent, and identically distributed observations. Gelman et al. (2014) and Le Cam and Yang (1990) provide reviews of this area. For convenience of exposition, we provide the standard result for a scalar θ. Let $\boldsymbol{y}$ be a $n \times 1$ random vector of finite population observations. Some aspect of the distribution of $\boldsymbol{y}$ depends on a parameter θ, contained in a parameter space Θ. Assume that Θ is closed set of points on the real line. Also assume that θ_0 is the true parameter and unique solution to the population maximization problem $\theta_0 = \max_{\theta \in \Theta} E\left(\log p\left(y \mid \theta\right)\right)$. For a random observed sample of size n, $\left\{y_i : i = 1, 2, \ldots, n\right\}$, we also draw I_i, which is a binary indicator variable that is equal to 1 of the observation i is used in estimation. The observation y_i is observed if and only if $I_i = 1$. The sampling weights are defined as the inverse of probability of inclusion

$w_i = 1/\pi_i$. Let π_i be the probability that unit i is in the sample, conditional on demographic characteristics $\boldsymbol{D}_i$, that is, $\pi_i = \Pr(I_i = 1 \mid \boldsymbol{D} = \boldsymbol{D}_i)$.

Given the data $\boldsymbol{y}' = (y_1, y_2, \ldots, y_n)$ and the sampling weights $\boldsymbol{w}' = (w_1, w_2, \ldots, w_n)$ and provided that the prior density $p(\theta)$ is continuous and positive, the pseudo posterior distribution can be written as:

$$\tilde{p}(\theta \mid \boldsymbol{y}, \boldsymbol{w}) \propto \prod_{i=1}^{n} p(y_i \mid \theta)^{I_i w_i} \, p(\theta)$$

Taking logs and dividing by n gives

$$\frac{1}{n} \log \tilde{p}(\theta \mid \boldsymbol{y}, \boldsymbol{w}) \propto \frac{1}{n} \sum_{i=1}^{n} I_i w_i \log p(y_i \mid \theta) + \frac{1}{n} \log p(\theta)$$

Let $\hat{\theta}$ be the posterior mode defined as:

$$\hat{\theta} = \max_{\theta \in \Theta} \left(\frac{1}{n} \sum_{i=1}^{n} I_i w_i \log p(y_i \mid \theta) + \frac{1}{n} \log p(\theta) \right)$$

As $n \to \infty$, the influence of the prior diminishes and the pseudo posterior is dominated by the influence of the pseudo likelihood. Given the prior $p(\theta)$ is non-zero at $\theta = \theta_0$, $n^{-1} \log p(\theta) \to 0$, and by the usual weak law of large numbers,

$$\frac{1}{n} \sum_{i=1}^{n} I_i w_i \log p(y_i \mid \theta) = E\left(\frac{I_i}{\pi_i} \log p(y_i \mid \theta) \right)$$

By using the law of iterated expectations, we have:

$$\begin{aligned} E\left(\frac{I_i}{\pi_i} \log p(y_i \mid \theta) \right) &= \iiint \left(\frac{I_i}{\pi_i} \log p(y_i \mid \theta) \right) p(y, I, \boldsymbol{D}) \, dy \, dI \, d\boldsymbol{D} \\ &= \iint \left(\frac{\int I_i p(I \mid y, \boldsymbol{D}) \, dI}{\pi_i} \log p(y_i \mid \theta) \right) p(y, \boldsymbol{D}) \, dy \, d\boldsymbol{D} \end{aligned}$$

$$= \iint \left\{ \frac{\pi_i}{\pi_i} \log p\left(y_i \mid \theta\right) \right\} p\left(y \mid \boldsymbol{D}\right) dy d\boldsymbol{D}$$
$$= \int \log p\left(y_i \mid \theta\right) p\left(y\right) dy \int p\left(\boldsymbol{D} \mid y\right) d\boldsymbol{D}$$
$$= E\left(\log p\left(y_i \mid \theta\right)\right)$$

where the third equality follows from $E\left(I_i \mid y_i, \boldsymbol{D}_i\right) = \Pr\left(I_i = 1 \mid \boldsymbol{D} = \boldsymbol{D}_i\right) = \pi_i$. Because θ_0 is assumed to uniquely maximize $E\left\{\log p\left(y_i \mid \theta\right)\right\}$ from assumption 1, we have $\operatorname{plim}_{n\to\infty} \hat{\theta} = \theta_0$.

