Shrinkage estimates for multi-level heteroscedastic hierarchical normal linear models

S.K. Ghoreishi
Department of Statistics, Faculty of Sciences, University of Qom
Qom, I. R. of Iran
atty_ghoreishi@yahoo.com

A. Mostafavinia
PhD student of Anatomy, Medical Faculty, Shahid Beheshti University
Tehran, I. R. of Iran
a.mostafavinia@gmail.com

Received 20 November 2014
Accepted 5 March 2015

Empirical Bayes approach is an attractive method for estimating hyperparameters in hierarchical models. But, under the assumption of normality for a multi-level heteroscedastic hierarchical model, which involves several explanatory variables, the analyst may often wonder whether the shrinkage estimators have efficient asymptotic properties in spite of the fact they involve numerous hyperparameters. In this work, we propose a methodology for estimating the hyperparameters whenever one deals with multi-level heteroscedastic hierarchical normal model with several explanatory variables. we investigate the asymptotic properties of the shrinkage estimators when the shrinkage location hyperparameter lies within a suitable interval based on the sample range of the data. Moreover, we show our methodology performs much better in real data sets compared to available approaches.

Keywords: Asymptotic optimality; Heteroscedasticity; Multiple linear regression; Shrinkage estimators; Stein's unbiased risk estimate(SURE).

2000 Mathematics Subject Classification: 62F15, 62F30

1. Introduction

Nowadays, hierarchical modeling has found vast applications in many disciplines such as biology, ecology, medicine and engineering. An extensive progress of hierarchical modeling has been the subject of attention over several decades. Undoubtedly Stein [1962] was the pioneer in developing such an important field of statistics. His initial work on shrinkage estimations of several normal means and later on their empirical Bayes interpretation founded an effective way for developing hierarchical models, James and Stein (1961). Empirical Bayes is an approach wherein known relationships among the coordinates of the parameters allow use of the data to estimate some features of the prior distribution. It is well known that the empirical Bayes methods can be categorized into parametric and nonparametric, Morris(1983). The parametric empirical Bayes interpretation of estimators as well as nonparametric ones have motivated various treatments of this problem.

Published by Atlantis Press

Although it seems that the parametric empirical Bayes estimators in comparison to the nonparametric empirical Bayes estimators are frequently used to analyze real data sets, they usually involve some hyperparameters which originate from prior distributions. Hence in practice, our biggest challenge is eliciting these hyperparameters.

There is a large body of theoretical literature on the empirical Bayes analysis, especially regarding asymptotic optimality of the empirical Bayes procedures; for more details see Berger (1985) and references therein. For dealing with both homoscedastic (equal subpopulation variances) and heteroscedastic (unequal subpopulation variances) hierarchical models, the available empirical Bayes procedures are powerful enough to handle them. For more detail on the subject, see Berger and Strawderman(1996) and Brown and Greenshtein (2009).

It should be emphasized that in exploring empirical Bayes methods, incorporation of a suitable loss function and investigation of its corresponding risk properties are of main interest. Thus, exploring shrinkage estimators, admissible minimax estimators, proper Bayes minimax estimators and their comparisons under different loss functions has attracted the attention of many authors.

Two popular methods of performing empirical Bayes analysis involve estimation of hyperparameters by maximum likelihood (EBML) and by method of moment (EBMM); for more details on evaluation and performance of these estimators, through a simulation study, see Brown(2008). Xie et al. (2012) proposed a class of shrinkage estimators that can be readily applied in the heteroscedastic hierarchical normal models. Their motivation in that work was to know whether it is possible to formally compare these different shrinkage estimators and identify the ’ optimal’ one. They called their shrinkage estimators SURE (Stein’s unbiased risk estimate). They also established the asymptotic optimality property of the SURE estimators. But modeling the parameters via alternate structural assumptions on the prior is a notable problem, too. For instance, one could let the prior variance differ according to some model or be stochastically dependent in some fashion. This subject may be common in general(generalized) linear model, especially whenever a hierarchical setting is assumed. However, Ghoreishi and Meshkani (2014) tried to solve the problem considering a class of weighted shrinkage estimators, called Mean General SURE(MGS)-estimators in the context of hierarchical models, assuming heteroscedasticity for both levels of a two-level normal hierarchical model. These estimates were obtained based on Stein’s unbiased estimate of risk. Moreover, the asymptotic properties of MGS-estimators were investigated.

