Shrinkage estimators for covariance matrices

Michael J. Daniels [1]

Robert E. Kass [2]

[1] Department of Statistics, Iowa State University

[2] Department of Statistics, Carnegie Mellon University

Department of Statistics

Iowa State University

102G Snedecor Hall

Ames, IA 50011

e-mail: mdaniels@iastate.edu

phone: (515) 294 7775

fax: (515) 294 4040

Summary

Estimation of covariance matrices in small samples has been studied by many authors. Standard estimators, like the unstructured maximum likelihood estimator (ML) or restricted maximum likelihood (REML) estimator can be very unstable with the smallest estimated eigenvalues being too small and the largest too big. A standard approach to more stably estimating the matrix in small samples is to compute the ML or REML estimator under some simple structure that involves estimation of fewer parameters, such as compound symmetry or independence. However, these estimators will not be consistent unless the hypothesized structure is correct. If interest focuses on estimation of regression coefficients with correlated (or longitudinal) data, a sandwich estimator of the covariance matrix may be used to provide standard errors for the estimated coefficients that are robust in the sense that they remain consistent under misspecification of the covariance structure. With large matrices, however, the inefficiency of the sandwich estimator becomes worrisome.

We consider here two general "shrinkage" approaches to estimating the covariance matrix and regression coefficients. The first involves shrinking the eigenvalues of the unstructured ML or REML estimator. The second involves shrinking an unstructured estimator toward a structured estimator. For both cases, the data determine the amount of shrinkage. These estimators are consistent and give consistent and asymptotically efficient estimates for regression coefficients. Simulations show the improved operating characteristics of the shrinkage estimators of the covariance matrix and the regression coefficients in finite samples. The final estimator chosen includes a combination of both shrinkage approaches, i.e., shrinking the eigenvalues and then shrinking towards structure. We illustrate our approach on a sleep EEG study which requires estimation of a 24×24 covariance matrix and for which inferences on mean parameters critically depend on the covariance estimator chosen. We recommend making inference using a particular shrinkage estimator that provides a reasonable compromise between structured and unstructured estimators.

1 Introduction

Since James and Stein (1961), many authors have explored better estimators for a covariance matrix. These have been derived from a decision-theoretic perspective (Stein, 1975; Dey and Svrinisan, 1985; Lin and Perlman, 1985; Haff, 1991) or by specifying an appropriate prior for the covariance matrix and choosing an estimator based on a particular loss function (Yang and Berger, 1994; Daniels and Kass, 1999). We will take the latter approach in this paper and build on the work Daniels and Kass (1999). The general approach will be to first generically stabilize an unstructured estimate of the covariance matrix and then to 'shrink' this estimate toward a parsimonious, structured form of the matrix. In this approach, the data will determine how much shrinkage is required.

We will now provide some more details on the approach in Daniels and Kass (1999); for a similar approach in the context of a covariance function in time series data, see Daniels and Cressie (2001). To develop more stable estimators for the covariance matrix, Daniels and Kass (1999) shrink the matrix toward a diagonal structure and obtain estimates (and posterior distributions) using combinations of importance sampling and MCMC. Here, we will extend this work in several ways. First, we consider shrinking toward any structure (not just diagonal), including non-stationary models, such as (structured) antedependence models (Zimmerman and Nunez-Anton, 1997) for longitudinal data, and stationary models, such as compound symmetry, AR, and ARMA models (see early work by Chen, 1979 and Lin and Perlman, 1985). Second, we propose estimators that can be computed by simple approximations, avoiding the fully MCMC approach, which can be computationally intensive for large matrices. These estimators will provide more stability than an unstructured estimator, but still provide robustness to mis-specification of the structure. These estimators might be called Empirical (or approximate) Bayes estimators (Kass and Steffey, 1989).

Another approach to inducing stability in estimating a covariance matrix is to shrink the eigenvalues (Stein, 1975; Dey and Srinivasan, 1985; Haff, 1991; Yang and Berger, 1994; Ledoit, 1997). It is well known that the sample eigenvalues are biased; the smallest is too small, and the largest, too large. We will review some of these estimators and propose another, derived from a simple prior distribution on the eigenvalues.

We will also examine the way these estimators influence covariate effects. Consider a fixed effects regression model with correlated errors, a common model for longitudinal and clustered data (Diggle, Liang, and Zeger, 1994),

$$Y_i \sim N(X_i\beta, \Sigma), \tag{1}$$

where Y_i is a $p \times 1$ vector and β is a $q \times 1$ vector; $i = 1, \ldots, n$. When n is small relative to p, estimation of the covariance matrix can be unstable. The most common approach to inducing stability is to assume some true structure and then estimate the relevant parameters, which will be fewer than those in the full covariance matrix, $p(p+1)/2$. However, the resulting variance of $\hat{\beta}$ will be incorrect if the hypothesized covariance structure is incorrect. To avoid this problem, a sandwich estimator (Huber, 1967) for the variance of $\hat{\beta}$ will result in asymptotically correct, but not efficient, variance estimates for $\hat{\beta}$. As a robust alternative to assuming some structure and as a way to induce stability over the unstructured estimator of the covariance matrix, we will use the covariance matrix estimators proposed above. Our strategy will be to use the following procedure to estimate Σ:

Step 1. Fit the model (1) using an unstructured covariance matrix, $\hat{\Sigma}$.

Step 2. Shrink the eigenvalues of the unstructured estimator to obtain a more stable estimate, $\hat{\Sigma}_{st}$.

Step 3. Fit the model (1) assuming some covariance structure, $\hat{\Sigma}_s$. This might be chosen, for example, using the Bayesian information criterion (BIC) (see, e.g., Zimmerman and Nunez Anton, 1997).

Step 4. Compute estimates of the parameters that determine the amount of shrinkage using $\hat{\Sigma}_{st}$ and $\hat{\Sigma}_{s}$.

Step 5. Compute a shrinkage estimator of the covariance matrix, $\hat{\Sigma}_{sh}$ using the estimates in steps (2)-(4).

Step 6. Re-fit model (1) conditional on $\hat{\Sigma}_{sh}$.

More details on the shrinkage parameters and shrinkage estimators will be given in Section 3. Steps 1-3 and 6 are easy and can be implemented in SAS proc mixed (SAS Institute, 1999). Steps 4 and 5 will require a simple macro. In section 4, we recommend two estimators. Both first shrink the eigenvalues together (Step 2 above). Then, this modified estimate is shrunk towards a chosen structure under two different parametrizations of the matrix (Steps 4 and 5 above).

In Section 2, we will explore estimators that shrink the eigenvalues and in Section 3, estimators that shrink an unstructured estimate of the covariance matrix toward some structure. In Section 4, we will do simulations to examine the risk of these estimators for estimating the sample covariance matrix and the mean squared error for estimating the fixed regression coefficients. Section 5 includes a theorem demonstrating the asymptotic properties of these estimators. In Section 6, we will show the importance of covariance matrix estimation on fixed regression coefficients in data from a sleep EEG study. We offer conclusions and extensions in Section 7.

2 Squeezing the eigenvalues together

In this section, we review and discuss two estimators to shrink the eigenvalues together and introduce a third based on a hierarchical model.

