GLS Estimation and Empirical Bayes Prediction for Linear Mixed Models with Heteroskedasticity and Sampling Weights

A Background Study for the POVMAP Project

Roy van der Weide

WORLD BANK GROUP

Development Research Group

Poverty and Inequality Team

September 2014

Abstract

This note adapts results by Huang and Hidirolou (2003) on Generalized Least Squares estimation and Empirical Bayes prediction for linear mixed models with sampling weights. The objective is to incorporate these results into

the poverty mapping approach put forward by Elbers et al. (2003). The estimators presented here have been implemented in version 2.5 of POVMAP, the custom-made poverty mapping software developed by the World Bank.

This paper is a product of the Poverty and Inequality Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at rvanderweide@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

GLS Estimation and Empirical Bayes Prediction for Linear Mixed Models with Heteroskedasticity and Sampling Weights

A background study for the POVMAP project

Roy van der Weide*

Keywords: linear mixed models, small area estimation, Empirical Bayes, sampling weights, poverty, inequality

JEL Classification: I32, C31, C43, C53

*World Bank Research Department. Email: rvanderweide@worldbank.org. A big thank you goes to Chris Elbers for providing comments on an earlier version of this note.

1 Introduction

The poverty mapping approach put forward by Elbers et al. (2003; henceforward ELL) makes it possible to estimate poverty and inequality at a highly disaggregated level. Depending on the geography of the country of interest, estimates of poverty might be obtained for areas as small as a city or community, which greatly facilitates the targeting of the poor among other applications (see e.g. Elbers et al., 2007). ELL achieve this by means of a massive out-of-sample prediction exercise that “imputes” income or consumption data for every household recorded in a population census. Once estimates of consumption are available for all households in the population this data can then be aggregated at almost any desired level of aggregation. The household consumption model used for prediction is estimated to data from a household income survey where the independent variables are restricted to those that are available in both the survey and the census.

A linear mixed model is assumed which is standard in the small area estimation literature (see e.g. Rao, 2003). Spatial correlation between the residuals is accounted for by means of a nested error structure that consists of a random area effect and an idiosyncratic household effect. ELL believed their approach would be most convincing if the assumptions about the errors are kept to a minimum. Specifically, the household errors are allowed to be heteroskedastic and by default no assumptions are made about the shape of the error distribution functions.

The ELL approach has been applied to obtain poverty maps in over 60 countries worldwide. Part of this success may arguably be attributed to its implementation in POVMAP, a custom-made software package developed by the World Bank that can be downloaded from the public domain at no cost.¹ The POVMAP project has made ELL, a computationally intensive approach, available to a large audience of applied users (and has thereby greatly lowered the threshold for adopting ELL). The first version of POVMAP (i.e. POVMAP 1.0) ran under MS-DOS. A graphical user interface was added with the second version (POVMAP 2.0). Both versions of POVMAP closely follow the procedures from the original ELL publication.

A decade has past since the original publication, a good time to take stock of new developments. The developments that we will focus on in this note are Empirical Bayes (EB) prediction married with the ELL approach (see Molina and Rao, 2010) while accounting for unequal sampling probabilities in the income survey (see Huang and Hidiroglou, 2003). EB prediction utilizes the survey data to narrow down the random area effects while non-EB prediction (i.e. conventional ELL) makes no such attempt.

¹POVMAP2 can be downloaded from: iresearch.worldbank.org/PovMap/

As such, EB will only make a difference for areas that are represented in the survey (for other areas EB reduces to conventional ELL prediction).²

The objective of this note is to adapt the results by Huang and Hidirolou (2003) on EB prediction for generalized linear mixed models with sampling weights to the ELL framework. Note that while the original paper by Molina and Rao (2010) implements EB prediction by assuming homoskedastic errors, this assumption is easily relaxed, as can be seen in this note. The introduction of sampling weights (as probability weights) also concerns the estimation of the model parameters, which in this case involves a modification to Generalized Least Squares (GLS). Our note functions as a background study for a milestone upgrade of POVMAP to version 2.5.