A.2 Asymptotic Normality of BPPE

Let $N_{\hat{\theta}}\left(\varepsilon\right) = \left\{\theta : \left|\theta - \hat{\theta}\right| < \varepsilon/\sqrt{n}\right\}$ be a neighborhood of $\hat{\theta}$ contained in Θ, where $\varepsilon > 0$ is a given fixed number. Using Taylor's theorem to expand $\log \tilde{p}\left(\theta \mid \boldsymbol{y}, \boldsymbol{w}\right)$ around $\hat{\theta}$ leads to

$$\log \tilde{p}\left(\theta \mid \boldsymbol{y}, \boldsymbol{w}\right) \approx \log \tilde{p}\left(\hat{\theta} \mid \boldsymbol{y}, \boldsymbol{w}\right) + \left(\theta - \hat{\theta}\right) \left. \frac{\partial \log \tilde{p}\left(\theta \mid \boldsymbol{y}, \boldsymbol{w}\right)}{\partial \theta} \right|_{\theta=\hat{\theta}}$$
$$+ \frac{1}{2}\left(\theta - \hat{\theta}\right) \left[\left. \frac{\partial^2 \log \tilde{p}\left(\theta \mid \boldsymbol{y}, \boldsymbol{w}\right)}{\partial \theta^2} \right|_{\theta=\hat{\theta}} \right] + R$$

where R is of higher order than $\left(\theta - \hat{\theta}\right)^2$ and the term $\partial \log \tilde{p}\left(\theta \mid \boldsymbol{y}, \boldsymbol{w}\right)/\partial\theta\big|_{\theta=\hat{\theta}}$ is zero since the log posterior density has zero first derivative at the posterior mode. The first term can be treated as constant since it does not involve θ. We can say that, as $n \to \infty$, any θ in $N_{\hat{\theta}}\left(\varepsilon\right)$ will approach $\hat{\theta}$ in probability. Thus, for arbitrary small $\delta > 0$,

$$\lim_{n\to\infty} P\left[\sup_{\theta \in N_{\hat{\theta}}(\varepsilon)} \left|R\right| \le \delta \right] = 1$$

In the neighborhood $N_{\hat{\theta}}\left(\varepsilon\right)$, we can express the pseudo posterior $\tilde{p}\left(\theta \mid \boldsymbol{y}, \boldsymbol{w}\right)$ as follows, as $n \to \infty$:

$$\tilde{p}\left(\theta \mid \boldsymbol{y}, \boldsymbol{w}\right) \propto \exp\left\{ -\frac{n}{2}\left(\theta - \hat{\theta}\right)^2 \left[-\frac{1}{n} \left. \frac{\partial^2 \log \tilde{p}\left(\theta \mid \boldsymbol{y}, \boldsymbol{w}\right)}{\partial \theta^2} \right|_{\theta=\hat{\theta}} \right] \right\}$$

Now,

$$-\frac{1}{n}\frac{\partial^2 \log \tilde{p}(\theta \mid \boldsymbol{y},\boldsymbol{w})}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}} = -\frac{1}{n}\frac{d^2 \log p(\theta)}{d\theta^2}\bigg|_{\theta=\hat{\theta}} - \frac{1}{n}\sum_{i=1}^{n} w_i \frac{d^2 \log p(y_i \mid \theta)}{d\theta^2}\bigg|_{\theta=\hat{\theta}}$$

As $n \to \infty$, the first term $-\frac{1}{n}\frac{d^2 \log p(\theta)}{d\theta^2}\Big|_{\theta=\hat{\theta}}$ goes to zero and the second term $-\frac{1}{n}\sum_{i=1}^{n} w_i \frac{d^2 \log p(y_i \mid \theta)}{d\theta^2}\Big|_{\theta=\hat{\theta}}$ is the estimated weighted negative Hessian matrix evaluated at $\theta = \hat{\theta}$. Therefore, as $n \to \infty$, $\tilde{p}(\theta \mid \boldsymbol{y},\boldsymbol{w})$ converges to a normal distribution with mean $\hat{\theta}$ and variance $\sigma^2_{BPP} = \frac{1}{n}\left(-\frac{1}{n}\sum_{i=1}^{n} w_i \frac{d^2 \log p(y_i \mid \theta)}{d\theta^2}\Big|_{\theta=\hat{\theta}}\right)^{-1}$ in the neighborhood of $N_{\hat{\theta}}(\varepsilon)$. The next step is to ensure that θ_0 is in the neighborhood of $\hat{\theta}$ which follows from the consistency of $\hat{\theta}$. Also, given the symmetry of the asymptotic normal distribution, the posterior mean will similarly have a large sample variance given by σ^2_{BPP}.

Appendices B, C and D

See online supplementary material.

References

Aitkin, M. (2008), "Applications of the Bayesian Bootstrap in Finite Population Inference," *Journal of Official Statistics*, 24(1), 25-51.

Albert, J.H. and S. Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of American Statistical Association*, 88(422), 669-679.