2.1 Stein Estimator

We first examine estimators that shrink the eigenvalues. Several authors have focused on orthogonally invariant estimators of the form, $\tilde{\Sigma} = O\Lambda^{\star}(\hat{\lambda})O^T$ where O is the matrix of normalized eigenvectors, $\hat{\lambda}$ is the vector of sample eigenvalues and $\Lambda^{\star}(\hat{\lambda}) = \text{diag}(\lambda_1^{\star}(\hat{\lambda}), \ldots \lambda_p^{\star}(\hat{\lambda}))$, where each $\lambda_j^{\star}$ is a real-valued non-negative function. Stein (1975) proposed setting $\lambda_j^{\star}(\hat{\lambda}) = n\hat{\lambda}_j/\alpha_j$ where $\alpha_j = n-p+1+2\hat{\lambda}_j \sum_{i \neq j} \frac{1}{\hat{\lambda}_j - \hat{\lambda}_i}$. This estimator minimizes an unbiased estimate of Stein's (entropy) loss under this class of estimators; this estimator has similar operating characteristics to the estimator derived from the Yang and Berger (1994) prior (see Yang and Berger, 1994). However, this estimator does not preserve the order of the eigenvalues and the resulting estimators of the eigenvalues can be negative. Thus, an isotonizing algorithm is often needed. Haff (1991) derived an estimator similar to Stein's, but computed under the constraint of maintaining the order of the sample eigenvalues; the estimator is identical when Stein's does not require the isotonozing algorithm. This estimator operates by adaptively shrinking the sample eigenvalues.

2.2 Ledoit Estimator

Ledoit (1997) introduced an estimator that is the optimal linear combination of the identity matrix and the sample covariance matrix under squared error loss. This is equivalent to finding the optimal linear shrinkage of the eigenvalues. This estimator has the advantage of being able to be computed when the dimension of the matrix is larger than the sample size. However, the use of squared error loss as the loss function for the covariance matrix can result in over-shrinkage of the eigenvalues (see the simulation in Section 4), especially the small ones. When the eigenvalues are very close together, this estimator performs very well, but when they are very far apart, it performs quite poorly. This is a by-product of deriving estimators using squared error loss which does not offer a severe penalty for overestimating the small eigenvalues; Stein's loss does penalize for this.

2.3 An estimator based on a simple hierarchical model

As an alternative to the above estimators, we suggest placing normal prior distribution on the logarithm of the eigenvalues, $\log(\lambda_i)$, $i = 1, \ldots, p$: $\log(\lambda_i)|\tau^2 \sim_{iid} N(\log(\lambda), \tau^2)$. To form a simple estimator based on this prior distribution, we can approximate the likelihood for the eigenvalues from the model, with $\log(\hat{\lambda}_i) \sim N(\log(\lambda_i), 2/n)$, the asymptotic distribution of the logarithm of the sample eigenvalues. A model formed from the approximation and the prior suggests a simple closed form estimator for the logarithm of the eigenvalues, the posterior mean, which here is also the posterior mode, of λ_i conditional on estimates $\log(\hat{\lambda})$ and $\hat{\tau}^2$. We exponentiate this estimator to obtain the following,

$$\tilde{\lambda}_i = \exp(\frac{2/n}{2/n + \hat{\tau}^2} \log(\hat{\lambda}) + \frac{\hat{\tau}^2}{2/n + \hat{\tau}^2} \log(\hat{\lambda}_i)). \tag{2}$$

We will call the estimator in (2) the log eigenvalue posterior mean estimator. We use the following approach to estimate the hyperparameters. For $\log(\lambda)$, we plug in the mean of the logarithm of the sample eigenvalues; this corresponds to the posterior mode under a uniform prior distribution on $\log(\lambda)$. For τ^2, we plug in

$$\hat{\tau}^2 = \sum_{i=1}^{p} (\log(\hat{\lambda}_i) - \log(\lambda))^2/(p+4) - 2/n.$$

This estimator was chosen to encourage shrinkage of the eigenvalues and corresponds the posterior mode under a uniform shrinkage prior on τ^2, $\pi(\tau^2) = \frac{2/n}{(2/n+\tau^2)^2}$ (see Daniels, 1999).

The estimation of τ^2 may be considered a special case of using the posterior mode based on a prior for τ^2 proportional to $(2/n + \tau^2)^{-k}$, which gives a posterior mode of

$$\hat{\tau}^2 = \sum (\log(\hat{\lambda}_i) - log(\lambda))^2/c - 2/n$$

where $c = p+2k$. Other possible approaches to estimate τ^2 include an estimator based on the posterior mode from a flat prior on τ^2, $k = 0$ or the James Stein estimator, $k = -5/6$ (see for example, Efron and Morris, 1973; Strawderman, 1971). However, preliminary simulations suggest that using these estimators for τ^2 tends to undershrink the eigenvalues. Note that our estimator (2), unlike Stein's, maintains the order of the sample eigenvalues.

3 Shrinking toward structure

We will focus on two parameterizations in which to shrink the covariance matrix as in Daniels and Kass (1999). The first will be useful for shrinking towards a correlation (or covariance) structure, but is not guaranteed to produce a positive definite matrix. The second can be used for shrinking toward a covariance structure and is guaranteed to give a positive definite matrix. The latter will involve decomposing the matrix into the Givens angles and the eigenvalues (see, e.g., Daniels and Kass, 1999). Early work by Chen (1979) and others suggests doing shrinkage with a Wishart prior, with scale matrix set to the structure and unknown degrees of freedom. However, this formulation only allows one parameter, the degrees of freedom, to characterize the shrinkage. In addition, there is a natural bound on the degrees of freedom to maintain a proper prior (i.e., the degrees of freedom must be greater than p), which can produce poor results when the hypothesized structure is incorrect (Daniels and Kass, 1999). The two approaches we suggest allow for two variance parameters: one for the correlations (angles) and one for the variances (eigenvalues).

The general approach will be the following. First, fit model (1) using maximum likelihood with an unstructured form for Σ. Then, conditional on $\hat{\beta}$, compute the observed information matrix for Σ based on one of the two parameterizations. For the first estimator, the correlation shrinkage, we simplify computations by treating the information matrix for the variances and correlations as if it was block diagonal, similar to Lin and Perlman (1985), while for the second estimator, the rotation shrinkage, the information matrix for the eigenvalues and angles is block diagonal (Yang and Berger, 1994). We then form the estimators, detailed in the next two sections, by developing a two-level model. At the first level, we approximate the distribution of the maximum likelihood estimator of Σ by a normal distribution with variance the inverse of the observed information matrix. At the second level, we introduce a normal prior whose purpose is to shrink the correlations/variance or angles/eigenvalues toward a structured form with unknown variance components, which will be estimated from the data as detailed in the next section. The unknown parameters of the structured form are estimated by fitting model (1) under that structure, and thereafter assumed known. As a result, we estimate only the two variance components, in addition to the parameters appearing in the structured model, which involves many fewer than the $p(p+1)/2$ parameters estimated in the unstructured case. From this two level normal-normal model, we can compute a simple shrinkage estimator, which will give consistent estimates of Σ, even under misspecification of the structure. We provide more details on these two estimators in the next two sections.

