

# An Empirical Bayesian Approach to Stein-Optimal Covariance Matrix Estimation\*

Ben Gillen  
California Institute of Technology

August 20, 2014

## Abstract

This paper proposes a conjugate Bayesian regression model to estimate the covariance matrix of a large number of securities. Characterizing the return generating process with an unrestricted factor model, prior beliefs impose structure while preserving estimator consistency. This framework accommodates economically-motivated prior beliefs and nests shrinkage covariance matrix estimators, providing a common model for their interpretation. Minimizing posterior finite-sample square error delivers a fully-automated covariance matrix estimator with beliefs that become diffuse as the sample grows relative to the dimension of the problem. In application, this Stein-optimal posterior covariance matrix performs well in a large set of simulation experiments.

---

\*Division of the Humanities and Social Sciences; MC 228-77; California Institute of Technology; Pasadena, CA 91125; bgillen@caltech.edu; tel: (626) 395-4061; fax: (626) 405-9841; This paper is taken from the third chapter of my doctoral thesis at the University of California, San Diego. I am grateful to Ayelen Banegas, Christian Brownlees, Gray Calhoun, Khai Chiong, Michael Ewens, Harry Markowitz, Alberto Rossi, Allan Timmermann, Michael Wolf, and Rossen Valkanov as well as participants in seminars at UC San Diego, UC Irvine, and the First Vienna Workshop on High Dimensional Time Series in Macroeconomics and Finance for helpful comments and discussion.

# 1 Introduction

In economic applications such as portfolio diversification and forecast combination, agent decisions depend upon a large covariance matrix summarizing the relationships between different returns or forecast errors. The sample size of the data available to the decision maker is typically quite limited relative to the dimensionality of the problem considered. As such, the unbiased sample covariance matrix estimator proves too imprecise to be practically useful in these applications, as its variance is magnified through an ill-posed optimization problem that yields highly unstable solutions.

The instability of the sample covariance matrix in portfolio diversification has been a long-studied topic since Markowitz (1952) first proposed the problem. Some of the first efforts to impose structure on the covariance matrix estimate itself through a restricted factor model were proposed in Sharpe (1963). Restricted factor models have evolved significantly since then to multi-factor models with a statistically defined number of potential factors in Connor & Korajczyk (1993) and Bai & Ng (2002).<sup>1</sup> A slightly different approach focuses on minimizing the finite-sample Stein (1955) mean square error, with a series of papers by Ledoit & Wolf (2003, 2004a,b) proposing shrinkage estimators that form a linear combination of the sample covariance matrix with a more structured model. This paper relates most directly to the shrinkage estimation strategy, presenting a Bayesian likelihood-based foundation of factor-based shrinkage models.

In parallel, a significant literature considers Bayesian analysis of the covariance matrix, anchored by the conjugate inverse-Wishart model to evaluate the sampling properties of the posterior covariance matrix.<sup>2</sup> While Yang & Berger (1994) present reference priors for the problem, a number of other researchers including Leonard & Hsu (1992) and Daniels & Kass (2001) have proposed informative priors that shrink the sample covariance matrix eigenvalues. Motivated by the difficulty interpreting the priors in these settings, a number of other papers seek to impose structure using clustering or a hierarchical Bayesian model, such as the analysis in Daniels & Kass (1999) and Liechty *et al.* (2004). Many of these techniques require MCMC simulation to characterize posterior expectations, a mechanism that can be computationally infeasible in extremely large models.

This paper builds on the Bayesian approach by analyzing posterior expectations for the covariance matrix in the natural conjugate setting with a standard Normal-Gamma data generating process. The statistical model represents the data generating process as a degenerate factor model,

---

<sup>1</sup>Fan *et al.* (2008) provide a theoretical foundation for establishing consistency of these estimators in sparse statistical models. Recent work, including Bickel & Levina (2008a,b); Lam & Fan (2009); Cai *et al.* (2010); Cai & Liu (2011); and Fan *et al.* (2011), extends the application of sparsity to derive regularization strategies for covariance matrix estimators.

<sup>2</sup>For examples, see Yang & Berger (1994) and Bensmail & Celeux (1996) for analyses based on the spectral decomposition of the matrix. Barnard *et al.* (2000) propose another approach, deriving informative priors for the covariance matrix in terms of its correlations and standard deviations. Liu (1993); Pinheiro & Bates (1996); and Pourahmadi (1999, 2000) can each be related to the Cholesky decomposition of the inverse of the covariance matrix, a device that is also used often in the analysis of sparse statistical models.

with a security's factor loadings determining its covariances with other assets. The factors are not the focus of inquiry in and of themselves, but rather only as a mechanism for characterizing the structure of the covariance matrix. For this reason, the analysis here treats the factors as fixed and observable, allowing for the number of factors to be potentially large. Conditional on these factors, I introduce an asymptotically-negligible perturbation of the likelihood for easily characterizing posterior expectations.

