SHRINKAGE ESTIMATES IN A TWO-LEVEL HETEROSCEDASTIC HIERARCHICAL NORMAL MODELS WITH LINEAR TREND MEAN

S. K. GHOREISHI

Department of Statistics
Faculty of Sciences
University of Qom
Iran
e-mail: Atty_ghoreishi@yahoo.com

Abstract

In heteroscedastic hierarchical normal models, often, the response variable may rather fluctuate around a nonconstant trend than a constant value. In this paper, we assume a linear trend for the second level of a two-level heteroscedastic hierarchical normal model. The empirical Bayes maximum likelihood estimates, Stein's unbiased risk estimates, and Bayesian estimates of the hyperparameters are elicited and the corresponding shrinkage estimates are constructed. The asymptotic properties of the Stein's unbiased risk estimates are illustrated. Since our primary purpose in Bayesian analysis is sampling from the posterior distributions, with assuming some priors for the hyperparameters, the posterior full conditions are given. A real data set is analyzed to demonstrate the proposed methodology.
1. Introduction

The primary objective of this paper is to develop a methodology for eliciting the hyperparameters of a two-level heteroscedastic hierarchical normal model where a linear trend is assumed for its second level. The acceptable performance of the corresponding shrinkage estimates are the secondary objective of this paper. It is well-known that the hierarchical modelling provides an efficient approach which combines partial information to achieve accurate and stable inference about parameters of interest. In this context, risk properties of shrinkage estimators, admissible minimax estimators, and many other estimators of hierarchical models have been noted by many authors under different loss functions. The literature covers both homoscedastic and heteroscedastic models. It seems heteroscedastic hierarchical modelling, which assumes unequal subpopulation variances, is more popular in real world than homoscedastic hierarchical modelling. For more detail on the subject, see Berger and Strawderman [1] and Brown and Greenshtein [3].

In the statistical modelling, it is well known that, often, the suitable choice of a trend, for the response variable, can efficiently control the heteroscedasticity. Ignoring such a trend makes the situation worst because it involves many locally unknown sources (or hyperparameters) which cause the heteroscedasticity. This phenomenon often occurs in the regression modelling where one or more explanatory variables may have an important role in the heteroscedasticity.

A class of shrinkage estimators that can be readily applied in homoscedastic hierarchical normal models was proposed by Xie et al. [9]. Their assumed model for subpopulations \( i = 1, 2, \ldots, n \) was

\[
\begin{align*}
y_i & \sim N(\theta_i, A_i), \\
\theta_i & \sim N(\mu, \lambda),
\end{align*}
\]

where \( A_i \)'s are some known and possibly distinct real values and \( \lambda \) and \( \mu \) are hyperparameters. In this line, Ghoreishi and Meshkani [5]
proposed and developed an alternative class of shrinkage estimators that involves one or several explanatory variables in the model. Their assumed model was given by

\[ y_i \sim N(\theta_i, k(z_i)), \quad i = 1, \cdots, n, \]
\[ \theta_i \sim N(\mu, \lambda g(z_i)), \quad (2) \]

where \( z_i \) is an explanatory variable (or a vector of explanatory variables of order \( k \) in \( D \)), global quantities \( \lambda > 0 \) and \( \mu \) are hyperparameters, \( k : D \subseteq \mathbb{R}^k \rightarrow \mathbb{R}^+ \) is a known function and finally \( g : D \subseteq \mathbb{R}^k \rightarrow \mathbb{R}^+ \) is either completely known or will be known by some plug-in robust estimators. The advantage of the model (2) in comparison with the model (1) is that it employs the effectible explanatory variable \( z \) in the model.