3.1 Correlation shrinkage estimator

One way to parameterize a covariance matrix is through the correlations and variances (Barnard, McCulloch, and Meng, 2000; Lin and Perlman, 1985). Daniels and Kass (1999) suggested placing normal priors on the z-transform of the correlations and shrinking these toward zero (equivalent to shrinking towards a diagonal matrix). We suggest replacing their prior with the following more general prior distribution on the z-transform of the correlations:

$$z(\rho) \sim N(z(\rho_s), \tau_\rho^2), \tag{3}$$

where ρ_s represents the correlations specified under the assumed covariance structure. Here, we assume a priori that the correlations are iid. Given the interplay of the correlations in the covariance matrix being positive definite, this may not be a realistic assumption. However, the simulations in Section 4 and in Daniels and Kass (1999) show that this is a reasonable specification in practice. To form an approximate estimator based on this prior, we need the asymptotic distribution of the correlations, given in Lin and Perlman (1985, pp. 417-418), who introduced a James-Stein estimator based on shrinking the z-transform of the correlations toward a common value. Given the asymptotic distribution of the correlations, we can approximate the specified model for ρ with a normal-normal model. Based on this model, we replace the sample correlations $\hat{\rho}$ with the following estimator:

$$\tilde{\rho} = \mathrm{z}^{-1}[(I(z(\hat{\rho}))^{-1} + \hat{\tau}^2 I)^{-1}\hat{\tau}^2 I z(\hat{\rho}) + (I(z(\hat{\rho}))^{-1} + \hat{\tau}^2 I)^{-1}(I(z(\hat{\rho}))^{-1} z(\hat{\rho}_s)] \tag{4}$$

where $I(z(\hat{\rho}))$ is the observed information matrix for the z-transform of the correlations and Fisher's z-transform of a vector is defined as the vector of z-transforms.

The estimated correlation under the assumed structure, $\hat{\rho}_s$, is computed by fitting the structured model (by maximum likelihood or restricted maximum likelihood) and extracting the relevant estimated correlations. To estimate τ_ρ^2, we generalize the moment estimator in DerSimonian and Laird (1986) to the multivariate case and use

$$\hat{\tau}_\rho^2 = ((z(\hat{\rho}) - z(\rho_s))^T I(z(\hat{\rho}))(z(\hat{\rho}) - z(\rho_s)) - p(p-1)/2)/tr(I(z(\hat{\rho}))). \tag{5}$$

For $p = 2$, this simplifies to their estimator.

To shrink towards a covariance rather than a correlation structure, we can also shrink the variances using $\log(\sigma^2) \sim N(\log(\sigma_s^2), \tau_\sigma^2)$ and form an approximate estimator using the asymptotic distribution of the log of the variances:

$$\tilde{\sigma}^2 = \log^{-1}[(I(\log(\hat{\sigma}^2))^{-1} + \hat{\tau}^2 I)^{-1}\hat{\tau}^2 I \log(\hat{\sigma}^2) + (I(\log(\hat{\sigma}^2))^{-1} + \hat{\tau}^2 I)^{-1}(I(\log(\hat{\sigma}^2))^{-1} \log(\hat{\sigma}^{2s})] \tag{6}$$

where $I(\log(\hat{\sigma}^2))$ is the observed information matrix for the logarithm of the variance and logarithm of a vector is defined as the vector of logarithms. Note that although the variances and correlations are not asymptotically independent, for simplicity we assume this when forming the shrinkage estimator of the covariance matrix. We will refer to (4) and (6) as the correlation shrinkage estimator.

We mention here that this estimator of the covariance matrix is not guaranteed to be positive definite. This is a deficiency of this estimator. However, this is a separate issue from the accuracy of the asymptotics. The asymptotics failing is not the reason the estimate is not positive definite. The problem is that we are using individual linear combinations of a transformation of the correlations in our estimator, so even though the unstructured and structured estimates are both positive definite, the resulting estimator may not be. However, the estimator proposed in the next section is guaranteed to be positive definite.

3.2 Rotation shrinkage estimator

Another way to parameterize the covariance matrix is through the eigenvalues and a decomposition of the rotation matrix through the $p(p-1)/2$ Givens angles (Hoffman, Raffenetti, and Ruedenberg, 1972). Similar to the prior underlying the correlation shrinkage estimator, we replace the prior in Daniels and Kass (1999) with the following more general prior distribution on the Givens angles:

$$\text{logit}(\theta) = \log(\frac{\pi/2+\theta}{\pi/2-\theta}) \sim N(\text{logit}(\theta_s), \tau_\theta^2).$$

where θ_s represents the Givens angles specified under the assumed covariance structure. To form an approximate estimator based on this prior, we will need to derive the asymptotic distribution of the Givens angles. The diagonal element of the information matrix corresponding to θ_{ij} takes the form:

$$\begin{aligned} I(\theta_{ij},\theta_{ij}) &= tr(\Sigma^{-1}d(O^t\Lambda O)/d\theta_{ij}\Sigma^{-1}d(O^t\Lambda O)/d\theta_{ij}) \\ &= -2+2tr(dO/d\theta_{ij}O^t\Lambda^{-1}OdO/d\theta_{ij}\Lambda)+tr(Od(O^t)/d\theta_{ij}d(O^t)/d\theta_{ij}) \end{aligned}$$

and the offdiagonal elements corresponding to $(\theta_{ij},\theta_{rs})$,

$$I(\theta_{ij},\theta_{rs}) \ = 2tr(dO/d\theta_{rs}O^t\Lambda^{-1}OdO/d\theta_{ij}\Lambda)+tr(OdO^t/d\theta_{rs}dO^t/d\theta_{ij}).$$

For additional details, see the Appendix.

Given the asymptotic distribution of the Givens angles, we can approximate the specified model for θ with a normal-normal model. Based on this model, we replace the sample Givens angles with the following estimator:

$$\tilde{\theta} = \text{logit}^{-1}[(I(\text{logit}(\hat{\theta}))^{-1} + \hat{\tau}^2 I)^{-1}\hat{\tau}^2 I(\text{logit}(\hat{\theta}))] + (I(\text{logit}(\hat{\theta}))^{-1} + \hat{\tau}^2 I)^{-1}(I(\text{logit}(\hat{\theta}))^{-1}\text{logit}(\hat{\theta}_s)] \quad (7)$$

where $I(\text{logit}(\hat{\theta}))$ is the observed information matrix for the logit of the Givens angles, the logit of a vector is the vector of the logits, and logit^{-1} is the inverse of the logit function. $\hat{\theta}_s$ is computed using the same procedure as $\hat{\rho}_s$; estimation of τ^2 uses (5) with θ in place of ρ.

To shrink toward a structure, we also need to shrink the eigenvalues. We will use a modification of log eigenvalue posterior mean estimator given in (2), replacing $\log(\hat{\lambda})$ with $\log(\hat{\lambda}_s)$,

$$\tilde{\lambda}_i = \exp(\frac{2/n}{2/n + \hat{\tau}^2}\log(\hat{\lambda}_s) + \frac{\hat{\tau}^2}{2/n + \hat{\tau}^2}\log(\hat{\lambda}_i)). \quad (8)$$

We will subsequently refer to the estimator, specified by (7) and (8), which shrinks the Givens angles and the log eigenvalues as the *rotation shrinkage estimator* and the estimator which only shrinks the log of the eigenvalues, specified by (8) as the *structured log eigenvalue estimator*.

4 Simulation Studies

We will perform two simulations to assess the operating characteristics of these estimators. The first will evaluate the risk in estimating a covariance matrix. The second will evaluate mean squared error of the estimators of the regression coefficients.