Prior beliefs on the factor loadings combine with the data to yield a structured, well-conditioned posterior expectation that remains consistent for the true covariance matrix. In the context where factors represent principal components of returns, I show the eigenvalues and eigenvectors of the sample covariance matrix, respectively, correspond to the variance of a factor and the associated vector of factor loadings across securities. Using this result, I show the posterior expected covariance matrix shrinks these eigenvectors toward their prior expectations and scales the corresponding eigenvalues to preserve orthonormality. This shrinkage representation is readily generalized, allowing the Bayesian framework I propose to nest any additive shrinkage estimator through empirically-determined priors.

As in Ledoit & Wolf (2004a), the shrinkage decomposition also facilitates deriving empirical prior beliefs to minimize finite-sample expected loss. Subject to a bandwidth parameter that can be effectively chosen via a simulated optimization algorithm, the Stein-optimal posterior covariance matrix is fully automated and easily implemented. This automation forgoes specifying a particular shrinkage target as the model for prior beliefs and allows for more robust performance of the posterior covariance matrix across a variety of settings. Recently, Ledoit & Wolf (2012) and Ledoit & Wolf (2013) have analyzed the nonlinear regularization of the eigenvalues for covariance matrices under different loss functions. Further, Bai & Liao (2012) consider the problem of extracting the principal components themselves in large problems. The exercise here considers a rather simpler question, focusing on solving for the optimal shrinkage under Frobenius loss proposed in Ledoit & Wolf (2004a) in a more flexible class of estimators, allowing for purely data-driven posterior regularization.

In application, the additional flexibility allows the Stein-optimal posterior estimator to perform effectively in a wider variety of settings than any of the individual methodologies presented in Ledoit & Wolf (2004a). Both in terms of mean-square error and in a portfolio optimization exercise, I show the Stein-optimal posterior performs as well as any currently available estimator and often performs better in a battery of simulation experiments. Though a given shrinkage estimator may perform better for specific data generating process, this performance may not prove to extend to other settings. Aggregating across a variety of asset universes, the stability of the Stein-optimal posterior's performance places it among the best estimators available in analyzing the covariances of returns in a large set of assets.

## 2 Statistical Model

This section develops the statistical model and derives posterior expectations for covariance matrices in a natural conjugate setting. The key innovation here lies in representing the sample covariance matrix as an unrestricted  $N$ -factor model, using prior beliefs in a structured factor model to impose structure in the posterior expectation of the covariance matrix.

The objective is to estimate the covariance matrix for the returns on  $N$  securities,  $r_{.,t} = [r_{1,t}, \dots, r_{N,t}]'$ , each of which are normally distributed with known means  $\mu = [\mu_1, \dots, \mu_N]'$  and an unknown covariance matrix  $\Sigma$ . To represent these returns in a linear model, assume that there are  $K$  observed factors  $F_{1,t}, \dots, F_{K,t}$  that represent all sources of variance across the securities and that these factors have known covariance matrix  $\Gamma$ . As the analysis focuses on the properties of covariance matrix estimators given a set of factors, I treat the factors as fixed and observable and ignore issues related to model identification. For example, these could correspond to the full set of derived principal components, with  $K = N$ , though the present analysis ignores any estimation error in deriving these factors or recovering their covariance matrix.

**Assumption 1** *The return generating process for returns satisfies the following conditions:*

- (a)  $r_{.,t} \sim N(\mu, \Sigma)$ , where  $\mu$  is known but  $\Sigma$  is unknown.
- (b)  $F_{.,t} = [F_{1,t}, \dots, F_{K,t}]' \sim N(\mu_F, \Gamma)$ , with both  $\mu_F$  and  $\Gamma$  known.
- (c)  $r_{i,\cdot} = [r_{i,1}, \dots, r_{i,T}]' \in \mathcal{S}(F)$ , the column space of the matrix  $F = [F'_{.,1}, \dots, F'_{.,T}]$  for all  $i$ .