Breunig, R. (2001), "Density Estimation for Clustered Data," *Econometric Reviews*, 20(3), 353-367.

Chib, S. (1992), "Bayesian Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, 51, 79-99.

Chotikapanich, D. and W.E. Griffiths (2008), "Estimating Income Distributions using a Mixture of Gamma Densities," in Duangkamon Chotikapanich (Ed.), *Modeling Income Distributions and Lorenz Curves* (New York: Springer).

Gelman, A. (2007), "Struggles with Survey Weighting and Regression Modeling," *Statistical Science*, 22(2), 153-164.

Gelman, A., J.B. Carlin, H.S. Stern, D. B. Dunson, A. Vehtari and D.B. Rubin (2014), *Bayesian Data Analysis*, 3rd edition, London: Chapman & Hall.

Geweke, J. (2007), "Interpretation and Inference in Mixture Models: Simple MCMC Works," *Computational Statistics and Data Analysis*, 51, 3529-3550.

Geweke, J. and G. Amisano, (2011), "Hierarchical Markov Normal Mixture Models with Applications to Financial Asset Returns," *Journal of Applied Econometrics*, 26, 1-29.

Geweke, J. and M. Keane (2007), "Smoothly Mixing Regressions," *Journal of Econometrics*, 138, 252-290.

Godambe, V.P. and M.E. Thompson (1986), Parameters of Superpopulation and Survey Population: their Relationships and Estimation," *International Statistical Review*, 54, 127-138.

Griffiths W.E. and G. Hajargasht (2012), "Using Mixtures to Estimate Income Distribution from Grouped Data," Technical Report, Department of Economics, University of Melbourne.

Hesketh, S.R., and A. Skrondal (2006), "Multilevel Modeling of Complex Surveu Data," *Journal of Royal Statistical Association Series A*, 169(4), 805-827.

Horvitz, D.G. and D.J. Thompson (1952), "A Generalisation of Sampling without Replacement from a Finite Universe," *Journal of American Statistical Association*, 47(260), 663-685.

Korinek, A., J.A. Mistiaen, and M. Ravallion (2007), "An Econometrics Method for Correcting for Unit Non-response Bias in Surveys," *Journal of Econometrics,* 136, 213-235.

Lazar, N.A. (2003), "Bayesian Empirical Likelihood", *Biometrika, 90,* 319-326.

Le Cam, L. and G.L. Yang (1990), *Asymptotics in Statistics: Some Basic Concepts*, New York: Springer-Verlag.

Molina, E.A. and C.J. Skinner (1992), "Pseudo-likelihood and Quasi-likelihood estimation for Complex Sampling Schemes," *Computational Statistics and Data Analysis*, 13, 395-505.

Monahan, J.F. and D.D. Boos (1992), "Proper likelihoods for Bayesian analysis", Biometrika, 79, 271-278.

Pfeffermann, D. (1996), "The Use of Sampling Weights for Survey Data Analysis," *Statistical Methods in Medical Research*, 5(3), 239-261.

Rao, J.N.K. and C. Wu (2010), "Bayesian Pseudo-empirical-likelihood Intervals for Complex Surveys," *Journal of the Royal Statistical Society Series B*, 72(4), 533-544.

Schennach, S.M. (2005), "Bayesian Exponentially Tilted Empirical Likelihood", *Biometrika*, 92 31-46.

Skinner, C.J. and B. Mason (2012), "Weighting in the Regression Analysis of Survey Data with a Cross-national Application," *Canadian Journal of Statistics*, 40, 697-711.

Tanner, A.M. and W.H. Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of American Statistical Association*, 82(398), 528-540.

Walker, A.M. (1969), "On the Asymptotic Behaviour of Posterior Distributions," *Journal of Royal Statistical Society Series B*, 31(1), 80-88.

Watson, N. and M. Wooden (2002), "The Household Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1 Survey Methodology," Technical Paper, HILDA project Technical Paper Series 1/02.

White, H. (1980), "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct test for Heteroskedasticity," *Econometrica*, 48(4), 817-838.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 1, 1-25.

Wiper, M., D.R. Insua, and F. Ruggeri (2001), "Mixtures of Gamma Distributions with Applications," *Journal of Computational and Graphical Statistics*, 10, 440-454.

Wooldridge, J.M. (1999), "Asymptotic Properties of Weighted M-estimators for Variable Probability Samples," *Econometrica*, 67, 1385-1406.

Wooldridge, J.M. (2001), "Asymptotic Properties of Weighted M-estimators for Standard Stratified Samples," *Econometric Theory*, 17, 451-470.

Wooldridge, J.M. (2007), "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 141, 1281-1301.