4.1 Estimation of the covariance matrix

4.1.1 Design

For the first simulation, we will examine the risk for each of these estimators using Stein's loss, $L_1 = tr(\hat{\Sigma}\Sigma^{-1}) - \log|\hat{\Sigma}\Sigma^{-1}| - p$. Specifically, we consider the following model:

$$Y_i \sim N(0, \Sigma).$$

Because the reason for using the shrinkage estimators is to improve upon the sample covariance estimator, for each estimator we have tabulated the percentage reduction in average loss (PRIAL), which is defined as the difference between the risk of the sample covariance matrix and the risk of the estimator divided by the risk of the sample covariance. We consider shrinking towards two structures, diagonal and AR(1), labelled as simulations A1 and A2, respectively. For A1, we consider the following six true matrices: Two diagonal covariance matrices, (I) equal eigenvalues, and (II) a somewhat ill-conditioned matrix, to examine how the estimators which shrink the eigenvalues perform. We also considered three non-diagonal matrices, two rotations of matrix II, R_1 and R_2 (in terms of Givens angles) and a matrix with all small correlations. The latter will be a situation that is "close" to diagonal. The matrices are given below:

I diag(1,1,1,1,1)

II $\text{diag}(1, .75, (.75)^2, (.75)^{10}, (.75)^{20}) = \text{diag}(1, .75, .56, .06, .003)$

IIR_1 matrix II with Givens angles all set to $\pi/4$

IIR_2 matrix II with Givens angles evenly spaced between $(-\pi/4, \pi/4)$

III variances all set to 1, correlations all set to .1

For A2, we also considered six true matrices: an AR(1) structure (IA), a structure close to AR(1) (IIA), and four matrices far from the AR(1) structure (IIIA-VIA), three with a Toeplitz correlation structure in which the elements are equal along subdiagonals (correlation bands). The matrices are given below:

I-A2 AR-1 structure, variances set to $(1, 1, 1, 1, 1)$ and correlation bands set to $(.7, .49, .34, .24)$

II-A2 close to AR-1 structure, variances set to $(.8, .9, 1.0, 1.1, 1.2)$ and correlation bands set to (.4, .15, .10, .01)

III-A2 far from AR-1 structure, variances set to $(1, 2, 3, 4, 5)$ and correlation bands set to $(.7, .7, .7, .7)$

IV-A2 far from AR-1 structure, variances set to $(1, .01, .01, .01, .01)$ and correlation bands set to $(0, 0, 0, 0)$

V-A2 far from AR-1 structure, variances set to $(1, .01, .01, .01, .01)$ and correlation bands set to $(.8, .7, .8, .9)$

VI-A2 far from AR-1 structure, eigenvalues set to $(1, .75^2, .75^3, .75^4)$ with Givens angles evenly spaced between $(-\pi/4, \pi/4)$

We ran simulation A1 for sample sizes of $n = 10, 20, 40$ and $p = 5$ and use the Stein, Ledoit, and log eigenvalue posterior mean estimators to shrink the eigenvalues and the rotation and correlation (without the variances, (4)) shrinkage estimators to shrink toward a structure (diagonal). We also re-ran simulation A1 for sample sizes of $n = 40, 100$ and $p = 20$. We ran simulation A2 for sample sizes of $n = 10, 20, 50$. Additional details on the estimators used in Simulation A2 are given in the next section.

4.1.2 Results

The results of simulation A1 appear in Table 1. The eigenvalue estimators all shrink the sample eigenvalues. Thus, the largest gains are evident when the eigenvalues are close together (with PRIAL as large as 67%). Specifically, the Stein performs well across all situations, especially when the true eigenvalues form one or several groups, while the log eigenvalue estimator does best when the true eigenvalues are close together. The Ledoit estimator also does very well for eigenvalues that are close together (with PRIAL as high as 90%), but considerably overshrinks when they are spread out, severely overestimating the smallest eigenvalues.

For the shrinkage-to-structure estimators, the correlation and rotation shrinkage estimators, when the structure is correct or "close" (matrix III), the estimators do very well (PRIAL from $30-60\%$); the rotation estimator has some problems regarding the accuracy of the asymptotics (which we will comment upon more in Section 4.3). When the structure is not close, they can do poorly in small samples (see cases II-R_1 and R_2); in addition, we see the the log eigenvalue posterior mean estimator is the same as the structured log eigenvalue estimator for matrix I (and it does well). However, as the sample size gets large, the data will dominate the incorrectly specified structure.

The results of simulation A1 for $p = 20$ appear in Table 2. In terms of introducing stability by shrinking the eigenvalues, the Stein and log eigenvalue posterior mean estimator gave smaller risks for all cases; the Ledoit estimator still overshrunk the small eigenvalues. The correlation shrinkage estimator did best when shrinking toward the correct structure (or close to it), matrices I, II, and III, with PRIAL 16% to 70%.

After some preliminary work for simulation A2, we decided to stabilize the correlation and rotation estimators by shrinking the Stein estimator, as opposed to the unstructured maximum likelihood (ML) estimator toward the AR(1) structure because the unstructured ML estimator is very unstable in small to medium size samples. In addition, in simulation A1, the Stein did best overall among the estimators introduced in Section 2 so in this simulation, we use the Stein to shrink the eigenvalues and the correlation, structured log eigenvalue, and rotation estimators to shrink toward a structure (AR(1)). We consider the structured log eigenvalue estimator separately from the rotation estimator as there is concern about the asymptotic approximations used in computing the rotation estimator.

The results of simulation A2 appear in Table 3. The Stein estimator performed well for all six cases (PRIAL from $3-50\%$). The structured log eigenvalue estimator also resulted in reductions in loss for all the cases, with the largest reductions when the true structure was close to AR(1) (PRIAL $15-50\%$). The correlation shrinkage estimator had large reductions when the structure was close to correct (Cases I, II, and VI) and the rotation shrinkage estimator had similar problems with asymptotics to simulation A1.

4.2 Estimation of the regression coefficients

The first simulation showed reductions in risk of as much as 70% in estimating Σ. We now focus on estimation of the regression coefficients (or fixed effects) in model (1).

4.2.1 Design

To evaluate our estimators for β,

$$\hat{\beta} = (\sum X_i \hat{\Sigma}^{-1} X_i')^{-1} \sum X_i \hat{\Sigma}^{-1} Y_i,$$

where $\hat{\Sigma}$ is our estimate of the covariance matrix, we will compute mean squared error. We leave examination of actual coverage probabilities of confidence regions for future work in which we will adjust the standard error of the regression coefficients for the uncertainty in estimating Σ.

We conduct two simulations, B1 and B2, varying the dimension of the covariance matrix, Σ and the regression vector, β. In simulation B1, we assume $X_{it} = (poly_t(1), g_i)$, where $t = 1, \ldots, p$, $poly_t(1)$ is a first order polynomial in t, and g_i is a binary indicator equal to 1 with probability 1/2 and -1 with probability 1/2. This design matrix was chosen to allow for both time-varying and baseline covariates. We set $\beta = (5, 3, 2)$. We consider $p = 5$ and sample sizes of $10, 20, 40, 100$. In simulation B2, we assume $X_{it} = (poly_t(3), g_i, a_i, b_i, c_i)$, where a_i is a binary indicator equal to 1 with probability 4/5, b_i is a standard normal, and c_i is a scaled chi-squared random variable with three degrees of freedom. Similar to the first simulation, this design matrix was chosen to allow for time-varying and baseline covariates with different distributions. We set $\beta = (4, 2, -2, 2, 1, 2, .5, .7)$ and consider $p = 10$ with $n = 20, 40, 100$.