Given Assumption 1, the return generating process for asset  $i$  in period  $t$  can be written as:

$$r_{i,t} = \alpha_i + \sum_{k=1}^K \beta_{i,k} F_{k,t} = \alpha_i + \beta'_{i,\cdot} F_{.,t} \quad (1)$$

In this return generating process, the vector  $\beta_{i,\cdot} = [\beta_{i,1}, \dots, \beta_{i,K}]'$  represents the factor loadings for asset  $i$ . Since the returns for asset  $i$  are fully explained by the set of factors, there is no idiosyncratic variation in the return generating process. Consequently, estimating  $\Sigma$  is equivalent to estimating these factor loadings.

Now consider a perturbation of the return generating process in which idiosyncratic noise is added to asset  $i$ 's return series after the factors have been extracted. Denote this white noise by  $\epsilon_{i,t}$ , which has a non-degenerate normal distribution with mean zero and idiosyncratic variance  $\sigma_{\epsilon,i}^2$ . This additional white noise is necessary to ensure the likelihood is well-behaved when conditioning on the factors  $F$ . In an analogy to the likelihood for nonparametric regression,  $\sigma_{\epsilon,i}^2$  can be interpreted

as a bandwidth parameter for the estimator.

$$\tilde{r}_{i,t} = \alpha_i + \beta'_{i,\cdot} F_{i,t} + \epsilon_{i,t} \quad (2)$$

The unrestricted covariance matrix implied by equation 2's return generating process takes the usual diagonalizable form. Let  $B$  denote a matrix with the factor loadings for all securities,  $\Gamma$  be the covariance matrix for the factors, and  $\Lambda$  be a diagonal matrix of idiosyncratic variances:

$$\Sigma = B\Gamma B' + \Lambda, \text{ where,} \quad (3)$$

$$B = \begin{bmatrix} \beta'_{1,\cdot} \\ \beta'_{2,\cdot} \\ \vdots \\ \beta'_{N,\cdot} \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,K} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{K,1} & \gamma_{K,2} & \cdots & \gamma_{K,K} \end{bmatrix}, \text{ and, } \Lambda = \begin{bmatrix} \sigma_{\epsilon,1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{\epsilon,2}^2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_{\epsilon,N}^2 \end{bmatrix}.$$

Factor models impose structure on the covariance matrix by implicitly restricting a subset of the factor loadings (typically those associated with non-economic factors) in the return generating process to equal zero. The alternative to this threshold-type restriction frames the factor model as the prior belief within a Bayesian regression framework. Deferring discussion of specific priors to sections 3 and 4, for now it suffices to represent the investor's prior beliefs satisfy conjugacy:

$$\beta_{i,\cdot}, \sigma_{\epsilon,i}^2 \sim_{\text{prior}} NG(\underline{\beta}_{i,\cdot}, \underline{\Omega}_i, \underline{\nu}_i, \underline{s}_i^2) \quad (4)$$

Here "NG" refers to the conditionally independent normal-gamma distribution. That is,  $\beta_{i,\cdot}$  has a Normal prior with mean  $\underline{\beta}_{i,\cdot}$  and covariance matrix  $\sigma_{\epsilon,i}^2 \underline{\Omega}_i$  conditional on the idiosyncratic variance  $\sigma_{\epsilon,i}^2$ , which has a Gamma distribution with  $\underline{\nu}_i$  degrees of freedom and expectation  $\underline{s}_i^2$ .

Given  $T$  observations from the normal return generating process,  $r_{i,\cdot} = [r_{i,1}, \dots, r_{i,T}]'$ , the likelihood of the data for specific values of  $\beta_{i,\cdot}$  and  $\sigma_{\epsilon,i}^2$  is given by a conditional Normal-Gamma distribution. That is, the likelihood for the true  $\beta_{i,\cdot}$  corresponds to a normal distribution with expectation given by the OLS estimates of factor loadings,  $\hat{\beta}_{i,\cdot}$ , and covariance matrix  $\sigma_{\epsilon,i}^2 (F'F)^{-1}$  conditional on  $\sigma_{\epsilon,i}^2$ , which has an unconditional gamma distribution with  $T - N$  degrees of freedom and expectation  $s_i^2$ , which is the OLS-computed standard error of residuals.

$$p(r_{i,\cdot} | \beta_{i,\cdot}, \sigma_{\epsilon,i}^2) = N(\hat{\beta}_{i,\cdot}, \sigma_{\epsilon,i}^2 (F'F)^{-1}), \text{ and, } p(\sigma_{\epsilon,i}^2 | \beta_{i,\cdot}) = G(T - N, s_i^2) \quad (5)$$