In this work, to control the heteroscedasticity, we survey the effect of the explanatory variable \( z \) on the mean of the second level, i.e., \( \mu \), by assuming the following heteroscedastic hierarchical model:

\[ y_i \sim N(\theta_i, k(z_i)), \]
\[ \theta_i \sim N(\alpha + \beta z_i, \lambda). \quad (3) \]

In model (3), \( \theta_i \)'s are considered as the mean parameters whereas \( \alpha, \beta \), and \( \lambda \) are the hyperparameters. The functions \( k \) is defined as before in model (2). Our main purpose in this work is, first, to construct a statistical inference methodology to estimate the hyperparameters \( \alpha, \beta \), and \( \lambda \) and, then, to estimate the shrinkage estimates of \( \theta_i \)’s. In this line, three approaches; the empirical Bayes maximum likelihood, Stein’s unbiased risk approach, and finally the hierarchical Bayesian approach are illustrated to elicit the hyperparameters.

The rest of the article is organized as follows. Section 2 gives the general notations and definitions. In Section 3, we develop our theoretical results. Section 4 illustrates the performance of our methodology by analyzing of a real dataset. Technical proofs are given in the Appendix.
2. Notations and Definitions

Let \( y_i \sim N(\theta_i, k(z_i)) \), \((i = 1, \ldots, n)\), be \( n \) independent normal observations for given \( z_i \in D \subseteq \mathbb{R} \). Again, consider the model (3)

\[
y_i \sim N(\theta_i, k(z_i)), \\
\theta_i \sim N(\alpha + \beta z_i, \lambda),
\]

where \( k : D \subseteq \mathbb{R} \rightarrow \mathbb{R}^+ \) is a known function and \( \alpha, \beta, \) and \( \lambda > 0 \) are the hyperparameters.

From Bayes’ theorem, the posterior distribution of \( \theta_i \) is

\[
N\left( \frac{\lambda}{\lambda + k(z_i)} y_i + \frac{k(z_i)}{\lambda + k(z_i)} (\alpha + \beta z_i), \frac{\lambda k(z_i)}{\lambda + k(z_i)} \right).
\]

The corresponding shrinkage estimate of \( \theta_i \) is given by

\[
\hat{\theta}_i^{a,\beta,\lambda} = \frac{\lambda}{\lambda + k(z_i)} y_i + \frac{k(z_i)}{\lambda + k(z_i)} (\alpha + \beta z_i). \tag{4}
\]

The empirical Bayes maximum likelihood estimators (EBMLE), Brown [2], are obtained by maximizing the marginal density of \( y_i \)'s with respect to \( \alpha, \beta, \) and \( \lambda \). The solutions satisfy in the following equations whenever the roots exist:

\[
\begin{align*}
\alpha \sum \frac{1}{\lambda + k(z_i)} + \beta \sum \frac{z_i}{\lambda + k(z_i)} &= \sum \frac{y_i}{\lambda + k(z_i)}, \\
\alpha \sum \frac{z_i^2}{\lambda + k(z_i)} + \beta \sum \frac{z_i y_i}{\lambda + k(z_i)} &= \sum \frac{z_i y_i}{\lambda + k(z_i)}, \\
\sum (\frac{1}{\lambda + k(z_i)})^2 (y_i - \alpha - \beta z_i)^2 - \frac{1}{\lambda + k(z_i)} &= 0.
\end{align*}
\]

Another approach to empirical Bayes method is something related to Stein’s unbiased risk estimate (SURE) which is based on the weighted mean of squared-loss,

\[
l_q(\theta_i, \hat{\theta}_i^{a,\beta,\lambda}) = \frac{1}{q_i} \sum q_i (\theta_i - \hat{\theta}_i^{a,\beta,\lambda})^2, \tag{5}
\]
where $q_i = \frac{1}{k(z_i)}$, see James and Stein [6]; Stein [8]; and Ghoreishi and Meshkani [5]. By applying (5) to the shrinkage estimator (4), the corresponding risk is given by

$$R(\theta, \hat{\lambda}, \alpha, \beta) = \sum_{q_i} \frac{1}{(\lambda + k(z_i))^2} \{k(z_i)(\theta_i - \alpha - \beta z_i)^2 + \lambda^2\}.$$  

An unbiased estimate for this risk is given by

$$U(\lambda, \alpha, \beta) = \sum_{q_i} \frac{k(z_i)(y_i - \alpha - \beta z_i)^2}{(\lambda + k(z_i))^2} + \frac{\lambda - k(z_i)}{\lambda + k(z_i)}.$$  