For simulation B1, we consider six matrices. The matrices, labelled with the suffix, -B1, are the same as the -A2 matrices, except matrix II-B1 is slightly different and appears below:

II-B1 close to AR-1 structure, variances set to $(.3, .7, 1.1, 1.5, 1.9)$ and correlation bands set to (.4, .25, .22, .24)

For simulation B2, we extend the simulation B1 true matrices to $p = 10$ as follows:

I-B2 AR-1 structure, variances set to $(1, 1, 1, 1, 1)$ and correlation bands set to $(.7, .7^2, \ldots, .7^9)$

II-B2 close to AR-1 structure, variances evenly spaced from .8 to 1.2 and correlation bands set to (.9, .8, .75, .68, .57, .52, .48, .42, .37)

III-B2 far from AR-1 structure, variances evenly spaced from 1 to 10 and correlation bands set to $(.7, .7, \ldots, .7)$

IV-B2 far from AR-1 structure, variances set to $(1, .01, .01, \ldots, .01)$ and correlation bands set to $(0, 0, \ldots, 0)$

V-B2 far from AR-1 structure, variances same as IV-A2 and correlation bands set to $(.8, .7, .7, \ldots, .7, .9)$.

VI-B2 far from AR-1 structure, eigenvalues set to $(1, .75, .75^2, \ldots, .75^9)$ with Givens angles evenly spaced between $(-\pi/4, \pi/4)$

We compute the unstructured restricted maximum likelihood estimator, the maximum likelihood estimator under AR(1), the Stein eigenvalue estimator, and the correlation, structured log eigenvalue, and rotation estimators. Similar to simulation A2 in section 4.1, we decided to stabilize the correlation and rotation estimators by shrinking the Stein estimator.

4.2.2 Results

The results of simulations B1 and B2 appear in Tables 4 and 5. First, we point out that when the structure is correct (case I-B1(2)), the structured estimator dominates in terms of MSE (reductions of over 50% over the unstructured estimator). However, the shrinkage estimators are competitive even for the very small sample size, $n = 10$ (reduction as large as 30%) and the shrinkage estimators dominate when the structure is incorrect; the structured estimator does very poorly in this case (in particular, see cases IV and V- B1(2)). When the structure is correct or close (case I-, II-B1(2)), the correlation and structured log eigenvalue outperform the unstructured (reductions as large as 50%) and for the small sample size, the structured log eigenvalues estimator does well against the unstructured even for matrices III-B1(2) through V-B1(2) (far from AR(1)) with reduction as large as 33% (and no increases). The Stein estimator is never worse than the unstructured and is better than the estimators which shrink toward AR(1) when the structure is far from correct (III-B1(2) through VI-B1(2)) for sample sizes of 20 and 40 (40 and 100). For case II-B2 of simulation B2, the correlation shrinkage estimator appears to do worse than the Stein and structured log eigenvalues estimator for $n = 100$. However, for this case, a high proportion (about 40%)

of the correlation shrinkage estimates were not positive definite. If we restrict the comparison of losses to datasets which resulted in positive definite correlation shrinkage estimates, the correlation shrinkage estimator did about as well as the Stein and structured log eigenvalue estimators. The rotation shrinkage estimator has trouble when the eigenvalues are close together and its sampling distribution is not well approximated by a normal distribution in non-large samples as discussed in section 4.1. Overall, across all the sample sizes, if the true matrix is close to the hypothesized structure, the correlation and structured log eigenvalue estimator do best with reductions in MSE as high as 30 to 40% for small sample sizes.

4.3 Conclusions from the simulations

Overall, the shrinkage estimators have much smaller risk than the unstructured maximum likelihood estimator (the sample covariance matrix in the simplest case). As the sample size grows, the shrinkage estimators and unstructured MLE become indistinguishable, and in small samples where the hypothesized structure is correct or nearly correct the shrinkage estimators can offer substantial improvements with reductions in risk for estimating Σ and β as large as 70%.

The simulations raise concern about the instability of the rotation estimator. Additional exploratory work showed that the logit of the sample Givens angles are slow to approach normality so the estimator does not work well. As a result, we believe it is most prudent to shrink the eigenvalues but not the angles.

Ultimately, we recommend to first shrink the eigenvalues of the unstructured estimator and thus, increase its stability, by replacing it with the Stein estimator, and then to shrink the Stein estimator toward a structure using the structured log eigenvalue or correlation shrinkage estimator. In general, the structured log eigenvalue estimator tends to be more conservative than the correlation shrinkage estimator; i.e., when the structure is correct or close, the correlation shrinkage does a little better, but when the structure is far from correct, the correlation shrinkage does a little worse. The example in Section 6 illustrates the magnitude of differences that can be seen in $\hat{\beta}$ and its standard error in practice when using different estimators for Σ.

5 Asymptotic Properties of the Estimators

All of the estimators for the covariance matrix discussed in Sections 2 and 3 and applied in the simulations are consistent and the resulting estimators for the regression coefficients are consistent and asymptotically normal. A theorem appears below and a sketch of the proof with additional details on regularity conditions (assumptions) appears in the appendix.

Theorem: Under regularity conditions (assumptions 3-11) as specified in Magnus (1978) and under the additional assumption that all the eigenvalues of the true matrix, Σ are distinct, all the shrinkage estimators proposed in Sections 2-4 are consistent and the resulting estimators of the regression coefficients, β, are consistent and asymptotically efficient. The asymptotics in this case correspond to the number of of subjects, n, going to infinity with p, the dimension of Σ, fixed.

6 Example

We illustrate our approach on a study of sleep electroencephalogram (EEG) in healthy men and women (Carrier et al., 2000). One of the goals of this study was to determine the effect of age on EEG spectra in the absence of psychiatric disorders. For this analysis, EEG measurements were first transformed from the time domain to the frequency domain; these functions of frequency are call power spectra. The power spectra were then log transformed and finely discretized into 32 one hertz bins. For our analysis, we only included males, for a total of 53 subjects, and only used the first 24 one hertz bins. Thus, we have to estimate a 24×24 covariance matrix with a sample size of 53. Following Carrier et al. (2000), we model the mean of the log power spectra using an intercept and 9 spline basis functions. The effect of age was examined by including an interaction of the intercept and each basis function with age.

Covariance structure selection in SAS proc mixed using the BIC suggested an AR(1) structure fit. Below, we provide estimates, standard errors, and confidence regions for a subset of the regression coefficients for the unstructured estimator, for the AR(1) structured estimator, and for the structured log eigenvalue estimator, using the algorithm discussed in the introduction. We do not present the results for the correlation shrinkage estimator since for this example, the estimate was not positive definite.

Table 6 shows two of the estimated coefficients, the interactions of age with the first and ninth spline basis functions, and their standard errors. Important differences are seen in the estimates and their standard errors. In particular, when the structured log eigenvalue estimator is used, both interactions are highly significant; for the interaction of age with the first spline basis function, this agrees with the AR(1) but not the unstructured model, while for the interaction of age with the ninth basis function it agrees with the unstructured model but not the AR(1).

Figure 1 shows the confidence regions for jointly estimating the aforementioned regression coefficients. The area of the region and its orientation are clearly changing when using the different estimators.

Because of the simulation results in section 4, we would recommend making inferences (and predictions) using the structured log eigenvalue estimators as we believe this estimator to produce estimates of β closer to the truth. So, in this case, we would conclude that the interaction of age and the first and ninth spline bases are significantly larger than zero, implying that these aspects of the EEG spectra do differ by age.