Following Huang and Hidirolou (2003) and Molina and Rao (2010) we assume normally distributed errors when we present Empirical Bayes prediction. For a treatment of EB under less restrictive assumptions, see the recent study by Elbers and van der Weide (2014). In POVMAP 2.5, the user will have a choice between normal EB (Molina and Rao, 2010) and non-normal non-EB (ELL). The relative performance of these two options will depend on: (a) the size of the random area effect; (b) the number of small areas represented in the survey; and (c) the degree of non-normality of the errors. Normal EB prediction is expected to do well if there are relatively large random area effects, if many of the small areas are covered by the survey, while the error distributions can be reasonably well approximated by a normal distribution.

The outline of the note is as follows. Section 2 introduces the model framework and some notation. Section 3 presents the modification to the GLS estimator due to the introduction of probability weights. EB prediction is presented in Section 4, where we explicitly allow both for sampling weights and heteroskedasticity.

2 Model and notation

Suppose that the (log) consumption data can be described by the following nested error regression model:

$$y_{ah} = \beta^T x_{ah} + u_a + \varepsilon_{ah}, \quad (1)$$

where the subscript ah refers to household h in area a , where y_{ah} denotes (log) household per capita consumption, x_{ah} denotes a vector containing m independent variables, and where u_a and ε_{ah} represent the area error and the household specific error with zero

²The challenge is to identify the conditional distribution for the area error. When both the area and the household errors are normally distributed, it follows that the area error conditional on the survey data will also be normally distributed. If we allow the errors to be non-normally distributed however, then working out the conditional distribution will no longer be a trivial exercise (see Elbers and van der Weide, 2014).

means and variances denoted by σ_u^2 and $\sigma_{\varepsilon,ah}^2$, respectively. The two errors are assumed independent from each other. Note that $\sigma_{\varepsilon,ah}^2$ is permitted to vary between households, while σ_u^2 is assumed to be a constant. For ease of exposition, we will assume that the variance parameters are known.³

Let n_a denote the number of households sampled in area a , so that $n = \sum_a n_a$ denotes the total sample size. Let w_{ah} denote the sampling weight for household ah . Let us also define W as the diagonal matrix with the sampling weights w_{ah} along the diagonal (sorted by area), and define Ω as a diagonal matrix with the following matrices along its diagonal (sorted by area): $\Omega_a = \left(\frac{\sum_h w_{ah}}{\sum_h w_{ah}^2} \right) I_{n_a}$, where I_{n_a} denotes the identity matrix of dimension n_a .

We will at times also represent the model in matrix notation:

$$y = X\beta + u + \varepsilon. \quad (2)$$

Let $R = E[\varepsilon\varepsilon^T]$ denote the diagonal matrix with the household error variances $\sigma_{\varepsilon,ah}^2$ along the diagonal (sorted by area). We will denote the diagonal block of R corresponding to area a by R_a . Similarly, let $Q = E[uu^T]$ be the block-diagonal matrix where the blocks are given by $Q_a = \sigma_u^2 1_{n_a} 1_{n_a}^T$, where 1_{n_a} denotes the unit vector of length n_a .

3 Estimation of β using GLS

You and Rao (2002) derive a GLS estimator for β with sampling weights under the assumption that $\sigma_{\varepsilon,ah}^2 = \sigma_\varepsilon^2$ for all households by solving weighted moment conditions. Huang and Hidirolou (2003) have relaxed this assumption by permitting heteroskedasticity, i.e. a non-constant $\sigma_{\varepsilon,ah}^2$. Their GLS estimator reduces to the estimator of You and Rao (2002) if one were to insert constant variances (which we will confirm).

The weighted GLS estimator for β from Huang and Hidirolou (2003) satisfies:

$$\hat{\beta}_w = (X^T \hat{V}_w^{-1} X)^{-1} X^T \hat{V}_w^{-1} y, \quad (3)$$

with:

$$\hat{V}_w = W^{-1} \hat{R} + \Omega \hat{Q}, \quad (4)$$

where the two matrices W and Ω are functions of the sampling weights only (see Section 2).