Since  $s_i^2 = 0$  in the sample, this likelihood is not well-defined. The singularity occurs because the data is perfectly described by the model, an event that also arises in non-parametric regression. To address this overfitting, consider the likelihood of the perturbed return generating process,

introducing noise to each security's return that prevents the factors from perfectly explaining each asset's return. The variance of this noise,  $\frac{h^2}{T}$ , can be interpreted as the bandwidth of the covariance matrix estimator and is scaled by the sample size to ensure estimator consistency. The likelihood for the perturbed model is then:

$$p(\tilde{r}_{i,\cdot} | \beta_{i,\cdot}, \sigma_{\epsilon,i}^2) = N(\hat{\beta}_{i,\cdot}, \sigma_{\epsilon,i}^2 (F'F)^{-1}), \text{ and, } p(\sigma_{\epsilon,i}^2 | \beta_{i,\cdot}) = G\left(T - N, s_i^2 + \frac{h^2}{T}\right) \quad (6)$$

With this likelihood, the prior and likelihood are natural conjugates, yielding analytical posterior expectations for each asset's factor loadings in closed-form. From textbook treatments on Bayesian econometrics such as Koop (2003) or Geweke (2005), the posterior expected factor loadings are the matrix-weighted average of prior expectations and the OLS estimated factor loadings:

$$\bar{\beta}_{i,\cdot} \equiv E_{post}[\beta_{i,\cdot}] = (\underline{\Omega}_i^{-1} + F'F)^{-1} \left( \underline{\Omega}_i \underline{\beta}_{i,\cdot} + F'F \hat{\beta}_{i,\cdot} \right) \quad (7)$$

Also, the posterior expected idiosyncratic variance ( $E_{post}[\sigma_{\epsilon,i}^2]$ , which is denoted  $\bar{s}_i^2$ ) is given by a weighted average of the prior expected idiosyncratic variance, the sample idiosyncratic variance, and a term that captures the disparity between the prior and OLS factor loadings:

$$(T + \underline{v}_i) \bar{s}_i^{*2} = \underline{v}_i s_i^2 + (T - N) \left( s_i^2 + \frac{h^2}{T} \right) + \left( \hat{\beta}_{i,\cdot} - \bar{\beta}_{i,\cdot} \right)' F'F \left( \hat{\beta}_{i,\cdot} - \bar{\beta}_{i,\cdot} \right) + \left( \underline{\beta}_{i,\cdot} - \bar{\beta}_{i,\cdot} \right)' \underline{\Omega}_i^{-1} \left( \underline{\beta}_{i,\cdot} - \bar{\beta}_{i,\cdot} \right) \quad (8)$$

Defining the matrices  $\bar{B}$  and  $\bar{\Lambda}$  as the posterior expectations for the matrices  $B$  and  $\Lambda$  defined above, the posterior expectation for the covariance matrix is:

$$\bar{\Sigma} = \bar{B}\Gamma\bar{B}' + \bar{\Lambda} \quad (9)$$

As is common with Bayesian estimators, as the amount of information in the data dwarfs the prior belief, the posterior expectation converges to the unbiased sample estimator. This convergence ensures that the estimator will be asymptotically consistent for the true covariance matrix.

**Proposition 1** *The posterior covariance matrix estimator is consistent:*

$$\text{p} \lim_{T \rightarrow \infty} \bar{\Sigma} = \Sigma \quad (10)$$

**Proof.** From equation 7, it's clear that  $\text{plim}_{T \rightarrow \infty} \bar{\beta} = \text{plim}_{T \rightarrow \infty} \hat{\beta}_{i,\cdot} = \beta_{i,\cdot}$ . This convergence implies that  $\text{plim}_{T \rightarrow \infty} \bar{B} = \text{plim}_{T \rightarrow \infty} \hat{B} = B$  and so, since  $\Gamma$  and  $\Lambda$  are known (the latter, given  $B$  and bandwidth  $h$ ), the result holds. ■

The model assumes that residuals are sampled independently over time, which is reasonably defensible in applications to data such as financial returns. The assumption could be relaxed to allow for autocorrelated residuals by adopting the sandwich covariance matrix in the likelihood. The Normal-Gamma conjugacy is more necessarily restrictive, as this property is essential to the analytical solutions for posterior expectations. A Harrison & West (1989) Dynamic Linear Model could move beyond conjugacy, allowing dynamic expectations and stochastic volatility and distributions with heavy tails.<sup>3</sup> Burda (2014) indicates a central practical challenge in such an extension is largely computational, since the estimation would require convergence to stationarity for a Markov Chain Monte Carlo sampler in extremely high-dimensions. From an analytical perspective, the lack of closed-form posteriors would make it difficult to characterize the optimal prior beliefs presented in section 5 beyond numerical solutions.