In our pervious work, Ghoreishi and Mehskani (2014), we tried to extend the assumption of constant variance for the second-level hierarchial model, considered by Xie et al. (2012), to the case where the variance of second-level model has a tendency to change through sub-populations. In that paper, we showed that negligence in considering this fundamental assumption could lead to substantial bias in the estimates of the parameters. To deal with our idea, we developed our methodology, assuming the following weighted simple regression hierarchical model:

$$\begin{aligned} A) : X_t|\theta_t &\sim N(\theta_t, k(z_t)),\\ \theta_t &= \beta_0 + \beta_1 z_t,\\ \beta_0 &\sim N(\mu, \lambda),\\ \beta_1 &\sim N(0, \lambda), \end{aligned}$$

where z_t is an explanatory variable and μ and $\lambda(\geq 0)$ are hyperparameters and finally $k : \mathscr{U} \subseteq \mathbb{R} \to \mathbb{R}^+$ is either completely known or will be known employing some plug-in robust estimators. It is easy to drive its equivalent two-level marginal model as:

$$X_t|\theta_t \sim N(\theta_t, k(z_t)),$$
$$\theta_t \sim N(\mu, \lambda(1+z_t^2)).$$

Our focus in this paper is to generalize model (A) to the model:

$$B): X_t|\theta_t \sim N(\theta_t, k(z_t)),$$
$$\theta_t = \beta_0 + \beta_1 z_t,$$
$$\beta_0 \sim N(\mu, \lambda_0 k_0(z_{0t})),$$
$$\beta_1 \sim N(0, \lambda_1 k_1(z_{1t})),$$

where k_0 and k_1 are some known positive functions and z_{0t} and z_{1t} are some explanatory variables which may only influence the variances of the simple regression coefficients β_0 and β_1. Assuming equal variances for all regression coefficients, i.e., model (A), is a special case of model (B). This generalized form is expected to perform better than model (A), in practice, since it is more flexible to fit to variable data sets.

By the above assumptions, the equivalent two-level marginal model (B) is

$$X_t|\theta_t \sim N(\theta_t, k(z_t)),$$
$$\theta_t \sim N(\mu, \lambda_0(k_0(z_{0t}) + \frac{\lambda_1}{\lambda_0} z_t^2 k_1(z_{1t}))).$$

Assuming $\lambda = \lambda_0 (> 0)$, $\delta = \frac{\lambda_1}{\lambda_0} (> 0)$, and $A_t = (z_t, z_{0t}, z_{1t})$, one can redefine the above model as

$$X_t|\theta_t \sim N(\theta_t, g(A_t)),$$
$$\theta_t \sim N(\mu, \lambda(g_0(A_t) + \delta g_1(A_t))),$$

where it is assumed that $A_t s$ are some known and possibly distinct points in $\mathscr{D} \subseteq \mathbb{R}^k$, and $g, g_i : \mathscr{D} \subseteq \mathbb{R}^k \to \mathbb{R}^+$; $i = 0, 1$. The quantities μ, λ, and δ are considered as the hyperparameters. Moreover, we define $g^*_\delta(A_t) = g_0(A_t) + \delta g_1(A_t)$. Therefore, we have

$$X_t|\theta_t \sim N(\theta_t, g(A_t)),$$
$$\theta_t \sim N(\mu, \lambda g^*_\delta(A_t)). \tag{1.1}$$

As mentioned above, model (**??**) can be reduced to that of Xie et al. (2012) if one assume $g^*_\delta(A_t)$ to be a constant function. Moreover, it will be equal to the model considered by Ghoreishi and Meshkani (2014) whenever $\delta = 1$. Thus, it applies to large domains of modeling.