³See the Annex for the estimation of σ_u^2 and $\sigma_\varepsilon^2 = E[\sigma_{\varepsilon,ah}^2]$. For the estimation of the conditional variances $\sigma_{\varepsilon,ah}^2$ we refer the reader to Elbers et al. (2003).

The variance of $\hat{\beta}_w$ can be estimated by:

$$\text{var}[\hat{\beta}_w] = (X^T \hat{V}_w^{-1} X)^{-1} (X^T \hat{V}_w^{-1} \hat{V} \hat{V}_w^{-1} X) (X^T \hat{V}_w^{-1} X)^{-1}, \quad (5)$$

with:

$$\hat{V} = \hat{R} + \hat{Q}. \quad (6)$$

Note that $\hat{V}$ and $\hat{V}_w$ are two different matrices. Also note that $\hat{\beta}_w$ reduces to the conventional GLS estimator if we insert constant sampling weights.

3.1 Expanding the expressions for $\hat{\beta}_w$ and $\text{var}[\hat{\beta}_w]$

In this subsection we will attempt to further work out the expressions for $\hat{\beta}_w$ and $\text{var}[\hat{\beta}_w]$ with the objective to ease implementation. Note that $\hat{\beta}_w$ is a function of $\hat{V}_w^{-1}$. We will drop the “hat” to ease notation. Due to the block-diagonal nature of V_w , we have that its inverse V_w^{-1} too will be block-diagonal where its blocks solve the inverse of the blocks of V_w .

This allows us to re-write the expression for $\hat{\beta}_w$ as follows:

$$\hat{\beta}_w = (X^T V_w^{-1} X)^{-1} X^T V_w^{-1} y \quad (7)$$

$$= \left(\sum_a X_a^T V_{a,w}^{-1} X_a \right)^{-1} \left(\sum_a X_a^T V_{a,w}^{-1} y_a \right), \quad (8)$$

where $V_{a,w}$ denotes the area a block of V_w , and where X_a and y_a denote the corresponding area a “blocks” of X and y , respectively, containing only the rows from area a . To further expand this expression let us work out the inverse of $V_{a,w}$. Note that $V_{a,w} = W_a^{-1} R_a + \Omega_a Q_a$, where W_a and R_a are both diagonal matrices of dimension n_a with w_{ah} and $\sigma_{\varepsilon,ah}^2$ along their diagonal, respectively. Recall that Q_a is defined as $Q_a = \sigma_u^2 1_{n_a} 1_{n_a}^T$, where 1_{n_a} denotes the unit vector of length n_a .

It will be convenient to represent the blocks $V_{a,w}$ as follows:

$$V_{a,w} = R_{a,w} + \sigma_u^2 \left(\frac{\sum_h w_{ah}}{\sum_h w_{ah}^2} \right) 1_{n_a} 1_{n_a}^T, \quad (9)$$

where $R_{a,w}$ is a diagonal matrix of dimension n_a with diagonal elements given by $\frac{\sigma_{\varepsilon,ah}^2}{w_{ah}}$. The inverse of $V_{a,w}$ is then seen to solve:

$$V_{a,w}^{-1} = R_{a,w}^{-1} - \left(\frac{\gamma_{a,w}}{1_{n_a}^T R_{a,w}^{-1} 1_{n_a}} \right) R_{a,w}^{-1} 1_{n_a} 1_{n_a}^T R_{a,w}^{-1}, \quad (10)$$

where:

$$\gamma_{a,w} = \frac{\sigma_u^2}{\sigma_u^2 + \frac{\sum_h w_{ah}^2}{\sum_h w_{ah}} (1_{n_a}^T R_{a,w}^{-1} 1_{n_a})^{-1}} \quad (11)$$

$$= \frac{\sigma_u^2}{\sigma_u^2 + \sum_h w_{ah}^2 \left(\sum_h w_{ah} \sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right)^{-1}}. \quad (12)$$