By minimizing of $U(\lambda, \alpha, \beta)$, the SURE estimates of $\alpha$, $\beta$, and $\lambda$ will be obtained. They satisfy the following equations whenever the roots exist:

$$\begin{align*}
\alpha \sum_{z_i} \frac{k(z_i)}{(\lambda + k(z_i))^2} + \beta \sum_{z_i} \frac{z_i k(z_i)}{(\lambda + k(z_i))^2} &= \sum_{z_i} \frac{k(z_i) y_i}{(\lambda + k(z_i))^2}, \\
\alpha \sum_{z_i} \frac{z_i k(z_i)}{(\lambda + k(z_i))^2} + \beta \sum_{z_i} \frac{z_i^2 k(z_i)}{(\lambda + k(z_i))^2} &= \sum_{z_i} \frac{z_i k(z_i) y_i}{(\lambda + k(z_i))^2}, \\
\sum_{z_i} \frac{k(z_i)}{(\lambda + k(z_i))^2} \{(y_i - \alpha - \beta z_i)^2 + \lambda + k(z_i)\} &= 0.
\end{align*}$$

An alternative approach to EBMLE and SURE is a fully Bayesian approach which assumes some second step priors for $\alpha$, $\beta$, and $\lambda$. Here, we assume normal distributions with large variances for $\alpha$ and $\beta$ and an inverse gamma distribution for $\lambda$. Therefore, the model (3) can be written as

$$y_i \sim N(\theta_i, k(z_i)),$$

$$\theta_i \sim N(\alpha + \beta z_i, \lambda),$$

$$\alpha \sim N(\mu_0, \gamma),$$

$$\beta \sim N(0, \gamma),$$

$$\lambda \sim Inv - gamma(r, \tau),$$

(6)
where the quantities $\mu_0, \gamma, r,$ and $\tau$ are assumed some known constants. Any inference for $\alpha, \beta,$ and $\lambda$ can be done using the corresponding marginal distributions. It is important to note that the marginal distribution of $\theta_i$ is a weighted sum of normal distributions and hence the corresponding shrinkage estimate is used to estimate the $\theta_i$. In the next section, we give the full posterior conditional distributions for Gibbs sampling purposes.

3. Main Results

In this section, we establish main results containing two theorems. The first is to illustrate the asymptotic properties of SURE estimators. In the previous section, we mentioned that our primary purpose in applying model (6) is how one can sample from posterior distributions. Therefore, the second theorem provides the full conditional posteriors.

**Theorem 3.1.** Assume the conditions (C1)-(C4) given in the Appendix. Then the SURE estimators satisfy in the following asymptotic property:

$$\sup_{0 \leq \lambda \leq M, 0 \leq \beta \leq M, 0 \leq \alpha \leq \infty} \left| U(\alpha, \beta, \lambda) - l_q(\theta, \hat{\theta}_i^{\alpha, \beta, \lambda}) \right| \rightarrow 0 \text{ in } L^2 \text{ as } n \rightarrow \infty.$$

The following result is obvious. It shows that in probability (as $n \rightarrow \infty$)

$$\left| l_q(\theta, \hat{\theta}_i^{\alpha, \beta, \lambda}) - l_q(\theta, \hat{\theta}_i^{\alpha, \beta, \lambda}) \right| \rightarrow 0.$$

This means that under conditions (C1)-(C4), the SURE estimators are asymptotically as good as the oracle estimators (OL) which obtain by minimizing (5) with respect to $\alpha, \beta,$ and $\lambda$.