Table 1: Estimates for parameter μ_y with true values $\mu_y = 10,\ \rho = 0,$ and $n \approx 4000$

Case		UBE	PMLE	BPPE	BWE
$\sigma_y^2 = 100$	$\bar{\mu}_y$	9.9989	10.0002	10.0002	9.9997
	$\bar{\sigma}_\mu^2$	0.0236	0.0367	0.0236	0.0458
	coverage	0.9502	0.9483	0.8835	0.9742

Table 2: Estimates for parameter μ_y with true values $\mu_y = 10,\ \rho = 0.8,$ and $n \approx 4000$

Case		UBE	PMLE	BPPE	BWE
$\sigma_y^2 = 100$	$\bar{\mu}_y$	14.9004	10.0037	10.0037	10.0037
	$\bar{\sigma}_\mu^2$	0.0226	0.0507	0.0236	0.0472
	coverage	0.0000	0.9464	0.8225	0.9386
$\sigma_y^2 = 16$	$\bar{\mu}_y$	11.9599	10.0019	10.0019	10.0019
	$\bar{\sigma}_\mu^2$	0.0036	0.0081	0.0038	0.0076
	coverage	0.0000	0.9465	0.8208	0.9405
$\sigma_y^2 = 4$	$\bar{\mu}_y$	10.9794	9.9997	9.9997	9.9997
	$\bar{\sigma}_\mu^2$	0.0009	0.0020	0.0009	0.0019
	coverage	0.0000	0.9494	0.8204	0.9431

Table 3: Estimates for parameter μ_y with true values $\mu_y = 10,\ \rho = 0.2,$ and $n \approx 4000$

Case		UBE	PMLE	BPPE	BWE
$\sigma_y^2 = 100$	$\bar{\mu}_y$	11.2248	10.0007	10.0007	10.0008
	$\bar{\sigma}_\mu^2$	0.0236	0.0376	0.0236	0.0472
	coverage	0.0000	0.9500	0.8793	0.9696
$\sigma_y^2 = 16$	$\bar{\mu}_y$	10.4913	10.0005	10.0005	10.0006
	$\bar{\sigma}_\mu^2$	0.0038	0.0060	0.0038	0.0076
	coverage	0.0000	0.9502	0.8761	0.9713
$\sigma_y^2 = 4$	$\bar{\mu}_y$	10.2446	9.9998	9.9998	9.9998
	$\bar{\sigma}_\mu^2$	0.0009	0.0015	0.0009	0.0019
	coverage	0.0000	0.9488	0.8810	0.9695

Table 4: Estimates for parameter μ_y with true values $\mu_y = 10,\ \rho = 0.8,$ and $n \approx 500$

Case		UBE	PMLE	BPPE	BWE
$\sigma_y^2 = 100$	$\bar{\mu}_y$	16.6706	10.0339	10.0339	10.0337
	$\bar{\sigma}_\mu^2$	0.1973	0.7285	0.2056	0.4091
	coverage	0.0000	0.9158	0.6945	0.8487
$\sigma_y^2 = 16$	$\bar{\mu}_y$	12.6668	10.0113	10.0113	10.0114
	$\bar{\sigma}_\mu^2$	0.0315	0.1169	0.0329	0.0656
	coverage	0.0000	0.9244	0.7096	0.8566
$\sigma_y^2 = 4$	$\bar{\mu}_y$	11.3329	10.0046	10.0046	10.0046
	$\bar{\sigma}_\mu^2$	0.0079	0.0290	0.0082	0.0164
	coverage	0.0000	0.9253	0.7020	0.8592

Table 5: Estimates for parameter μ_y with true values $\mu_y = 10,\ \rho = 0.2,$ and $n \approx 500$

Case		UBE	PMLE	BPPE	BWE
$\sigma_y^2 = 100$	$\bar{\mu}_y$	11.6657	10.0045	10.0045	10.0045
	$\bar{\sigma}_\mu^2$	0.2066	0.4759	0.2069	0.4121
	coverage	0.0450	0.9425	0.7976	0.9303
$\sigma_y^2 = 16$	$\bar{\mu}_y$	10.6661	10.0041	10.0041	10.0041
	$\bar{\sigma}_\mu^2$	0.0330	0.0752	0.0330	0.0657
	coverage	0.0439	0.9449	0.8019	0.9278
$\sigma_y^2 = 4$	$\bar{\mu}_y$	10.3332	10.0012	10.0012	10.0012
	$\bar{\sigma}_\mu^2$	0.0083	0.0189	0.0083	0.0165
	coverage	0.0447	0.9454	0.8087	0.9347

Table 6: BWE for gamma mixture parameters

		ξ	μ_1	μ_2	ν_1	ν_2
True		0.6000	208.0000	700.0000	3.0000	2.0000
BWE	$\bar{\theta}$	0.5926	208.1776	701.3392	3.1381	2.0814
	$\bar{\sigma}_\theta^2$	0.0019	38.8275	2650.0000	0.0428	0.1061
	coverage	0.8900	0.9520	0.8960	0.8310	0.8970