### 3 Empirical Bayesian Priors for Shrinkage Estimators

This section presents empirical Bayesian priors consistent with Ledoit & Wolf shrinkage estimators. I begin by decomposing the posterior expected covariance matrix into an additive factor structure, providing a shrinkage representation for posterior expectations. In section 5, this representation is useful in characterizing prior beliefs that yield admissible posterior expectations.

#### 3.1 A Shrinkage Representation of Posterior Expectations

To further characterize the properties of the posterior covariance matrix, consider the special case when factors and beliefs are orthogonal. Here  $\underline{\Omega}_i = \text{diag}(\omega_1^2, \dots, \omega_N^2)$  and  $\Gamma$  is a diagonal matrix with the  $k$ -th entry  $\sigma_{F_k}^2$ . With the cross-factor independence, equation 7 implies the  $k$ -th posterior expected posterior factor loading is a weighted average of the prior expected factor loading,  $\underline{\beta}_{i,k}$ , and the OLS-estimated factor loading,  $\hat{\beta}_{i,k}$ . Let  $\delta_k$  denote the weight assigned to the OLS-estimated factor  $k$  loading be defined as:

$$\delta_k = \frac{T\sigma_{F_k}^2}{\omega_k^{-2} + T\sigma_{F_k}^2} \quad (11)$$

These weights depend only on the total variation observed in the factor,  $T\sigma_{F_k}^2$  and prior variance  $\omega_k^2$ , so  $\delta_k$  is constant across all securities. Denote by  $\underline{B}_k$  the  $N \times 1$  vector of each asset's prior expected  $k$  factor loadings and let  $\hat{B}_k$  and  $\bar{B}_k$  be the vector of each asset's OLS-estimated and

---

<sup>3</sup>Since the empirical exercise here focuses on the static problem, such dynamic features are beyond the scope of the current analysis. One could introduce GARCH effects into the factors themselves as a conditional Bayesian extension of Alexander (2001) O-GARCH or Engle (2002) DCC-models. Voev (2008) considers shrinkage approaches based on O-GARCH. Other Bayesian approaches to dynamic factor models in asset allocation problems have been considered by Aguilar & West (2000), Ebner & Neumann (2008), and Zhou *et al.* (2014).

posterior expected  $k$  factor loadings. The cross-sectional posterior expected factor loadings are:

$$\bar{B}_k = (1 - \delta_k) \underline{B}_k + \delta_k \hat{B}_k \quad (12)$$

This formula for posterior expected factor loadings links the posterior covariance matrix with existing shrinkage estimators, allowing the posterior covariance matrix to be written as:

$$\Sigma^* = \bar{B} \Gamma \bar{B}' + \bar{\Lambda} = \sum_{k=1}^N \delta_k \sigma_{F_k}^2 \hat{B}_k \hat{B}_k' + \sum_{k=1}^N (1 - \delta_k) \sigma_{F_k}^2 \underline{B}_k \underline{B}_k' + \bar{\Lambda} \quad (13)$$

This decomposition provides an analytically useful device for deriving empirical prior beliefs consistent with shrinkage-based estimators. To illustrate this approach, I present the prior beliefs consistent with the Ledoit & Wolf (2004a) estimator. Appendix A2 extends this analysis to a case where the sample covariance matrix is shrunk towards any positive-semidefinite prior covariance matrix or even a linear combination of positive-semidefinite prior covariance matrices.

### 3.2 Empirical Bayesian Priors for Ledoit and Wolf Shrinkage

The Ledoit & Wolf (2004a) Single-Factor Shrinkage estimator is defined as a linear combination of the sample covariance matrix ( $\Sigma_S$ ) and the single-factor covariance matrix ( $\Sigma_{SF}$ ):

$$\begin{aligned} \Sigma_{LW}^* &= (1 - \delta) \Sigma_{SF} + \delta \Sigma_S \\ &= (1 - \delta) (B_{SF} \sigma_{SF}^2 B_{SF}' + \Lambda_{SF}) + \delta (B \Gamma B' + \Lambda) \end{aligned} \quad (14)$$

Here,  $B_{SF}$  denotes the vector of factor loadings for each asset in a restricted single-factor covariance matrix ( $\Sigma_{SF}$ ) with factor variance  $\sigma_{SF}^2$  and diagonal matrix of idiosyncratic variances  $\Lambda_{SF}$  and, as before,  $B$ ,  $\Gamma$ , and  $\Lambda$  represent the parameters of an  $N$  factor covariance matrix. Ledoit & Wolf (2004a) set the shrinkage intensity,  $\delta$ , to minimize the estimator's expected square error.