7 Discussion

We have proposed classes of estimators for a covariance matrix that shrink some functional of the matrix to obtain better properties. The structured log eigenvalue, rotation, and correlation shrinkage estimators offer a compromise between completely unstructured and structured estimators. The Stein and log eigenvalue posterior mean estimators attempt to overcome the distortion of the eigenvalues of the unstructured estimators. Substantial gains were seen in estimating the matrix while gains were also seen in the mean squared error of estimators of the regression coefficients. In both situations, we obtain asymptotically optimal estimators. Because they are both effective and easy to compute using simple macros in standard statistical software, we would recommend the Stein, structured log eigenvalue (8), or correlation shrinkage (4, 6) estimators in practice.

These results extend and generalize those in Daniels and Kass (1999) and show that these simpler estimators can be competitive with the more computationally intensive estimators discussed by Daniels and Kass (1999). The accuracy of the asymptotics in small samples is a limiting factor in the gains produced by these new estimators. The approach in Daniels and Kass (1999) produces estimates which do not rely on any asymptotic approximations. The accuracy of the approximations in our approach might be assessed by comparing the inverse of the observed information matrix to an estimate of variability generated by a bootstrap approach. Clearly these shrinkage estimators are superior to the unstructured estimator in small to medium sized samples and superior to the structured covariance estimator when the hypothesized structure is incorrect. When a particular covariance structure is plausible and may be close to correct, but there are doubts, these estimators allow the data to be informative.

We plan on extending this work in a variety of directions. We are pursuing an alternative parameterization with which to shrink a covariance matrix toward a structure that does not require any asymptotic approximations and should be computationally efficient, as suggested by Daniels and Pourahmadi (2000). We also want to extend this work to estimating a random effects covariance matrix in two-level models and to assessing the subsequent impact on estimates of random effects. The asymptotics needed to form the approximate shrinkage estimators need to be worked out, especially for generalized linear random effects models. In addition, accounting for the estimation of the covariance matrix when estimating standard errors of random effects might be implemented using the approach discussed in Kass and Steffey (1989).

These methods can be extended to the case of Generalized Estimating Equations (GEE's) for longitudinal data with common measurement times across subjects. In this case, since we often think of a working correlation structure, we could focus on the estimator that shrinks the correlations. In future work we intend to examine the operating characteristics of such estimators for use with binary and count data.

Appendix

Part 1: Asymptotic distribution of the Givens angles

Here, we will derive an expression for the asymptotic variance of the Givens angles. First, define $vech(\Sigma)$ to be the vector composed of the lower triangular elements of the covariance matrix and define the spectral composition of Σ to be $\Sigma = O^t \Lambda O$. We can derive the general form for the information matrix of the Givens angles from the information matrix for $vech(\Sigma)$, $I(vech(\Sigma)) = (1/2)U^T(\Sigma^{-1} \otimes \Sigma^{-1})U$, where U is the gradient vector correponding to the derivative of Σ with respect to the lower triangular elements of Σ, $vech(\Sigma)$ and is of dimension $p(p+1)/2$. We replace U by G, where G is the gradient vector corresponding to the derivative of $vech(\Sigma)$ with respect to the eigenvalues and the

Givens angles. Based on results in Yang and Berger (1994), the information matrix for $vech(\Sigma)$, parameterized in terms of the eigenvalues and Givens angles is block diagonal with the upper block a known form, so we only need to compute the lower block. Using the identity $tr(ABCE) = (vech(E)^t)^t(C^t \otimes A)(vech(B))$, the diagonal elements corresponding to the Givens angles take the form $\frac{1}{2}b_{ij}^t(\Sigma^{-1} \otimes \Sigma^{-1})b_{ij}$ and the off-diagonal take the form $\frac{1}{2}b_{ij}^t(\Sigma^{-1} \otimes \Sigma^{-1})b_{rs}$ where $b_{ij} = \frac{dvech(\Sigma)}{d\theta ij}$. So, we have

$$\begin{aligned} b_{ij}^t(\Sigma^{-1} \otimes \Sigma^{-1})b_{ij} &= tr(\Sigma^{-1}d(O^t\Lambda O)/d\theta_{ij}\Sigma^{-1}d(O^t\Lambda O)/d\theta_{ij}) \\ &= -2 + 2tr(dO/d\theta ij O^t\Lambda^{-1}OdO/d\theta ij\Lambda) + tr(Od(O^t)/d\theta_{ij}d(O^t)/d\theta_{ij}) \end{aligned}$$

Using a similar derivation, the off-diagonal elements simplify to:

$$b_{ij}^t(\Sigma^{-1} \otimes \Sigma^{-1})b_{rs} = 2tr(dO/d\theta rs O^t\Lambda^{-1}OdO/d\theta ij\Lambda) + tr(OdO^t/d\theta_{rs}dO^t/d\theta_{ij}).$$

For the logit of θ_{ij}, we can multiply the information matrix by a diagonal matrix with elements $\frac{(\pi/2+\theta)(\pi/2-\theta)}{\pi}$.

When $\theta_{ij} = 0$ for all i, j, the information matrix simplifies to a diagonal matrix with diagonal element corresponding to θ_{ij}, $\frac{(\lambda_i-\lambda_j)^2}{\lambda_i\lambda_j}$. When considering the logit of the angles, the diagonal element is $\frac{\pi^2(\lambda_i-\lambda_j)^2}{16\lambda_i\lambda_j}$. As a consequence, this suggests a choice of the constant when using the uniform shrinkage prior for τ^2 in these models. We might choose the harmonic mean of the variances of the logits of θ_{ij}'s, evaluated at $\theta = 0$ (see e.g., Daniels, 1999 or Christiansen and Morris, 1997 for a discussion of choosing these constants). This form also illustrates a problem when any of the λ are equal: the variance is infinite, since the θ_{ij} are no longer unique. For example, it is known that there is not a unique rotation matrix for the identity matrix.

Part 2: Sketch of proof of theorem in Section 5

We give a brief outline of the proof here by proceeding through the six steps given in Section 1 to compute the shrinkage estimator for the covariance matrix and the resulting estimator for the regression coefficients. The main assumptions required in the theorem include the design matrix being full rank and $n > p$ and several assumptions regarding either the continuity or uniform convergence, or both, of functions of the design matrix (X), Σ^{-1}, and its gradient and Hessian. We refer the reader to Magnus (1978) for specific details.

- Steps 1 and 3: By Theorem 5 in Magnus (1978), the maximum likelihood estimates of Σ are consistent.

- Step 2: Using the following theorem: If X_n (vector) converges in probability to X (vector) and g is continuous, then $g(X_n)$ converges in probability to $g(X)$. (e.g., Serfling, 1980, p. 24). Given the consistency of the estimator for Σ, through application of this theorem, and under the condition that all the eigenvalues of the true matrix are distinct, the Stein estimator is consistent. This is clear from the form of the Stein estimator on p. 5.

- Step 4: The estimators for the parameters that determine the amount of shrinkage, τ^2, all converge to a constant.

- Step 5: Under the assumptions specified in Theorem 5 in Magnus (1978), with the additional assumption of distinct eigenvalues for the rotation shrinkage estimator, and by application of the convergence in probability theorem, $\hat{\Sigma}_{sh}$ converges in probability to Σ. The form of the estimators $\hat{\Sigma}_{sh}$ are the standard form seen in Empirical Bayes and as with the Stein estimator, the consistency is clear.

- step 6: Through application of the corollary to the Theorem 4 in Magnus (1978), $\hat{\beta}$ will be consistent and asymptotically efficient.