From Bayes' theorem, the posterior distribution of θ_t is

$$\theta_t \sim N(\frac{\lambda g^*_\delta(A_t)}{\lambda g^*_\delta(A_t) + g(A_t)} X_t + \frac{g(A_t)}{\lambda g^*_\delta(A_t) + g(A_t)} \mu, \frac{\lambda g^*_\delta(A_t) g(A_t)}{\lambda g^*_\delta(A_t) + g(A_t)}). \tag{1.2}$$

It is easy to see that the marginal distribution of X_t is

$$X_t \sim N(\mu, \lambda g^*_\delta(A_t) + g(A_t)). \tag{1.3}$$

Our main purposes in this setting are:

i) Deriving some suitable estimates for the unknown quantities μ, λ, and δ and plugging-in these estimates to obtain the corresponding shrinkage MGS-estimators.
ii) Investigating the asymptotic properties of MGS-estimators.
iii) Extending model (**??**) from a simple linear model framework to a multiple one with r regressors, while assuming the shrinkage location hyperparameter μ lies in the interval $[-\max_t |X_t|, \max_t |X_t|]$. Practically, this assumption is not a restriction since no sensible shrinkage estimator would attempt to shrink toward a location that lies outside the sample range of the data.

The structure of this paper is as follows. Section 2 presents some preliminary results including some basic concepts and necessary notations. Section 3 contains the main results for establishing the asymptotic properties for MGS(GS)-estimators. The theoretical results are illustrated on a real data set in Section 4.

2. Preliminaries

Consider the functions $g, g_0, g_1 : \mathscr{D} \subseteq \mathbb{R}^k \to \mathbb{R}^+$. Let $X_1, X_2, \cdots, X_n$ be a collection of independent normal variables with means $\theta_1, \theta_2, \cdots, \theta_n$ and variances $g(A_1), g(A_2), \cdots, g(A_n)$, respectively. Here, as in elsewhere, we assume $A_t s$ are some known and possibly distinct points in $\mathscr{D} \subseteq \mathbb{R}^k$. That is,

$$X_t|\theta_t \sim N(\theta_t, g(A_t)), \quad t = 1, 2, \cdots, n.$$

Moreover, for given g_0 and g_1, assume θ_ts are independent and normally distributed with mean μ and variance $\lambda g^*_\delta(A_t)$. That is,

$$\theta_t \sim N(\mu, \lambda g^*_\delta(A_t)),$$

where $g^*_\delta(A_t) = g_0(A_t) + \delta g_1(A_t)$. In this setting, the quantities μ, λ, and δ are unknown hyperparameters and therefore, they need to be estimated. Below, three approaches are adopted for estimating these quantities.

2.1. *Empirical Bayes moment method*

The EBMM estimators are obtained as the solutions of the following equations:

$$\begin{aligned}
\mu &= \frac{\sum_t \sqrt{q_t \frac{h_\delta(A_t)}{\lambda + h_\delta(A_t)}} X_t}{\sum_t \sqrt{q_t \frac{h_\delta(A_t)}{\lambda + h_\delta(A_t)}}}, \\
\lambda &= \frac{1}{n} [\sum_t h_\delta(A_t) \{q_t (X_t - \mu)^2 - 1\}]^+, \\
\delta &= \frac{[\sum_t \{2(X_t - \mu)^2 - \lambda g_0(A_t) - g(A_t)\}]^+}{\lambda \sum g_1(A_t)}.
\end{aligned}$$

where $h_\delta(A_t) = \frac{g(A_t)}{g^*_\delta(A_t)}$ and $q_t = \frac{1}{g(A_t)}$. Whenever $\mu = 0$, the EBMM estimator of λ is given by

$$\lambda = \frac{1}{n}\mathbf{[}\sum_t h_\delta(A_t)\{q_t X_t^2 - 1\}\mathbf{]}^+,$$
$$\delta = \frac{\mathbf{[}\sum_t\{2X_t^2 - \lambda g_0(A_t) - g(A_t)\}\mathbf{]}^+}{\lambda \sum g_1(A_t)}.$$