Table 7: Posterior Summary Statistics for the Parameters of Individual Disposable Income 2009 (posterior standard deviation in brackets)

	ξ_1	ξ_2	μ_1	μ_2	μ_3	ν_1	ν_2	ν_3
BWE	0.0566 (0.0061)	0.9114 (0.0080)	750.41 (103.6512)	751.35 (8.2639)	163.36 (2.3460)	0.2302 (0.0213)	2.7242 (0.0845)	88.8678 (22.2388)
UBE	0.0571 (0.0051)	0.8999 (0.0071)	630.36 (73.7848)	723.43 (6.2133)	164.81 (1.8207)	0.2120 (0.0161)	2.6102 (0.0630)	90.1537 (18.5516)

Table 8: Posterior Summary Statistics of Mean Income, Gini, and Headcount for 2009 (95% credible intervals in brackets)

	Unweighted	Weighted
μ ($'00)	694.09 (681.75, 706.93)	732.51 (714.85, 751.24)
G	0.3828 (0.3758, 0.3905)	0.3747 (0.3656, 0.3850)
HC	0.1380 (0.1306, 0.1456)	0.1162 (0.1078, 0.1252)

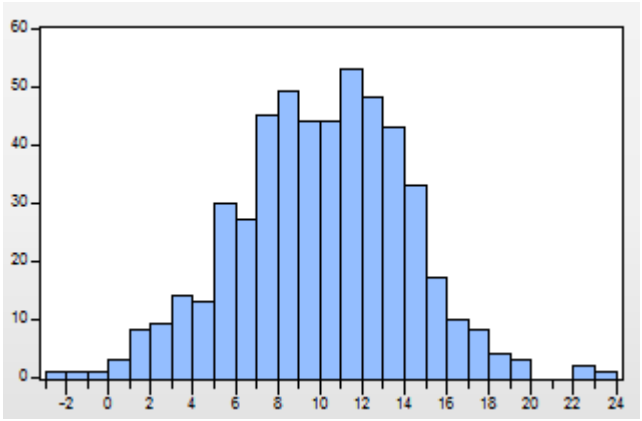

Figure 1(a) Sample with no selection, $\rho = 0$, $\bar{y} = 10.09$, $s_y = 3.97$, $n = 511$

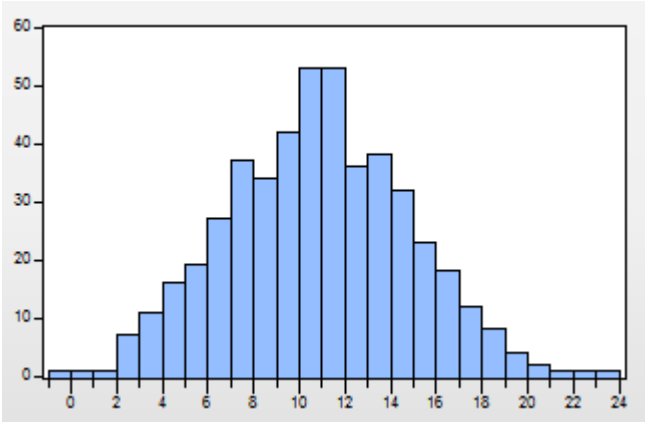

Figure 1(b) Selected sample with $\rho = 0.2$, $\bar{y} = 10.81$, $s_y = 3.98$, $n = 478$

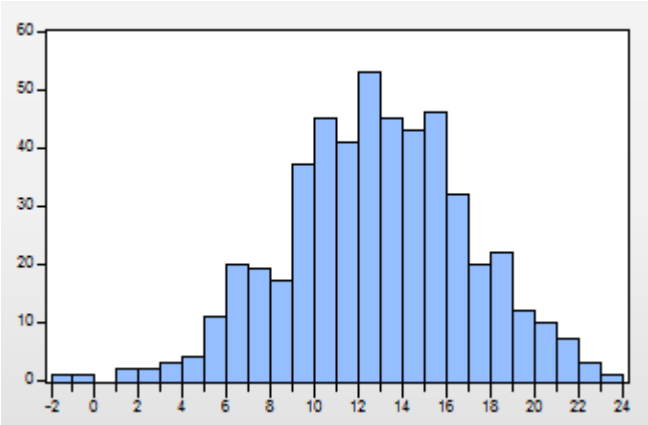

Figure 1(c) Selected sample with $\rho = 0.8$, $\bar{y} = 12.81$, $s_y = 4.11$, $n = 497$