Part 3: Matrix IIR_1 and IIR_2 of Section 4.1 and matrix VI-A2(B1) of Sections 4.1 and 4.2 represented in terms of correlations

Correlation Matrix IIR_1:

1.000				
-.871	1.000			
-.223	.097	1.000		
-.357	.165	.651	1.000	
-.339	- .054	.013	.590	1.000

Correlation Matrix IIR_2:

1.000				
-.146	1.000			
.358	.306	1.000		
.598	.535	.245	1.000	
.593	.217	-.269	.608	1.000

Correlation Matrix VI-A2:

1.000				
-0.286	1.000			
-0.138	0.101	1.000		
-0.087	0.113	0.233	1.000	
-0.149	-0.022	0.087	0.296	1.000

Acknowledgments

Research of the first author was partially supported by National Institute of Health grant:

1-R01-CA85295-01A1; reseach of the second author, by National Institute of Health grant: 1-R01-CA54852-08. We would like to thank Yan Zhao and Jason Sinwell for coding and runnning the simulations and Dr. Stephanie Land for providing us with the data.

References

Barnard J, McCulloch R, Meng X. (2000) A Natural Strategy for modeling covariance matrices with Application to shrinkage. To appear in *Statistica Sinica.*

Carrier J., Land, S., Buyse, D.J., Kupfer, D.J., Monk, T.H. (2000) Age effects of sleep EEG power spectral density, To appear in *Psychophysiology.*

Chen, C-F. (1979) Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *Journal of the Royal Statistical Society, B*, 41:235-248.

Christiansen CL, Morris, CN (1997) Hierarchical Poisson regression modeling. *Journal of the American Statistical Association.* 92:618-632.

Daniels, MJ (1999) A prior for the variance in hierarchical models. *Canadian Journal of Statistics*, 27:569-580.

Daniels, M.J, Cressie, N.A.C. (2001) A hierarchical approach to covariance function estimation for time series. *Journal of Time Series Analysis*, 22:253-266.

Daniels, MJ, Kass RE (1999) Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, 1254-1263.

Daniels, M.J, Pourahmadi, M. (2000) Dynamic Models and Bayesian Modelling of Covariance Matrices in Longitudinal Data. Submitted.

DerSimonian, R and Laird, N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7:177-188.

Dey DK, Srinivasan C. (1985) Estimation of a covariance matrix under Stein's loss. *Annals of Statistics*. 13:1581-1591.

Diggle, P., Liang, K-Y, Zeger, S. (1994) *Analysis of Longitudinal Data*, Oxford University Press.

Efron, B. and Morris, C. (1973) Stein's estimation rule and its competitors — An empirical Bayes approach. *Journal of the American Statistical Association*, 68:117-130.

Haff, L.R. (1991) The variational form of certain Bayes estimators. *Annals of Statistics*. 19:1163-1190.

Hoffman DK, Raffenetti RC, Ruedenberg K. (1972) Generalization of Euler angles to N-dimensional orthogonal matrices. *Journal of Mathematical Physics*. 13:528-533.

Huber, P.J. (1967) The behaviour of the maximum likelihood estimator under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, LeCam, L.M. and Neyman, J. editors, University of California Press, pp. 221-33.

Kass, R.E., Steffey, D. (1989) Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models), *Journal of the American Statistical Association*, 84:717-726.

James, W. and Stein, C. (1961) Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, LeCam, L.M. and Neyman, J. editors, University of California Press, pp. 361-379.

Ledoit, O (1996). A well-conditioned estimator for large dimensional covariance matrices. Working paper. Anderson Graduate School of Management, UCLA.

Lin SP, and Perlman MD (1985). A Monte Carlo comparison of four estimators for a covariance matrix. In *Multivariate Analysis 6* P.R. Krishnaiah, ed.) p.411-429. North Holland.

Magnus, J.R. (1978) Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics*, 7, 281-312.

SAS Institute Inc. (1999) SAS/STAT User's Guide, Version 7. SAS Institute Inc., Cary, NC.

Serfling, R.J. (1980) Approximation theorems of mathematical statistics. Wiley.

Stein C. (1975) Estimation of a covariance matrix. Rietz Lecture, 39th Annual meeting IMS. Atlanta, Georgia.

Strawderman, W.E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Annals of Mathematical Statistics*, 42, 385-388.

Yang R, Berger J.O. (1994) Estimation of a covariance matrix using the reference prior. *Annals of Statistics*. 22:1195-1211.

Zimmerman, D. and Nunez-Anton, V. (1997) Structured antedependence models for longitudinal data. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions*. Springer Lecture Notes in Statistics, No. 122, Ed. T.G. Gregoire et al., pp. 63-76. New York: Springer-Verlag.

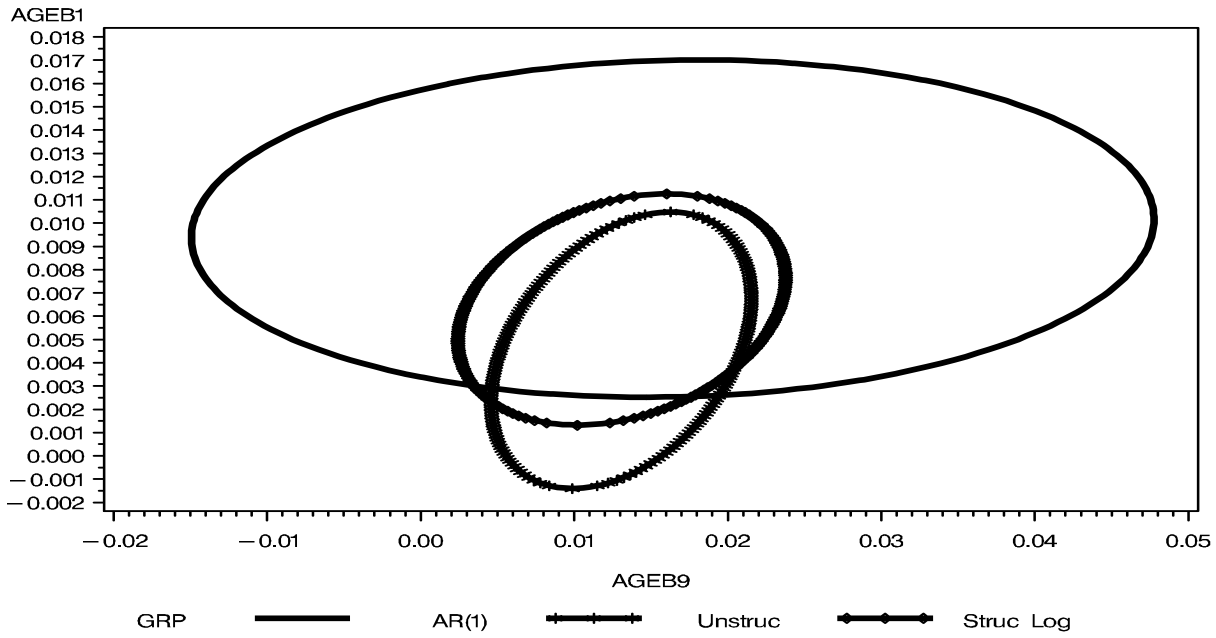

Figure 1: Confidence regions for the coefficients corresponding to the interaction of age and the first and ninth spline bases functions for the three models. AR(1): AR(1) structured estimator model, UN: Unstructured estimator model, Struc Log: Structured Log eigenvalue shrinkage estimator model