2.2. *Empirical Bayes maximum likelihood method*

The EBML estimators are obtained by maximizing the marginal density of $X_t s$ with respect to μ, λ, and δ. They satisfy the following equations whenever the roots exist,

$$\sum q_t \frac{h_\delta(A_t)(X_t - \mu)}{\lambda + h_\delta(A_t)} = 0,$$
$$\sum\{\frac{1}{\lambda + h_\delta(A_t)} - q_t \frac{h_\delta(A_t)(X_t - \mu)^2}{(\lambda + h_\delta(A_t))^2}\} = 0,$$
$$\sum \frac{g_1(A_t)h_\delta(A_t)}{g(A_t)}\{\frac{1}{\lambda + h_\delta(A_t)} - q_t \frac{h_\delta(A_t)(X_t - \mu)^2}{(\lambda + h_\delta(A_t))^2}\} = 0.$$

Whenever $\mu = 0$, these equations can be written in a suitable form

2.3. *Stein's unbiased risk estimate(SURE) method*

An alternative approach to empirical Bayes method is something related to Stein's unbiased risk estimate (SURE) which is based on the weighted mean of squared error-loss, Ghoreishi and Meshkani (2014),

$$l_q(\theta_t, \hat{\theta}_t) = \frac{1}{\sum q_t}\sum q_t(\hat{\theta}_t - \theta_t)^2, \tag{2.1}$$

where $q_t = \frac{1}{g(A_t)}$. Assuming $q_t = 1$ it corresponds to the case considered by Xie et al. (2012). Ghoreishi and Meshkani (2014) applied it to the shrinkage estimator

$$\hat{\theta}_t^{\lambda,\mu,\delta} = \frac{\lambda}{\lambda + h_\delta(A_t)}X_t + \frac{h_\delta(A_t)}{\lambda + h_\delta(A_t)}\mu, \tag{2.2}$$

to estimate θ_t, where $\delta = 1$. However, here we assume δ as an unknown quantity. From SURE approach perspective, if one is interested in using the shrinkage estimator (2.2) as an estimator for θ_t with fixed λ, μ, and δ, he/she can first estimate λ, μ, and δ by minimizing the unbiased estimator of $E[l_q(\theta, \hat{\theta})]$. In this case, the corresponding SURE estimate for θ_t is given by

$$\hat{\theta}_t^{SURE} = \frac{\hat{\lambda}^{SURE}}{\hat{\lambda}^{SURE} + h_{\hat{\delta}^{SURE}}(A_t)}X_t + \frac{h_{\hat{\delta}^{SURE}}(A_t)}{\hat{\lambda}^{SURE} + h_{\hat{\delta}^{SURE}}(A_t)}\hat{\mu}^{SURE}. \tag{2.3}$$

It is natural to expect that adding one more parameter δ we would have smaller $E[l_q(\theta, \hat{\theta})]$ in comparison to our pervious work, Ghoreishi and Meshkani (2014).

In order to compare the performance of the estimator (2.3), one may use the oracle loss (OL) estimator

$$\hat{\theta}_t^{OL} = \frac{\tilde{\lambda}^{OL}}{\tilde{\lambda}^{OL} + h_{\tilde{\delta}^{OL}}(A_t)} X_t + \frac{h_{\tilde{\delta}^{OL}}(A_t)}{\tilde{\lambda}^{OL} + h_{\tilde{\delta}^{OL}}(A_t)} \tilde{\mu}^{OL}, \tag{2.4}$$

where $\tilde{\lambda}^{OL}$, $\tilde{\mu}^{OL}$, and $\tilde{\delta}^{OL}$ are obtained by minimizing

$$\frac{1}{n}\sum\left(\frac{\lambda}{\lambda + h_\delta(A_t)} X_t + \frac{h_\delta(A_t)}{\lambda + h_\delta(A_t)} \mu - \theta_t\right)^2, \tag{2.5}$$