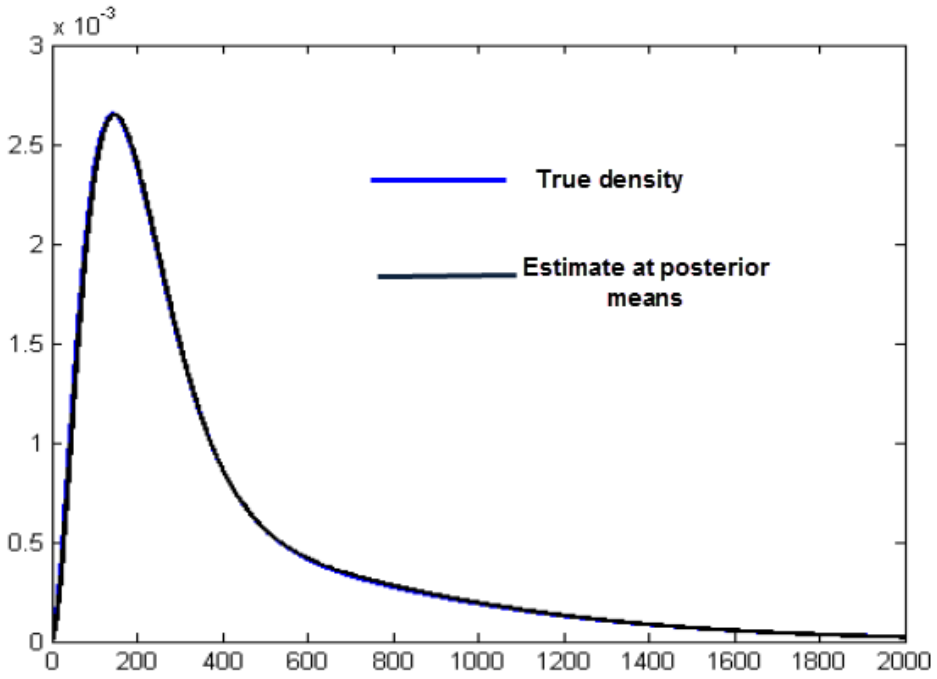

Figure 2 A gamma mixture density and its estimate from the posterior means of the Bayesian weighted estimator

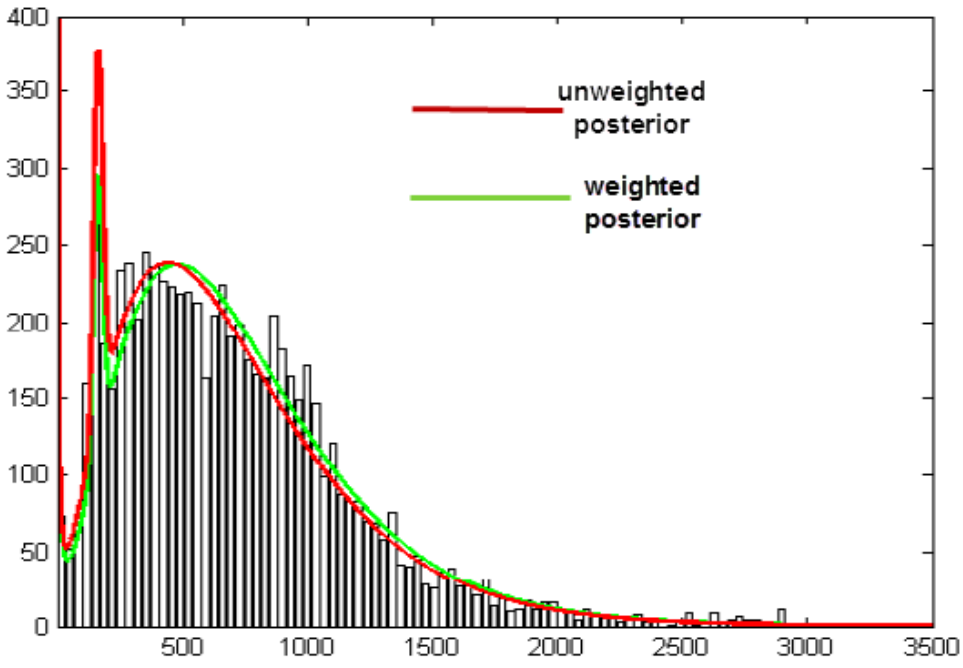

Figure 3 Weighted histogram and unweighted and weighted gamma mixture densities (at posterior means of parameters) for Australian household disposable income in 2009 ($'00).