Relating equation 13 to 14 simply requires specifying prior beliefs so each factor's shrinkage coefficient,  $\delta_k$ , equals  $\delta$ . Let  $\hat{\beta}_{i,SF}$  be the single factor OLS parameter estimate for asset  $i$ , then:

**Proposition 2** *Suppose the likelihood of the data is given by equation 6 and an investor's prior belief is given by equation 4 with parameters:*

$$\underline{\beta}_{i,k} = \begin{cases} \hat{\beta}_{i,SF}, & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases}, \underline{\Omega}_{i,\{j,k\}} = \begin{cases} \frac{1-\delta}{\delta} T \hat{\sigma}_{F_k}^2, & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

*Then the posterior covariance matrix is given by the Ledoit and Wolf estimator in equation 14.*



Proposition 2’s proof is in appendix A1, with the only technical bit showing the priors for idiosyncratic variance are well-defined. The result illustrates how prior variances for a factor loading scale with the empirical variance of that factor so the shrinkage intensity will be constant across all factors. When  $N$  is fixed, Ledoit & Wolf (2003) show that the asymptotically optimal value of  $\delta$  behaves like a constant over  $T$ . Consequently, when  $\delta$  is chosen to minimize finite-sample expected loss,  $\omega_k$  grows as  $T$  becomes large and the priors implied by the optimal shrinkage become diffuse as the sample size itself grows. In this sense, the Ledoit and Wolf estimator converges to the sample covariance matrix faster than the posterior covariance matrix with fixed prior beliefs.

To this point, the model has abstracted from the problem of identifying the factors and their data generating processes. While the posterior analysis flexibly adapts to any factor specification, the equivalence result in proposition 2 relies on the factor structure embedded in Ledoit & Wolf (2004a) shrinkage. In particular, the single factor defining the shrinkage target must match the first of the  $N$  factors in the sample covariance matrix representation, which must also be orthogonal to the other factors in the model. This requirement is not too restrictive, since the factors can be defined so that it holds by construction.<sup>4</sup>

## 4 Economically Motivated Prior Belief Specifications

Beyond prior beliefs supporting shrinkage estimators, we may wish to consider other models for adding structure to covariance matrix estimation. This section presents two such models for prior beliefs based on economic intuition and empirical regularities in factor models.

### 4.1 Benchmark Driven Correlation Prior

To incorporate the structure of a  $K < N$  factor model of covariance, consider a prior that is diffuse over the first  $K$  factor loadings but shrinks the remaining  $N - K$  factor loadings toward zero. As a further simplifying assumption, assume the prior for each factor loading is independent of one another and that the prior standard deviation is constant for each of the remaining  $N - K$  factors.

$$\beta_{i,\cdot}, \sigma_{\epsilon,i}^2 \sim_{\text{prior}} NG \left( 0, \begin{bmatrix} \underline{\sigma}_\alpha^2 & 0 & 0 \\ 0 & \infty I_K & 0 \\ 0 & 0 & \underline{\sigma}_C^2 I_{N-K} \end{bmatrix}, \underline{\underline{\nu}}, \underline{\underline{s}}^2 \right) \quad (15)$$

This prior relates to Bayesian pricing models in Pastor (2000) and Pastor & Stambaugh (2002),

---

<sup>4</sup>For instance, suppose the shrinkage target uses an equal-weighted factor. Taking the equal-weighted factor as the first factor, orthogonalize security returns with respect to this equal weighted factor. From the orthogonalized returns, extract the remaining factors using principal components analysis. This basis of factors will satisfy the conditions for both Assumption 1 and Proposition 2.

modeling prior beliefs in a benchmark asset pricing model as diffuse over the factor loadings while shrinking the alphas toward zero. The variance parameter  $\underline{\sigma}_\alpha$  controls the extent to which assets' expected returns vary independently of the priced factors. Assuming the  $N - K$  derived factors have zero expected return, the present approach nests Pastor and Stambaugh's model as a special case where  $\underline{\sigma}_C^2 = 0$ .<sup>5</sup> The parameter  $\underline{\sigma}_C$  characterizes the amount of influence non-benchmark factors have in driving correlations, with larger values allowing posterior factor loadings for augmented factors to deviate further from zero. In the extreme case where  $\underline{\sigma}_C \rightarrow \infty$ , the extra-benchmark factor loadings become freely variable and the posterior covariance matrix converges to the unbiased sample estimate. Diffuse prior beliefs for the idiosyncratic variance set degree of freedom parameter  $\underline{\nu} \approx 0$  with any finite  $\underline{s}$ .