with respect to μ, λ, and δ. The notation $\tilde{\theta}_t^{OL}$ rather than $\hat{\theta}_t^{OL}$ is used to emphasize that $\tilde{\theta}_t^{OL}$ depends on unknown θ_t and hence is not really an estimator. However, since it has smaller loss or risk within the class of estimators of the form

$$\frac{\lambda}{\lambda + h_\delta(A_t)} X_t + \frac{h_\delta(A_t)}{\lambda + h_\delta(A_t)} \mu,$$

it is suitable for evaluating the performance of the SURE estimator.

Let us consider the general Bayes shrinkage estimator (2.2). Under the weighted sum of squared-error loss (2.1), if one uses the shrinkage estimator (2.2) to estimate θ with fixed μ, λ, and δ then an unbiased estimate for its risk,

$$R(\theta, \hat{\theta}^{\lambda,\mu,\delta}) = E[l_q(\theta, \hat{\theta}^{\lambda,\mu,\delta})] = \frac{1}{\sum q_t} \sum \frac{1}{(\lambda + h_\delta(A_t))^2} \{q_t(\theta_t - \mu)^2 + \lambda^2\},$$

would be

$$\begin{aligned} MGS(\lambda,\mu,\delta) = \frac{1}{\sum q_t} \sum \{\frac{g(A_t)(X_t-\mu)^2}{(\lambda g_\delta^*(A_t) + g(A_t))^2} + \frac{\lambda g_\delta^*(A_t) - g(A_t)}{\lambda g_\delta^*(A_t) + g(A_t)}\} = \\ \frac{1}{\sum q_t} \sum \{q_t \frac{h_\delta^2(A_t)(X_t-\mu)^2}{(\lambda + h_\delta(A_t))^2} + \frac{\lambda - h_\delta(A_t)}{\lambda + h_\delta(A_t)}\}. \end{aligned}$$

One can obtain the estimates $\hat{\mu}^{SURE}$, $\hat{\lambda}^{SURE}(\geq 0)$ and $\hat{\delta}^{SURE}(\geq 0)$ as the minimizers of $MGS(\lambda,\mu,\delta)$. They satisfy the following equations whenever the solutions exist,

$$\begin{aligned} &\sum q_t \frac{h_\delta^2(A_t)(X_t-\mu)}{(\lambda + h_\delta(A_t))^2} = 0, \\ &\sum \{\frac{2h_\delta(A_t)}{(\lambda + h_\delta(A_t))^2} - q_t \frac{h_\delta^2(A_t)(X_t-\mu)^2}{(\lambda + h_\delta(A_t))^3}\} = 0, \\ &\sum \frac{g_1(A_t)(X_t-\mu)^2}{g_\delta^*(A_t)(\lambda + h_\delta(A_t))^3} = 0. \end{aligned}$$

Here, it is important to note that one may be interested in extending model (**??**) from a simple linear model framework to a multiple one with r regressors, i.e.,

$$\begin{aligned} X_t|\theta_t &\sim N(\theta_t, g(A_t)), \\ \theta_t &\sim N(\mu, \lambda g_\delta^*(A_t)), \end{aligned} \tag{2.6}$$

where $g_\delta^*(A_t) = g_0(A_t) + \delta_1 g_1(A_t) + \delta_2 g_2(A_t) + \cdots, +\delta_r g_r(A_t)$, $h_\delta(A_t) = \frac{g(A_t)}{g_\delta^*(A_t)}$, and $\delta = (\delta_1, \delta_2, \cdots, \delta_r)$. In this case, we assert that all estimates EBMM, EBML and SURE of λ, μ, and