Online Supplementary material for *Bayesian Weighted Inference from Surveys*

David Gunawan
University of New South Wales

Anastasios Panagiotelis
Monash University

William Griffiths
University of Melbourne

Duangkamon Chotikapanich
Monash University

23 March 2017

Appendix B: Covariance Matrix for PMLE For Gaussian Model

For the covariance matrix of the PMLE, we can write $\hat{\boldsymbol{V}}_w$ and the estimated Hessian $\hat{\boldsymbol{H}}_w$ as:

$$\hat{\boldsymbol{V}}_w = \frac{1}{n}\sum_{i=1}^{n} w_i^2 \left[\begin{array}{cc} \left(\frac{\partial \log p\left(y_i \mid \mu, \sigma^2\right)}{\partial \mu}\right)^2 & \frac{\partial \log p\left(y_i \mid \mu, \sigma^2\right)}{\partial \mu}\frac{\partial \log p\left(y_i \mid \mu, \sigma^2\right)}{\partial \sigma^2} \\ \frac{\partial \log p\left(y_i \mid \mu, \sigma^2\right)}{\partial \mu}\frac{\partial \log p\left(y_i \mid \mu, \sigma^2\right)}{\partial \sigma^2} & \left(\frac{\partial \log p\left(y_i \mid \mu, \sigma^2\right)}{\partial \sigma^2}\right)^2 \end{array} \right]\Bigg|_{\mu=\hat{\mu}_{y,PMLE},\sigma^2=\hat{\sigma}^2_{y,PMLE}}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \left[\begin{array}{cc} \left(\frac{w_i}{\sigma^2}(y_i-\mu)\right)^2 & \frac{w_i}{\sigma^2}(y_i-\mu)\frac{w_i}{2\sigma^4}\left[(y_i-\mu)^2-\sigma^2\right] \\ \frac{w_i}{\sigma^2}(y_i-\mu)\frac{w_i}{2\sigma^4}\left[(y_i-\mu)^2-\sigma^2\right] & \left(\frac{w_i}{2\sigma^4}\left[(y_i-\mu)^2-\sigma^2\right]\right)^2 \end{array} \right]\Bigg|_{\mu=\hat{\mu}_{y,PMLE},\sigma^2=\hat{\sigma}^2_{y,PMLE}}$$

and

$$\hat{\boldsymbol{H}}_w = \frac{1}{n}\sum_{i=1}^{n} \left[\begin{array}{cc} -\frac{w_i}{\sigma^2} & -\frac{w_i}{\sigma^4}(y_i-\mu) \\ -\frac{w_i}{\sigma^4}(y_i-\mu) & \frac{w_i}{2\sigma^4}-\frac{w_i}{\sigma^6}(y_i-\mu)^2 \end{array} \right]\Bigg|_{\mu=\hat{\mu}_{y,PMLE},\sigma^2=\hat{\sigma}^2_{y,PMLE}}$$

After substitution of the estimators for μ and σ^2, direct multiplication shows that the first diagonal element of $n^{-1}\hat{\boldsymbol{H}}_w^{-1}\hat{\boldsymbol{V}}_w\hat{\boldsymbol{H}}_w^{-1}$ is given by $\sum_{i=1}^{n} w_i^2\left(y_i-\hat{\mu}_{y,PMLE}\right)^2 \Big/ n^2$.

Appendix C: Derivation of BPPE for Gaussian Model

The joint pseudo posterior density is given by

$$\tilde{p}\left(\mu_y, \sigma_y^2 \mid \boldsymbol{y}, \boldsymbol{w}\right) \propto \prod_{i=1}^{n}\left(\left(2\pi\sigma_y^2\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma_y^2}\left(y_i-\mu_y\right)^2\right)\right)^{w_i}\left(\frac{1}{\sigma_y^2}\right)$$

$$\propto \prod_{i=1}^{n}\left(\left(\sigma_y^2\right)^{-\frac{w_i}{2}}\exp\left(-\frac{w_i}{2\sigma_y^2}\left(y_i-\mu_y\right)^2\right)\right)\left(\frac{1}{\sigma_y^2}\right)$$

$$\propto \left(\left(\sigma_y^2\right)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma_y^2}\sum_{i=1}^{n} w_i\left(y_i-\mu_y\right)^2\right)\right)\left(\frac{1}{\sigma_y^2}\right)$$

Let $\bar{y}^* = n^{-1}\sum_{i=1}^{n} w_i y_i$. We can rewrite the exponent as

$$\begin{aligned}\sum_{i=1}^{n} w_i\left(y_i-\mu_y\right)^2 &= \sum_{i=1}^{n} w_i\left(\left(y_i-\bar{y}^*\right)-\left(\mu_y-\bar{y}^*\right)\right)^2 \\ &= \sum_{i=1}^{n} w_i\left(y_i-\bar{y}^*\right)^2+\sum_{i=1}^{n} w_i\left(\mu_y-\bar{y}^*\right)^2-2\left(\mu_y-\bar{y}^*\right)\sum_{i=1}^{n} w_i\left(y_i-\bar{y}^*\right) \\ &= \sum_{i=1}^{n} w_i\left(y_i-\bar{y}^*\right)^2+n\left(\mu_y-\bar{y}^*\right)^2\end{aligned}$$