## 4.2 Mean Reverting Factor Loading Prior

A common approach to generating factor models extracts latent factors from the returns themselves, introducing potentially valuable information for prior beliefs about factor loadings. For instance, if a latent factor is defined by positive weights for each security, a zero prior expectation for that factor's loadings may be inappropriate. Define the cross-sectional average beta,  $\beta_0 = \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{i,\cdot}$ , and average idiosyncratic variance  $s_0^2 = \frac{1}{N} \sum_{i=1}^N \hat{s}_i^2$ , where  $\hat{\beta}_{i,\cdot}$  and  $\hat{s}_i^2$  denote the OLS-estimated factor loadings and residual variance for the  $i$ -th security, respectively. The Mean Reverting prior beliefs shrink the factor loadings for an individual security toward these grand means.

$$\beta_{i,\cdot}, \sigma_{\epsilon,i}^2 \sim_{\text{prior}} NG \left( \beta_0, \begin{bmatrix} \underline{\sigma}_\alpha^2 & 0 \\ 0 & \underline{\sigma}_C^2 I_N \end{bmatrix}, \underline{\nu}, s_0^2 \right) \quad (16)$$

This prior belief is rooted in Blume (1971, 1975)'s empirical observation that factor loadings exhibit mean reversion in the cross-section. As with the Benchmark Driven Correlation prior,  $\underline{\sigma}_C$  represents the degree to which the model allows for cross-sectional variation in factor loadings. As  $\underline{\sigma}_C \rightarrow 0$ , all factor loadings become identical and all covariances converge to a single constant. If  $\underline{\nu}$  becomes large, the idiosyncratic variances also converge to a constant. The limiting posterior covariance matrix is defined by two parameters with diagonal entries equaling the average variance and off-diagonal entries equaling the average covariance for all assets.

As an empirical Bayesian procedure, the Mean-Reverting prior fails the statistical assumption that the prior is independent of the likelihood. A more formal approach could follow hierarchical

---

<sup>5</sup>In papers that apply Pastor & Stambaugh (2002) to conditional settings, Avramov & Wermers (2006) and Banegas *et al.* (2013) utilize prior beliefs to limit the influence of macroeconomic factors to an investment's expected return. These approaches can also be nested in the current context by treating the macroeconomic factor as any other factor. Since the applications considered here focus solely on minimizing volatility, the level of expectations and specification for  $\sigma_\alpha$  is irrelevant. One appeal of the Benchmark Driven Correlation prior lies in its ability to conveniently nest this sort of flexible Bayesian pricing model without restricting non-benchmark correlations.

Bayesian approach motivated by Jones & Shanken (2005) where the cross-section is informative about an individual asset's factor loadings. In this sense, the pricing parameter  $\underline{\sigma}_\alpha$  measures the degree to which an investor believes individual fund alphas can vary from the grand mean alpha (for example, as in Frost & Savarino (1986)), with large values of  $\underline{\sigma}_\alpha$  allowing alphas to be effectively unrestricted in the cross-section.

## 5 The Stein-Optimal Posterior Covariance Matrix Expectation

This section derives optimal prior variance specifications for any fixed prior expected factor loading, extending the optimal shrinkage intensity analysis from Ledoit & Wolf (2003) to the current setting. The analysis builds on the shrinkage representation from equation 13, treating the shrinkage weights themselves as free parameters for tuning prior beliefs. After solving for the admissible shrinkage weights, proposition 5 provides a natural way to construct prior beliefs consistent with these weights.