$\delta_j;\ j = 1,\cdots,r$ are elicitable in the same way which is discussed in this section. The SURE estimates are the solutions of the following equations:

$$\begin{aligned}&\sum q_t \frac{h_\delta^2(A_t)(X_t-\mu)}{(\lambda+h_\delta(A_t))^2} = 0,\\ &\sum\{\frac{2h_\delta(A_t)}{(\lambda+h_\delta(A_t))^2} - q_t\frac{h_\delta^2(A_t)(X_t-\mu)^2}{(\lambda+h_\delta(A_t))^3}\} = 0,\\ &\sum \frac{g_j(A_t)(X_t-\mu)^2}{g_\delta^*(A_t)(\lambda+h_\delta(A_t))^3} = 0;\ j = 1,2,\cdots,r.\end{aligned}$$

The corresponding SURE estimate for θ_t is given by

$$\hat{\theta}_t^{SURE} = \frac{\hat{\lambda}^{SURE}}{\hat{\lambda}^{SURE}+h_{\hat{\delta}^{SURE}}(A_t)}X_t + \frac{h_{\hat{\delta}^{SURE}}(A_t)}{\hat{\lambda}^{SURE}+h_{\hat{\delta}^{SURE}}(A_t)}\hat{\mu}^{SURE}. \tag{2.7}$$

3. Theoretical results

In this section, we establish the theoretical results through two following theorems which are essential for evaluating the performance of the weighted SURE estimators of the form (2.3) and (2.7) under the weighted loss function (2.1). Their proofs are easily verifiable because they are mostly in the same lines as the theorems in Ghoreishi and Meshkani (2014).
For establishing the asymptotic results, the following two conditions are necessary,

C1) $\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^n g(A_t) < \infty$.
C2) $\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^n \theta_t^{2+\eta} < \infty$ for some $\eta > 0$.

Moreover, without loss of generality, assume the sub-populations were re-indexed such that we have $0 < h_\delta(A_1) \leq h_\delta(A_2) \leq \cdots \leq h_\delta(A_n)$. Also, the relationship between the harmonic and arithmetic means leads to the following inequality

$$\frac{n}{\sum q_t} \leq \frac{1}{n}\sum g(A_t) \Leftrightarrow \frac{1}{\sum q_t} \leq \frac{1}{n^2}\sum g(A_t) \tag{3.1}$$

The following theorems reveal the asymptotic optimality of SURE estimators. These theorems show that the SURE estimators (2.7) are asymptotically as good as the general oracle loss (OL) estimator (2.4).

Theorem 3.1. *For model (**??**), under conditions (C1) and (C2) we have*

$$\sup_{\substack{\lambda\geq 0,\\ |\mu|\leq \max_t|X_t|,\\ \delta\geq 0}} \mid MGS(\lambda,\mu,\delta) - l_q(\theta,\hat{\theta}_t^{\lambda,\mu,\delta}) \mid \to 0$$

in L^1 and in probability, as $n\to\infty$.

Theorem 3.2. *For model (**??**), under conditions (C1) and (C2) we have*

$$\sup_{\substack{\lambda\geq 0,\\ |\mu|\leq \max_t|X_t|,\\ \delta_j\geq 0; j=1,\cdots,r}} \mid MGS(\lambda,\mu,\delta) - l_q(\theta,\hat{\theta}_t^{\lambda,\mu,\delta}) \mid \to 0$$

in L^1 and in probability, as $n\to\infty$.

In these two theorems, we impose a restriction on location hyperparameter μ to lie in the interval $[-\max_t |X_t|, \max_t |X_t|]$. This assumption is only for sake of easier proof. Practically, this assumption is not a restriction since no sensible shrinkage estimator would attempt to shrink toward a location that lies outside the sample range of the data.
These two theorems show that our shrinkage estimators have asymptotic optimality properties within the class of estimators (2.3) and (2.7). Thus, they can be widely used in generalized linear models which are applied in many disciplines. Here, we avoid diving more into technical details and provide justification of our results via analyzing two real data sets. However, as it is evident from the results of these examples, our SURE estimates for hyperparameters, which are based on regressor effects, perform much better in comparison to EBMM, EBML and SURE-estimates proposed by Xie et al. (2012).