Let $s^{*2}=\left(n-1\right)^{-1}\sum_{i=1}^{n} w_i\left(y_i-\bar{y}^*\right)^2$. The joint pseudo posterior density $\tilde{p}\left(\mu_y,\sigma_y^2\mid \boldsymbol{y}\right)$ can be written as:

$$\tilde{p}\left(\mu_y,\sigma_y^2\mid \boldsymbol{y},\boldsymbol{w}\right)\propto\left(\sigma_y^2\right)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma_y^2}\left(\left(n-1\right)s^{*2}+n\left(\mu_y-\bar{y}^*\right)^2\right)\right)\left(\frac{1}{\sigma_y^2}\right)$$

Integrating out σ_y^2 and μ_y from this joint posterior yields the marginal pseudo posterior densities $\tilde{p}\left(\mu_y\mid \boldsymbol{y},\boldsymbol{w}\right)=t\left(\bar{y},\frac{\nu}{\nu-2}\frac{s^{*2}}{n}\right)$ and $\tilde{p}\left(\sigma_y^2\mid \boldsymbol{y},\boldsymbol{w}\right)=IG\left(\nu/2,\nu s^{*2}/2\right)$.

Appendix D: Steps for Generating Population Values for Gamma Mixture Model

1. Generate $N=100{,}000$ observations on $\left(q_{1i},q_{2i}\right), i=1,2,\ldots,N$ from the bivariate normal distribution with correlation parameter ρ

$$\begin{pmatrix}q_1\\ q_2\end{pmatrix}\sim BVN\left(\begin{pmatrix}0\\ 0\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

2. Compute the copula data $u_{1i}=\Phi\left(q_{1i}\right)$ and $u_{2i}=\Phi\left(q_{2i}\right), i=1,2,\ldots,N.$

3. Compute $\left(Y_i,X_i\right), i=1,2,\ldots,N$ by setting $X_i=F_N^{-1}\left(u_{1i}\right)$ and $Y_i=F_G^{-1}\left(u_{2i}\right)$ where F_N and F_G are the distribution functions for the $N\left(\mu_X,\sigma_X^2\right)$ distribution, and the gamma mixture distribution in (10), respectively. Computing $F_N^{-1}\left(u_{1i}\right)$ is straightforward, but computing $F_G^{-1}\left(u_{2i}\right)$ is not. For this purpose we use the following algorithm.

Algorithm 4: Computing Gamma Mixture Quantiles

Given values $u_{2i}, i = 1,2,\ldots,n$ that lie in the interval (0,1), the objective is to find quantiles $Y_i = F_G^{-1}(u_{2i})$ where F_G is the distribution function for the two-component gamma mixture whose density function, given in equation (18), is

$$p(y \mid \xi, \boldsymbol{\mu}, \boldsymbol{\nu}) = \xi G(y \mid \nu_1, \mu_1) + (1-\xi) G(y \mid \nu_2, \mu_2) \tag{D.1}$$

1. Generate a large number of draws, say 100,000, from (D.1) and sort them from lowest to highest. Let the vector of these draws be denoted by **x** and the *j*th ordered value by x_j.

2. Find the 100,000 cumulative proportions $\boldsymbol{F}_G(\mathbf{x})$.

3. Set a tolerable error ε. We used $\varepsilon = 10^{-4}$.

4. For $i = 1:n$

   4a. Find the smallest j, call it j_i, for which $F_G(x_j) \geq u_{2i}$.

   4b. Find an initial value $\hat{y}_1 = F_G^{-1}(u_{2i})$ as a draw from the uniform distribution $U(x_{j_i-1}, x_{j_i})$.

   4c. Compute $F_G(\hat{y}_1)$. If $|F_G(\hat{y}_1) - u_{2i}| > \varepsilon$, go to step 4d; otherwise, go to step 4e.

   4d. We improve on the initial value $\hat{y}_1$ as follows. If $F_G(\hat{y}_1) - u_{2i} > 0$, generate a random draw $\hat{y}_2 = U(x_{j_i-1}, \hat{y}_1)$. If $F_G(\hat{y}_1) - u_{2i} < 0$, generate $\hat{y}_2 = U(\hat{y}_1, x_{j_i})$. Set $\hat{y}_1 = \hat{y}_2$ and go back to step 4c.

   4e. Set $Y_i = \hat{y}_1$.

   4f. End