Given this expression for V_w^{-1} , let us work out what this means for $X^T V_w^{-1} X$ and $X^T V_w^{-1} y$ separately, and then put these back together to obtain the alternative representation for $\hat{\beta}_w$. We begin with $X^T V_w^{-1} X$.

$$\begin{aligned} X^T V_w^{-1} X &= \sum_a X_a^T V_{a,w}^{-1} X_a \\ &= \sum_a X_a^T \left(R_{a,w}^{-1} - \left(\frac{\gamma_{a,w}}{1_{n_a}^T R_{a,w}^{-1} 1_{n_a}} \right) R_{a,w}^{-1} 1_{n_a} 1_{n_a}^T R_{a,w}^{-1} \right) X_a \\ &= \sum_a X_a^T R_{a,w}^{-1} X_a - \gamma_{a,w} (1_{n_a}^T R_{a,w}^{-1} 1_{n_a}) X_a^T \left(\frac{R_{a,w}^{-1}}{1_{n_a}^T R_{a,w}^{-1} 1_{n_a}} \right) 1_{n_a} 1_{n_a}^T \left(\frac{R_{a,w}^{-1}}{1_{n_a}^T R_{a,w}^{-1} 1_{n_a}} \right) X_a \\ &= \sum_a \left(\sum_h \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) x_{ah} x_{ah}^T - \gamma_{a,w} \left(\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) \bar{x}_{a,w} \bar{x}_{a,w}^T \right), \end{aligned}$$

with:

$$\bar{x}_{a,w} = \left(\frac{1}{\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2}} \right) \sum_h \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) x_{ah}. \quad (13)$$

By similar logic we obtain the following expression for $X^T V_w^{-1} y$:

$$X^T V_w^{-1} y = \sum_a \left(\sum_h \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) x_{ah} y_{ah} - \gamma_{a,w} \left(\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) \bar{x}_{a,w} \bar{y}_{a,w} \right), \quad (14)$$

with:

$$\bar{y}_{a,w} = \left(\frac{1}{\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2}} \right) \sum_h \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) y_{ah}. \quad (15)$$

Combining the expressions we obtained for $X^T V_w^{-1} X$ and $X^T V_w^{-1} y$ yields the following expression for $\hat{\beta}_w$:

$$\begin{aligned} \hat{\beta}_w &= \left(\sum_a \sum_h \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) x_{ah} x_{ah}^T - \gamma_{a,w} \left(\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) \bar{x}_{a,w} \bar{x}_{a,w}^T \right)^{-1} \\ &\times \left(\sum_a \sum_h \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) x_{ah} y_{ah} - \gamma_{a,w} \left(\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) \bar{x}_{a,w} \bar{y}_{a,w} \right). \end{aligned}$$

If we assume constant variance $\sigma_{\varepsilon,ah}^2 = \sigma_\varepsilon^2$, we have that σ_ε^2 drops from the equation altogether, in which case our expression for $\hat{\beta}_w$ is seen to coincide with the expression obtained by You and Rao (2002) under the same assumptions.

Let us next try to re-write the expression for the variance of $\hat{\beta}_w$ in a way that will make it easier to compute. Due to the block-diagonal nature of both V_w and V , it follows that $\text{var}[\hat{\beta}_w]$ can be written as:

$$\begin{aligned}\text{var}[\hat{\beta}_w] &= (X^T V_w^{-1} X)^{-1} (X^T V_w^{-1} V V_w^{-1} X) (X^T V_w^{-1} X)^{-1} \\ &= \left(\sum_a X_a^T V_{a,w}^{-1} X_a \right)^{-1} \left(\sum_a X_a^T V_{a,w}^{-1} V_a V_{a,w}^{-1} X_a \right) \left(\sum_a X_a^T V_{a,w}^{-1} X_a \right)^{-1},\end{aligned}$$