For numerical computation, we need the following theorem:
Theorem 3.2. The full posterior conditional distributions needed for efficient posterior computation of model (6) are as follows:

(i) \( \theta_i | (\alpha, \beta, \lambda, z_i, y_i) \sim N \left( \frac{\lambda}{\lambda + k(z_i)} y_i + \frac{k(z_i)}{\lambda + k(z_i)} (\alpha + \beta z_i), \frac{\lambda k(z_i)}{\lambda + k(z_i)} \right) \),

(ii) \( \alpha | (\theta, \beta, \lambda, z_i) \sim N \left( \frac{\gamma}{n + \lambda} \sum (\theta_i - \beta z_i) + \frac{\lambda}{n + \lambda} \mu_0, \frac{\lambda \gamma}{n + \lambda} \right) \),

(iii) \( \beta | (\theta, \alpha, \lambda, z_i) \sim N \left( \frac{\gamma}{\sum z_i^2 + \lambda} \sum (\theta_i - \alpha z_i), \frac{\lambda \gamma}{\sum z_i^2 + \lambda} \right) \),

(iv) \( \lambda | (\theta, \alpha, \beta, z_i) \sim Inv - gamma \left( \frac{n}{2} + r, \tau + \sum (\theta_i - \beta z_i)^2 \right) \).

The last three items show that the joint distribution of \( (\alpha, \beta, \lambda) | z_i \) depends on the response \( y_i \) through \( \theta_i \). So the joint posterior distribution of \( (\alpha, \beta, \lambda) | z_i \) is conditionally independent of \( y_i \) given \( \theta_i \).

It is important to note that the results of the paper can easily be extend from a simple linear model, \( \alpha + \beta z_i \), for the mean of \( \theta_i \) to a multiple one with \( p \) regressors, i.e., \( \alpha + \beta_1 z_{i1} + \cdots + \beta_p z_{pi} \).

4. Application

To apply our theoretical results for real data, consider Bid at auction data in Table 1, see Lari [7]. These data belong to a big state company which wanted to survey its recent 12 auctions. It contains an explanatory variable \( z \): Bid at auction (in million dollars) and response variable \( y \): Cost of auction (in million dollars). The least squares regression shows \( R^2 = 0.880 \) and the mean square error MSE = 34.45, whereas the results of the weighted least square, where the weights are proportional to \( \frac{1}{z^2} \), lead to \( R^2 = 0.915 \) and the weighted mean square error WMSE = 0.792. This means the weighted least square provides a suitable fit. Therefore, we consider the following simple hierarchical heteroscedastic regression model:
\[ y_i \sim N(\theta_i, \sigma_i^2), \]
\[ \theta_i \sim N(\alpha + \beta z_i, \lambda). \]

We complete our model by substituting the weighted mean square error \( WMSE = 0.792 \) for \( \sigma^2 \). To evaluate the performance of shrinkage estimates (4), we use three introduced strategies: empirical Bayes maximum likelihood estimates, Stein’s unbiased risk estimates, and Bayesian approach to elicit the hyperparameters and in consequence the shrinkage estimates. For Bayesian approach, the following extra priors are assumed:

\[ \alpha \sim N(\bar{y}, 100), \]
\[ \beta \sim N(0, 100), \]
\[ \lambda \sim \text{Inv-gamma}(2, 2). \]

We apply the Gibbs sampler of Theorem 3.2 for \( N = 10000 \) draws. Moreover, the weighted least square (WLS) estimates are elicited. Our evaluative criterion in this work is the mean predictive errors,

\[ MPE = \frac{1}{n} \sum (\hat{\theta}_{i}^{\alpha, \beta, \lambda} - y_i)^2. \]

The results are given in Table 2. For the weighted least square \( MPE = 29.87 \).
Table 1. The Bid at auction data and corresponding shrinkage estimates