Σ	n	Ledoit	Stein	Log Eigen	Rotation	Corr
I	10	91	63	51	51	39
	20	90	67	64	64	51
	40	90	67	70	70	56
II	10	—	19	11	–	39
	20	—	17	7	–	51
	40	—	13	3	—	57
IIR_1	10	—	10	11	—	-118
	20	—-	16	6	—	-63
	40	—	13	4	—	-31
IIR_2	10	—	19	11	0	-19
	20	—	17	7	-4	-3
	40	—	13	4	-13	0
III	10	87	61	50	49	37
	20	82	61	59	56	43
	40	74	58	57	47	41

Table 1: Percentage reduction in average loss (PRIAL) for estimating Σ ($p = 5$) for samples of size 10, 20 and 40. Negative values indicate increases in risk. The dashes correspond to a PRIAL < -120. Details on the true Σ can be found in Section 4.1. Ledoit: Ledoit estimator, Stein: Stein estimator, Log Eigen: log eigenvalue posterior mean estimator, Rotation: rotation shrinkage estimator, Corr: correlation shrinkage estimator.

Σ	n	Ledoit	Stein	Log Eigen	Rotation	Corr
I	40	99	84	13	13	36
	100	99	76	18	18	68
II	40	—	24	4	—	35
	100	—	9	2	—	70
IIR_1	40	—	26	4	—	—
	100	—	9	2	—	—
IIR_2	40	—	25	4	-63	20
	100	—	9	2	—	16
III	40	90	81	12	2	31
	100	75	78	14	-17	45

Table 2: Percentage reduction in average loss (PRIAL) for estimating Σ ($p = 20$) for samples of size 40 and 100. Negative values indicate increases in risk. The dashes correspond to a PRIAL < -120. Details on the true Σ can be found in Section 4.1. Ledoit: Ledoit estimator, Stein: Stein estimator, Log Eigen: log eigenvalue posterior mean estimator, Rotation: rotation shrinkage estimator, Corr: correlation shrinkage estimator.

Σ	n	Stein	Struc Log	Rotation	Corr
I-A2	10	33	34	-52	57
	20	19	20	—	48
	50	7	15	—	38
II-A2	10	46	52	47	79
	20	34	33	18	73
	50	15	15	-33	62
III-A2	10	33	26	-83	-54
	20	22	13	—	-78
	50	9	5	—	-81
IV-A2	10	47	48	—	—
	20	46	46	—	—
	50	44	44	—	-
V-A2	10	26	26	—	—
	20	13	13	—	—
	50	3	3	—	—
VI-A2	10	52	64	61	65
	20	43	50	42	44
	50	24	24	4	3

Table 3: Percentage reduction in average loss (PRIAL) for estimating Σ ($p = 5$) for samples of size 10, 20, and 50. Negative values indicate increases in risk. The dashes correspond to a PRIAL < -120. Details on the true Σ can be found in Section 4.1. Stein: Stein estimator, Struc Log: structured log eigenvalue shrinkage estimator, Rotation: rotation shrinkage estimator, Corr: correlation shrinkage estimator.

Σ	n	Struc	Unstruc	Stein	Struc Log	Rotation	Corr
I-B1	10	.042	.085	.068	.063	.057	.050
	20	.021	.027	.024	.024	.025	.023
	40	.010	.011	.011	.011	.011	.010
	100	.0042	.0044	.0043	.0043	.0045	.0043
II-B1	10	.042	.060	.049	.046	.045	.041
	20	.021	.019	.018	.018	.019	.018
	40	.0104	.0087	.0085	.0084	.0091	.0083
	100	.0040	.0031	.0031	.0031	.0032	.0031
III-B1	10	.151	.193	.161	.152	.154	.138
	20	.073	.064	.061	.059	.068	.066
	40	.038	.028	.028	.027	.032	.028
	100	.015	.011	.011	.011	.012	.011
IV-B1	10	.0038	.0005	.0004	.0004	.0004	.0010
	20	.0020	.0002	.0001	.0001	.0001	.0005
	40	.0011	.0001	.0001	.0001	.0001	.0003
	100	.0004	.0000	.0000	.0000	.0000	.0001
V-B1	10	.0100	.00027	.00024	.00024	.0023	.0020
	20	.0053	.00009	.00009	.00009	.0022	.00062
	40	.0027	00004	.00004	.00004	.00038	.00020
	100	.00091	.00001	.00001	.00001	.00008	.00004
VI-B1	10	.010	.019	.015	.013	.013	.012
	20	.0052	0063	.0056	.0054	.0054	.0052
	40	.0025	.0026	.0025	.0025	.0025	.0025
	100	.0011	.0010	.0010	.0010	.0010	.0010

Table 4: Mean squared error (MSE) for estimating the regression coefficients for samples of size 10, 20, 40, and 100. Details on the true Σ can be found in Section 4.2. Struc: AR(1) structured estimator, Unstruc: Unstructured estimator, Stein: Stein estimator, Struc Log: structured log eigenvalue shrinkage estimator, Rotation: rotation shrinkage estimator, Corr: correlation shrinkage estimator.

Σ	n	Struc	Unstruc	Stein	Struc Log	Rotation	Corr
I-B2	20	.035	.082	.064	.059	.047	.042
	40	.016	.022	.020	.020	.021	.017
	100	.0053	.0055	.0054	.0054	.0069	.0053
II-B2	20	.063	.14	.11	.10	.075	.063
	40	.027	.033	.030	.030	.032	.029
	100	.0092	.0097	.0094	.0094	.0107	.0127
III-B2	20	.279	.162	.135	.130	.189	.132
	40	.141	.042	.040	.039	.096	.048
	100	.062	.014	.015	.015	.038	.018
IV-B2	20	.12	.024	.016	.016	.052	.036
	40	.051	.0065	.0054	.0054	.037	.027
	100	.017	.0018	.0017	.0017	.015	.0026
V-B2	20	.69	.026	.022	.022	.24	.12
	40	.265	.0071	.0071	.0071	.127	.024
	100	.089	.0017	.0017	.0017	.029	.0086
VI-B2	20	.0044	.0053	.0043	.0041	.0038	.0033
	40	.0020	.0014	.0013	.0013	.0015	.0013
	100	.00065	.00036	.00035	.00035	.00037	.00035

Table 5: Mean squared error (MSE) for estimating the regression coefficients for samples of size 20, 40, and 100. Details on the true Σ can be found in Section 4.2. For Matrix IV-B2 and V-B2, the MSE are multiplied by 100. Struc: AR(1) structured estimator, Unstruc: Unstructured estimator, Stein: Stein estimator, Struc Log: structured log eigenvalue shrinkage estimator, Rotation: rotation shrinkage estimator, Corr: correlation shrinkage estimator.

	AR(1)	Unstruc	Struc Log
age*B1	.0097 (.0029)	.0045 (.0024)	.0063 (.0020)
ratio	3.30	1.87	3.10
age*B9	.0164 (.0128)	.0131 (.0034)	.0131 (.0043)
ratio	1.28	3.83	3.02

Table 6: Estimates, standard errors, and the ratios of the estimate to the standard error for the regression coefficients corresponding to the interaction of age and the first (age*B1) and ninth (age*B9) spline bases functions under three covariance structure models. AR(1): AR(1) structured estimator, Unstruc: Unstructured estimator, Struc Log: structured log eigenvalue shrinkage estimator