where for ease of notation we have dropped the “hat” from the right-hand-side (RHS). Note that we have already expanded $X_a^T V_{a,w}^{-1} X_a$ when we revisited the expression for $\hat{\beta}_w$, which leaves only $X_a^T V_{a,w}^{-1} V_a V_{a,w}^{-1} X_a$. Let us first examine the matrix $V_{a,w}^{-1} V_a V_{a,w}^{-1}$. Writing out the matrix multiplication yields:

$$\begin{aligned}V_{a,w}^{-1} V_a V_{a,w}^{-1} &= R_{a,w}^{-1} R_a R_{a,w}^{-1} + \sigma_u^2 R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} - \left(\frac{\gamma_{a,w}}{\mathbf{1}_{n_a}^T R_{a,w}^{-1} \mathbf{1}_{n_a}} \right) R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} R_a R_{a,w}^{-1} \\ &\quad - \sigma_u^2 \gamma_{a,w} R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} - \left(\frac{\gamma_{a,w}}{\mathbf{1}_{n_a}^T R_{a,w}^{-1} \mathbf{1}_{n_a}} \right) R_{a,w}^{-1} R_a R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} \\ &\quad - \sigma_u^2 \left(\frac{\gamma_{a,w}}{\mathbf{1}_{n_a}^T R_{a,w}^{-1} \mathbf{1}_{n_a}} \right) R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} \\ &\quad + \left(\frac{\gamma_{a,w}}{\mathbf{1}_{n_a}^T R_{a,w}^{-1} \mathbf{1}_{n_a}} \right)^2 R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} R_a R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} \\ &\quad + \sigma_u^2 \gamma_{a,w} \left(\frac{\gamma_{a,w}}{\mathbf{1}_{n_a}^T R_{a,w}^{-1} \mathbf{1}_{n_a}} \right) R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1}.\end{aligned}$$

After rearranging terms we obtain:

$$V_{a,w}^{-1} V_a V_{a,w}^{-1} = \sigma_u^2 (1 - \gamma_{a,w})^2 R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1} + B_a R_a B_a^T,$$

where:

$$B_a = R_{a,w}^{-1} - \left(\frac{\gamma_{a,w}}{\mathbf{1}_{n_a}^T R_{a,w}^{-1} \mathbf{1}_{n_a}} \right) R_{a,w}^{-1} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^T R_{a,w}^{-1}. \quad (16)$$

Inserting this into $X_a^T V_{a,w}^{-1} V_a V_{a,w}^{-1} X_a$ yields:

$$\begin{aligned}
X_a^T V_{a,w}^{-1} V_a V_{a,w}^{-1} X_a &= \sigma_u^2 (1 - \gamma_{a,w})^2 (X_a^T R_{a,w}^{-1} 1_{n_a}) (1_{n_a}^T R_{a,w}^{-1} X_a) + X_a^T B_a R_a B_a^T X_a \\
&= \sigma_u^2 (1 - \gamma_{a,w})^2 \left(\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right)^2 \bar{x}_{a,w} \bar{x}_{a,w}^T \\
&+ \left(\sum_h \frac{w_{ah}^2}{\sigma_{\varepsilon,ah}^2} \right) \left[\sum_h \check{w}_{a,h} x_{ah} x_{ah}^T - \gamma_{a,w} \bar{x}_{a,w} \check{x}_{a,w}^T - \gamma_{a,w} \check{x}_{a,w} \bar{x}_{a,w}^T + \gamma_{a,w}^2 \bar{x}_{a,w} \bar{x}_{a,w}^T \right],
\end{aligned}$$

where:

$$\check{x}_{a,w} = \sum_h \check{w}_{a,h} x_{a,h},$$

with:

$$\check{w}_{a,h} = \frac{\frac{w_{ah}^2}{\sigma_{\varepsilon,ah}^2}}{\sum_h \frac{w_{ah}^2}{\sigma_{\varepsilon,ah}^2}}.$$