| No. | \( z \) | \( Y \) | \( \hat{\theta}^{WLS} \) | \( \hat{\theta}^{SURE} \) | \( \hat{\theta}^{Bayesian} \) | \( \hat{\theta}^{EBMLE} \) |
|-----|-----|-----|----------------|----------------|----------------|----------------|
| 1   | 2.13| 15.5| 14.58          | 15.47          | 15.49          | 15.49          |
| 2   | 1.21| 11.1| 10.73          | 11.06          | 11.11          | 11.07          |
| 3   | 11.00| 62.6| 51.75          | 62.58          | 62.6           | 62.60          |
| 4   | 6.00| 35.4| 30.80          | 35.38          | 35.4           | 35.40          |
| 5   | 5.60| 24.9| 29.12          | 24.92          | 24.91          | 24.90          |
| 6   | 6.91| 28.1| 34.61          | 28.12          | 28.11          | 28.10          |
| 7   | 2.97| 15.0| 18.10          | 15.06          | 15.01          | 15.00          |
| 8   | 3.35| 23.2| 19.70          | 23.15          | 23.19          | 23.19          |
| 9   | 10.39| 42.0| 49.20          | 42.01          | 42.00          | 42.00          |
| 10  | 1.10| 10.0| 10.27          | 10.03          | 10.01          | 9.98           |
| 11  | 4.36| 20.0| 23.93          | 20.04          | 20.01          | 20.00          |
| 12  | 8.00| 47.5| 39.18          | 47.48          | 47.50          | 47.49          |

Table 2. The prediction errors and hyperparameter estimates for various methods

| Method    | \( \hat{\alpha} \) | \( \hat{\beta} \) | \( \hat{\lambda} \) | MPE          |
|-----------|----------------|----------------|----------------|-------------|
| EBMLE     | 4.11 | 4.53          | 28.79          | 1.3 \times 10^{-4} |
| SURE      | 5.63 | 4.21          | 4.42           | 1.2 \times 10^{-3} |
| Bayesian  | 6.22 | 4.24          | 29.77          | 6.6 \times 10^{-5} |

Table 2 shows that although all three strategies work well in comparison with the weighted least square approach, the hierarchical Bayesian approach has the least MPE, additionally, its results are robust to the second step prior parameters. The corresponding shrinkage estimates of various methods are reported in Table 1.
References

[1] J. O. Berger and W. E. Strawderman, Choice of hierarchical priors: Admissibility in estimation of normal means, The Annals of Statistics 24(3) (1996), 931-951.
DOI: https://doi.org/10.1214/aos/1032526950

[2] L. D. Brown, In-season prediction of batting average: A field test of empirical Bayes and Bayes methodologies, The Annals of Applied Statistics 2(1) (2008), 113-152.

[3] L. D. Brown and E. Greenshtein, Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of means, The Annals of Statistics 37(4) (2009), 1685-1704.
DOI: https://doi.org/10.1214/08-AOS630

[4] S. K. Ghoreishi, Bayesian analysis of hierarchical heteroscedastic linear models using Dirichlet-Laplace priors, Journal of Statistical Theory and Applications 16(1) (2017), 53-64.
DOI: https://doi.org/10.2991/jsta.2017.16.1.5

[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, Journal of Multivariate Analysis 132 (2014), 129-137.
DOI: https://doi.org/10.1016/j.jmva.2014.08.001

[6] W. James and C. M. Stein, Estimation with quadratic loss, Proceedings of the 4th Berkeley Symposium on Probability and Statistics 1 (1961), 367-379.

[7] A. B. Lari, Applied Linear Regression, Shiraz University Press, Shiraz, 2006.

[8] C. M. Stein, Confidence sets for the mean of a multivariate normal distribution (with discussion), Journal of the Royal Statistical Society, Series B 24(2) (1962), 265-296.

[9] X. Xie, S. C. Kou and L. D. Brown, SURE estimates for a heteroscedastic hierarchical model, Journal of the American Statistical Association 107(500) (2012), 1465-1479.
DOI: https://doi.org/10.1080/01621459.2012.728154
Appendix

The proof of Theorem 3.2 is straightforward. For establishing Theorem 3.1, the following conditions are required:

(C1) \( \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} k(z_i) < \infty \).

(C2) \( \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \theta_i^2 < \infty \).

(C3) For given \( M > 0, |\alpha| \leq M < \infty \), and \( |\beta| \leq M < \infty \).

(C4) For given \( b > 0, |z_i| \leq b < \infty \).

Practically, assumptions (C3) and (C4) are not a restriction since no sensible shrinkage estimator would attempt to shrink toward a location that lies outside the sample range of the data.