4. Application

To illustrate the obtained theoretical results, we have considered two examples of simple and multiple linear regression. Our first example has been taken from our previous work, Ghoreishi and Meshkani (2014). Here, we have revisited this example to demonstrate the usefulness of the methodology presented in this work.

4.1. *Simple regression model*

Consider Bid at Auction data, Table 1 in Ghoreishi and Meshkani(2014). These data belong to a big state company which wanted to survey its recent 12 auctions. It contained an explanatory variable z:Bid at Auction(in million dollars) and response variable X:Cost of Auction(in million dollars). There, we argued that the weighted least squares provide a suitable fit, whenever the weights are proportional to $\frac{1}{z^2}$. Therefore, we considered the following weighted simple regression hierarchical model:

$$\begin{aligned} X_t|\theta_t &\sim N(\theta_t, \sigma^2 \frac{1}{z_t^2}), \\ \theta_t &= \beta_0 + \beta_1 z_t, \\ \beta_0 &\sim N(\mu, \lambda), \\ \beta_1 &\sim N(0, \lambda). \end{aligned}$$

Now, to apply our more heteroscedastic approach, we consider the following generalized weighted simple regression hierarchical model:

$$\begin{aligned} X_t|\theta_t &\sim N(\theta_t, \sigma^2 \frac{1}{z_t^2}), \\ \theta_t &= \beta_0 + \beta_1 z_t, \\ \beta_0 &\sim N(\mu, \lambda_1), \\ \beta_1 &\sim N(0, \lambda_2), \end{aligned}$$

or, equivalently,

$$\begin{aligned} X_t|\theta_t &\sim N(\theta_t, g(A_t)), \\ \theta_t &\sim N(\mu, \lambda(g_0(A_t) + \delta g_1(A_t))), \end{aligned}$$

where $g(A_t) = \frac{\sigma^2}{z_t^2}$, $g_0(A_t) = 1$, $g_1(A_t) = z_t^2$, and reanalyze the data. Moreover, we adopt the following weighted prediction errors (WPE) in terms of θ and X_t to evaluate the performance of our methodology.

$$WPE = \frac{1}{n}\sum q_t(X_t - \hat{\theta}_t)^2. \tag{4.1}$$

We consider three different models. Model M_1 with $\delta = 0$, ignoring the explanatory variable effect in heteroscedastic model. Model M_2 with $\delta = 1$, assuming equal variances for the intercept and the slope. Finally, model M_3 with letting the model have unequal variances for the intercept and the slope. In these three models, we use the weighted mean square estimate $WMSE = 0.792$ as a plug-in estimate of σ^2. Table 1 shows the results for these three settings. It is easy to see that model M_1 produces a large WPE and an unrealistic estimate for λ. In comparison, model M_2 and model M_3 give small WPEs and sensible estimates for λ. However, from practical point of view, we prefer model M_3, since it provides smaller values for both WPE and proportional variances of the intercept and the slope.

Table 1. Various estimates for the three simple regression models

Model	$\hat{\mu}^{SURE}$	$\hat{\lambda}^{SURE}$	$\hat{\delta}^{SURE}$	WPE
M_1	10.4810	44.936	0	0.236
M_2	10.8228	0.759	1	0.041
M_3	10.6709	0.253	3.569	0.009

4.2. *Multiple regression model*

Consider Systolic Blood Pressure data which is available at *http://college. cengage.com/ mathematics/ brase/ understandable-statistics/ 7e/ students/ datasets/ mlr/ frames/ frame.html*. It contains two explanatory variables z_1: Age in years and z_2 Weight in pounds. The response variable X is the systolic blood pressure of 11 patients. A multiple Classical regression analysis shows $MSE = 42.993$, which is used as a plug-in estimate of σ^2. Therefore, we consider the following multiple regression hierarchical model:

$$X_t|\theta_t \sim N(\theta_t, \sigma^2),$$
$$\theta_t \sim N(\mu, \lambda(g_0(A_t) + \delta_1 g_1(A_t) + \delta_2 g_2(A_t))),$$