Putting the terms together gives us the following elaborate expression for the variance of $\hat{\beta}_w$:

$$\begin{aligned}
\text{var}[\hat{\beta}_w] &= \left(\sum_a X_a^T V_{a,w}^{-1} X_a \right)^{-1} \left(\sum_a X_a^T V_{a,w}^{-1} V_a V_{a,w}^{-1} X_a \right) \left(\sum_a X_a^T V_{a,w}^{-1} X_a \right)^{-1} \\
&= C \left(\sum_a \sigma_u^2 (1 - \gamma_{a,w})^2 \left(\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right)^2 \bar{x}_{a,w} \bar{x}_{a,w}^T \right) C^T \\
&+ C \left(\sum_a \left(\sum_h \frac{w_{ah}^2}{\sigma_{\varepsilon,ah}^2} \right) \left[\sum_h \check{w}_{a,h} x_{ah} x_{ah}^T - \gamma_{a,w} \bar{x}_{a,w} \check{x}_{a,w}^T - \gamma_{a,w} \check{x}_{a,w} \bar{x}_{a,w}^T + \gamma_{a,w}^2 \bar{x}_{a,w} \bar{x}_{a,w}^T \right] \right) C^T.
\end{aligned}$$

where:

$$C = \left(\sum_a \sum_h \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) x_{ah} x_{ah}^T - \gamma_{a,w} \left(\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) \bar{x}_{a,w} \bar{x}_{a,w}^T \right)^{-1}.$$

3.2 Probability weighted OLS nested as a special case

The weighted OLS estimator is effectively obtained by setting $V_w = \sigma^2 W$ and $V = \sigma^2 I_n$, where σ^2 denotes the variance of the total error term. This yields the following expression for $\tilde{\beta}_w$:

$$\begin{aligned}
\tilde{\beta}_w &= (X^T W X)^{-1} X^T W y \\
&= \left(\sum_a \sum_h w_{ah} x_{ah} x_{ah}^T \right)^{-1} \left(\sum_a \sum_h w_{ah} x_{ah} y_{ah}^T \right).
\end{aligned}$$

The corresponding variance solves:

$$\begin{aligned}\text{var}[\tilde{\beta}_w] &= \sigma^2 (X^T W X)^{-1} (X^T W^2 X) (X^T W X)^{-1} \\ &= \sigma^2 \left(\sum_a \sum_h w_{ah} x_{ah} x_{ah}^T \right)^{-1} \left(\sum_a \sum_h w_{ah}^2 x_{ah} x_{ah}^T \right) \left(\sum_a \sum_h w_{ah} x_{ah} x_{ah}^T \right)^{-1}.\end{aligned}$$

Note that in Stata this estimator can be obtained by including the sampling weights as “probability weights” in the regular regression function (without using robust standard errors).

4 Empirical Bayes prediction assuming normality

Here we are interested in identifying the distribution of the area random error u_a conditional on the residuals e_a for the households sampled from area a .⁴ This task is greatly simplified by assuming that both u_a and ε_{ah} are normally distributed, as is done by Huang and Hidirolou (2003) and You and Rao (2002). It then follows that the distribution of u_a conditional on e_a too will be normal. What remains is to identify the mean and variance of this distribution.

Huang and Hidirolou (2003) offer an estimate of the conditional mean $E[u_a|e_a]$ for the general linear mixed model. Applying their results to our nested error regression model with potentially non-constant variances $\sigma_{\varepsilon,ah}^2$, we obtain the following:

$$E[u_a|e_a] = \hat{u}_a = \left(\frac{\sum_h w_{ah}}{\sum_h w_{ah}^2} \right) \sigma_u^2 1_{n_a}^T V_{a,w}^{-1} e_a, \quad (17)$$

where $e_a = (e_{a1}, \dots, e_{an_a})^T$ denotes the vector of area a residuals coming out of the (weighted) GLS regression. Note that we dropped the “hat” from the RHS to ease notation. Substituting the expression we derived for $V_{a,w}^{-1}$ (see eq. (10)) into eq. (17) yields:

$$\begin{aligned}\hat{u}_a &= \left(\frac{\sum_h w_{ah}}{\sum_h w_{ah}^2} \right) \sigma_u^2 1_{n_a}^T \left[R_{a,w}^{-1} - \left(\frac{\gamma_{a,w}}{1_{n_a}^T R_{a,w}^{-1} 1_{n_a}} \right) R_{a,w}^{-1} 1_{n_a} 1_{n_a}^T R_{a,w}^{-1} \right] e_a \\ &= \gamma_{a,w} \left(\frac{1_{n_a}^T R_{a,w}^{-1} e_a}{1_{n_a}^T R_{a,w}^{-1} 1_{n_a}} \right) \\ &= \gamma_{a,w} \left(\frac{\sum_h \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2} \right) e_{ah}}{\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2}} \right).\end{aligned}$$

⁴For ease of exposition we will treat the residuals e_a as if they were observed data, i.e. as if β was known. In practice of course we will be working with estimates of e_a .

Huang and Hidirolou (2003) unfortunately do not offer an estimate of the variance of u_a conditional on $\hat{u}_a$, which is what we would need to implement Empirical Best estimation. One way to compute $\text{var}[u_a|\hat{u}_a]$ is to appeal to the law of total variance:

$$\begin{aligned}\text{var}[u_a] &= E[\text{var}[u_a|\hat{u}_a]] + \text{var}[E[u_a|\hat{u}_a]] \\ &= E[\text{var}[u_a|\hat{u}_a]] + \text{var}[\hat{u}_a].\end{aligned}$$

To compute $\text{var}[\hat{u}_a]$ it will be convenient to define $\alpha_{ah} = \left(\frac{w_{ah}}{\sigma_{\varepsilon,ah}^2}\right) / (\sum_h \frac{w_{ah}}{\sigma_{\varepsilon,ah}^2})$:

$$\begin{aligned}\text{var}[\hat{u}_a] &= \text{var}[\gamma_{a,w} \sum_h \alpha_{ah} e_{ah}] \\ &= \gamma_{a,w}^2 \text{var}[u_a + \sum_h \alpha_{ah} \varepsilon_{ah}] \\ &= \gamma_{a,w}^2 \left(\sigma_u^2 + \sum_h \alpha_{ah}^2 \sigma_{\varepsilon,ah}^2 \right).\end{aligned}$$

Inserting this into eq. (18) gives us:

$$E[\text{var}[u_a|\hat{u}_a]] = \sigma_u^2 - \gamma_{a,w}^2 \left(\sigma_u^2 + \sum_h \alpha_{ah}^2 \sigma_{\varepsilon,ah}^2 \right). \quad (18)$$

It can be verified that under the assumption of constant variance $\sigma_{\varepsilon,ah}^2 = \sigma_\varepsilon^2$, we have that $\sigma_u^2 + \sum_h \alpha_{ah}^2 \sigma_{\varepsilon,ah}^2$ simplifies to $\sigma_u^2 + \sigma_\varepsilon^2 \sum_h \alpha_{ah}^2 = \frac{\sigma_u^2}{\gamma_{a,w}}$. In this case the conditional variance is seen to take the form: $\text{var}[u_a|\hat{u}_a] = (1 - \gamma_{a,w})\sigma_u^2$, which coincides with the expression derived by You and Rao (2002) under the same assumptions. Interestingly, $\text{var}[u_a|\hat{u}_a]$ will be of the same form when we allow $\sigma_{\varepsilon,ah}^2$ to vary but assume the sampling weights to be constant. This representation of the conditional variance does not apply to the more general case however where both $\sigma_{\varepsilon,ah}^2$ and the sampling weights will vary across households.