It is easy to see

\[
U(\alpha, \beta, \lambda) - l_q(\theta, \hat{\alpha}, \hat{\beta}, \hat{\lambda}) = \frac{1}{\sum q_i} \sum \left\{ \frac{y_i^2 - k(z_i) - \theta_i^2}{k(z_i)} \right\} - \frac{2\lambda}{\lambda + k(z_i)} \\
\times \left( \frac{y_i^2 - k(z_i) - \theta_i y_i}{k(z_i)} \right) - \frac{2\alpha}{\sum q_i} \sum q_i \frac{k(z_i) (y_i - \theta_i)}{\lambda + k(z_i)} \\
- \frac{2\beta}{\sum q_i} \sum q_i \frac{k(z_i) (y_i - \theta_i)}{\lambda + k(z_i)} ,
\]

and hence, we have

\[
\sup_{\substack{0 \leq |\alpha| \leq M, \\ 0 \leq |\beta| \leq M, \\ 0 \leq \lambda < \infty}} |U(\alpha, \beta, \lambda) - l_q(\theta, \hat{\alpha}, \hat{\beta}, \hat{\lambda})| \\
\leq \sup_{0 \leq \lambda < \infty} \left| \frac{1}{\sum q_i} \sum q_i \left\{ \frac{y_i^2 - k(z_i) - \theta_i^2}{k(z_i)} \right\} - \frac{2\lambda}{\lambda + k(z_i)} \right|
\]
\[
\begin{align*}
\left( \frac{y_i^2 - k(z_i) - \theta_i y_i}{k(z_i)} \right) &+ \sup_{0 \leq |t| \leq M, 0 \leq \lambda < \infty} \left| \sum_{i=1}^n q_i \sum_{i=1}^n q_i \frac{k(z_i)(y_i - \theta_i)}{\lambda + k(z_i)} \right| \\
&+ \sup_{0 \leq |t| \leq M, 0 \leq \lambda < \infty} \left| \sum_{i=1}^n q_i \sum_{i=1}^n q_i \frac{k(z_i)(y_i - \theta_i)}{\lambda + k(z_i)} \right|
\end{align*}
\]

The first and second parts in right hand side are the straightforward results of Theorems 3.1 and 3.2 in Ghor eishi and Meshkani [5], so it is enough to consider the third part,

\[
\sup_{0 \leq |t| \leq M, 0 \leq \lambda < \infty} \left| \sum_{i=1}^n q_i \sum_{i=1}^n q_i \frac{k(z_i)(y_i - \theta_i)}{\lambda + k(z_i)} \right| < \frac{2M}{\sum_{i=1}^n q_i} \sup_{0 \leq \lambda < \infty} \left| \sum_{i=1}^n q_i \frac{k(z_i)(y_i - \theta_i)}{\lambda + k(z_i)} \right|
\]

\[
\leq \frac{2M}{\sum_{i=1}^n q_i} \max_{0 \leq j \leq n} \left| \sum_{i=1}^n q_i \frac{k(z_i)(y_i - \theta_i)}{\lambda + k(z_i)} \right|
\]

By the $L^2$ maximum inequality on martingales, one can have

\[
E\left[ \sup_{0 \leq \lambda < \infty} \sum_{i=1}^n q_i \sum_{i=1}^n q_i \frac{k(z_i)(y_i - \theta_i)}{\lambda + k(z_i)} \right]^2 \leq E\left[ \max_{0 \leq j \leq n} \sum_{i=1}^n q_i \sum_{i=1}^n q_i \frac{k(z_i)(y_i - \theta_i)}{\lambda + k(z_i)} \right]^2
\]

\[
\leq 4 \left( \frac{1}{\sum_{i=1}^n q_i} \right)^2 \sum_{i=1}^n q_i^2 k(z_i)
\]

\[
= 4 \frac{b^2}{\sum_{i=1}^n q_i} \leq 4 \frac{b^2}{n^2} \sum_{i=1}^n k(z_i).
\]