where $g_0(A_t) = 1$, $g_1(A_t) = z_{1t}^2$, and $g_2(A_t) = z_{2t}^2$. Again to analyze these data we consider three different models. Model M_1 with $\delta_1 = 0$ and $\delta_2 = 0$, ignoring the explanatory variables effects in the heteroscedastic model. Model M_2 with $\delta_1 = 1$ and $\delta_2 = 1$, assuming equal variances for the intercept and two slopes. Finally, model M_3 with letting the model have unequal variances for the intercept and the slopes. Table 2 shows the results for these three settings. Clearly, model M_1 produces a large WPE and an unrealistic estimate for λ. In comparison, models M_2 and M_3 give small WPEs and sensible estimates for λ. However, from practical point of view, model M_3 is preferred, since it entails smaller WPE and also smaller proportional variances for the intercept and the slopes.

Table 2. Various estimates for the three multiple regression models

Model	$\hat{\mu}^{SURE}$	$\hat{\lambda}^{SURE}$	$\hat{\delta}_1^{SURE}$	$\hat{\delta}_2^{SURE}$	WPE
M_1	149.87	125.94	0	0	0.254
M_2	145.31	0.253	1	1	0.704×10^{-4}
M_3	145.82	0.253	0.152	0	0.448×10^{-5}

Summary

One topic which has received much attention during the last decades is how to elicit SURE estimates for hyperparameters in heteroscedastic models. Neglecting to find the true values of the hyperparameters may lead to bias estimates of the model parameters. Xie et al. (2012) assumed the constant variance for the second-level of a two-level hierarchial model. We tried to extend this assumption to the case where the variance of the second-level model has a tendency to change across subpopulations, Ghoreishi and Mehskani (2014). There, we investigated the usefulness of considering an explanatory variable in eliciting the model hyperparameters. In this work, we generalized our previous work to a multiple regression and the settings where each regressor coefficient has a different variance and plays an appreciable role in determining the SURE estimates of hyperparameters. We also discuss the asymptotic optimality of the shrinkage estimators of parameters.

Acknowledgements

We wish to thank Ghazal Ghoreishi, Professor M.R. Meshkani, the editor and two anonymous referees whose comments greatly improved the article.

References

[1] J.O. Berger, *Statistical decision theory and Bayesian analysis*, (Springer, New York, 1985).

[2] J. Berger and W.E. Strawderman, Choice of Hierarchical Priors:Admissibility in Estimation of Normal Means, *Annals of Statistics* **24** (1996) 931–951.

[3] L.D. Brown, In-Season Prediction of Batting Average: A Field Test of Empirical Bayes and Bayes Methodologies, *Annals of Applied Statistics* **2** (2008) 113–152

[4] L.D. Brown and E. Greenshtein, Nonparametric Empirical Bayes and Compound Decision Approaches to Estimation of a High-Dimensional Vector of Means, *Annals of Statistics* **37** (2009) 1685–1704.

[5] S.K. Ghoreishi and M.R. Meshkani, On SURE estimates in hierarchical models assuming heteroscedasticity for both levels of a two-level normal hierarchical model, *J. of Multivariate Analysis* **132** (2014) 129–137.

[6] W. James and C.M. Stein, Estimation With Quadratic Loss, *Proceedings of the 4th Berkeley Symposium on Probability and Statistics* **I** (1961) 367–379.

[7] K.C. Li, Asymptotic optimality of C_L and Generalized Cross Validation in Ridge Regression with Application to Spline Smoothinge, *Annals of Statistics* **14** (1986) 1101–1112.

[8] C. Morris, Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78** (1983) 47–65.

[9] C.M. Stein, Confidence Sets for the Mean of a Multivariate Normal Distribution (with discussion), *J. Roy. Statist. Soc. Ser. B,* **24** (1962) 265–296.

[10] X. Xie, S.C. Kou, and L.D. Brown, SURE Estimates for a Heteroscedastic Hierarchical Model, *J. Amer. Statist. Assoc.* **107** (2012) 1465–1479.