5 Annex: Estimation of σ_u^2 and σ_ε^2

A variation of Henderson's method III estimator for the variance parameters that permits the use of sampling weights can be found in Huang and Hidirolou (2003). Let us define (borrowing notation from Huang and Hidirolou, 2003):

$$\begin{aligned}
SSE &= \sum_{ah} w_{ah} (y_{ah} - \bar{y}_{a,w})^2 - \sum_{ah} w_{ah} (y_{ah} - \bar{y}_{a,w}) (x_{ah} - \bar{x}_{a,w})^T \times \\
&\quad \times \left(\sum_{ah} w_{ah} (x_{ah} - \bar{x}_{a,w}) (x_{ah} - \bar{x}_{a,w})^T \right)^{-1} \sum_{ah} w_{ah} (y_{ah} - \bar{y}_{a,w}) (x_{ah} - \bar{x}_{a,w}) \\
t_2 &= tr \left[\left(\sum_{ah} w_{ah} (x_{ah} - \bar{x}_{a,w}) (x_{ah} - \bar{x}_{a,w})^T \right)^{-1} \sum_{ah} w_{ah}^2 (x_{ah} - \bar{x}_{a,w}) (x_{ah} - \bar{x}_{a,w})^T \right] \\
t_3 &= tr \left[\left(\sum_{ah} w_{ah} x_{ah} x_{ah}^T \right)^{-1} \sum_{ah} w_{ah}^2 x_{ah} x_{ah}^T \right] \\
t_4 &= tr \left[\left(\sum_{ah} w_{ah} x_{ah} x_{ah}^T \right)^{-1} \sum_a \left(\sum_h w_{ah} \right)^2 \bar{x}_{a,w} \bar{x}_{a,w}^T \right],
\end{aligned}$$

where $\bar{y}_{a,w} = \sum_h w_{ah} y_{ah} / (\sum_h w_{ah})$ and $\bar{x}_{a,w} = \sum_h w_{ah} x_{ah} / (\sum_h w_{ah})$ denote the weighted mean of y_{ah} and x_{ah} , respectively. (Note however that these weighted mean variables are different from those defined in the main text for they use different weights.)

The estimators for the unconditional variances σ_u^2 and σ_ε^2 can then be obtained as:

$$\begin{aligned}
\hat{\sigma}_{\varepsilon,w}^2 &= \frac{SSE}{\sum_{ah} w_{ah} - \sum_a \left(\frac{\sum_h w_{ah}^2}{\sum_h w_{ah}} \right) - t_2} \\
\hat{\sigma}_{u,w}^2 &= \frac{\sum_{ah} w_{ah} y_{ah}^2 - (\sum_{ah} w_{ah} y_{ah} x_{ah}^T) (\sum_{ah} w_{ah} x_{ah} x_{ah}^T)^{-1} (\sum_{ah} w_{ah} y_{ah} x_{ah}) - (\sum_{ah} w_{ah} - t_3) \hat{\sigma}_{\varepsilon,w}^2}{\sum_{ah} w_{ah} - t_4}.
\end{aligned}$$

You and Rao (2002) find that the use of sampling weights makes little difference for the estimation of the variance parameters (they opt for leaving out the sampling weights for this purpose).

References

- Elbers, C., Fujii, T., Lanjouw, P., Ozler, B. and Yin, W. (2007). Poverty alleviation through geographic targeting: How much does disaggregation help? *Journal of Development Economics*, **83**, 198–213.
- Elbers, C., Lanjouw, J. and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, **71**, 355–364.
- Elbers, C. and van der Weide, R. (2014). Estimation of normal mixtures in a nested error model with an application to small area estimation of poverty and inequality. *Policy Research World Bank Working Paper*, no. 6962.
- Huang, R. and Hidirolou, M. (2003). Design consistent estimators for a mixed linear model on survey data. *Joint Statistical Meetings, Section on Survey Research Methods* 1897–1904.
- Molina, I. and Rao, J. (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics*, **38**, 369–385.
- Rao, J. (2003). *Small area estimation*. London: Wiley.
- You, Y. and Rao, J. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Canadian Journal of Statistics*, **30**, 431–439.