### 5.1 Optimal Priors for Stein Loss

Consider optimal prior beliefs under the expected Frobenius Loss measure, which also corresponds to the loss function chosen by Ledoit & Wolf (2004a,b) in solving for optimal shrinkage intensities:

$$\mathcal{L} = \left\| \Sigma - \hat{\Sigma} \right\|^2 = \sum_{i=1}^N \sum_{j=1}^N (\sigma_{i,j} - \hat{\sigma}_{i,j})^2 \quad (17)$$

This loss function is a natural measure of mean-square error based on the  $\mathcal{L}_2$  norm for matrices, a common loss function for statistical problems. The optimization problem balances bias and variance from the shrinkage estimator in equation 13 to minimize the risk function:

$$\mathcal{R}(\delta_1, \delta_2, \dots, \delta_N) \equiv \quad (18)$$

$$E \left[ \left\| \Sigma - \sum_{k=1}^N \delta_k \sigma_{F_k}^2 \hat{B}_k \hat{B}_k' + \sum_{k=1}^N (1 - \delta_k) \sigma_{F_k}^2 \underline{B}_k \underline{B}_k' + \Lambda \right\|^2 \right]$$

Squared summations quickly become cumbersome, so denote the total prior bias and sample variance for the  $(k, l)$  entry of the covariance matrix as:

$$\mathcal{B}_{k,l} = \sum_{i=1}^N \sum_{j=1}^N E[(\beta_{i,k} \beta_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) (\beta_{i,l} \beta_{j,l} - \underline{\beta}_{i,l} \underline{\beta}_{j,l})]$$

$$\mathcal{V}_{k,l} = \sum_{i=1}^N \sum_{j=1}^N cov(\hat{\beta}_{i,k} \hat{\beta}_{j,k}, \hat{\beta}_{i,l} \hat{\beta}_{j,l})$$

This notation compactly expresses the optimal finite-sample shrinkage intensities (and consequently, the optimal empirical prior beliefs) in the following proposition.

**Proposition 3** *The risk function in equation 18 is minimized when  $\delta_1, \dots, \delta_N$  are chosen to equal the solution to the following set of  $N$  linear equations:*

$$\Psi\delta = \xi \tag{19}$$

where:

$$\begin{aligned} \xi_k &= \sum_{l=1}^N \sigma_{F,l}^2 \mathcal{B}_{k,l} \\ \Psi_{\{k,l\}} &= \sigma_{F,l}^2 (\mathcal{B}_{k,l} + \mathcal{V}_{k,l}) \end{aligned}$$

The formula in proposition 4, which is proved in appendix A1, captures the familiar tradeoff in shrinkage estimators between bias introduced by a misspecified model (represented by  $\xi$ ) with the total Mean Square Error of an estimator (reflected by  $\Psi$ ).

## 5.2 Feasible Estimation of Stein Optimal Priors

Feasibly implementing optimal priors a consistent estimate for the biases and covariances in  $\mathcal{B}_{k,l}$  and  $\mathcal{V}_{k,l}$ . Following the approach of Ledoit & Wolf (2003), the bias terms in  $\mathcal{B}_{k,l}$  can be consistently estimated by replacing population moments with unbiased sample moments and taking the difference between the estimated factor loadings and the prior expected factor loadings:

$$\hat{\mathcal{B}}_{k,l} = \sum_{q=1}^N \sum_{r=1}^N \left( \hat{\beta}_{q,k} \hat{\beta}_{r,k} - \underline{\beta}_{q,k} \underline{\beta}_{r,k} \right) \left( \hat{\beta}_{q,l} \hat{\beta}_{r,l} - \underline{\beta}_{q,l} \underline{\beta}_{r,l} \right) \tag{20}$$

For the covariance terms, note that  $\mathcal{V}_{k,l} = 0, \forall k \neq l$ , since the covariance between OLS-estimates of loadings on two orthogonal factors will always be zero. A bit of algebra reveals the sum defining  $\mathcal{V}_{k,k}$  to include  $N$  terms corresponding to the kurtosis of  $\hat{\beta}_{i,k}$  and  $N(N-1)$  terms corresponding to the product of the variances for  $\hat{\beta}_{i,k}$  and  $\hat{\beta}_{j,k}$ :

$$T\mathcal{V}_{k,k} = 3\sigma_{F_k}^{-4} \sum_{i=1}^N \sigma_{\epsilon,i}^4 + \sigma_{F_k}^{-4} \sum_{i=1}^N \sum_{j \neq i}^N \sigma_{\epsilon,i}^2 \sigma_{\epsilon,j}^2 \tag{21}$$

Substituting in the bandwidth parameter  $\frac{h}{\sqrt{T}}$  for  $\sigma_{\epsilon,i}$ , this reduces to:

$$T\mathcal{V}_{k,k} = \frac{(N^2 + 2N)h^4}{T^2\sigma_{F_k}^4} \quad (22)$$

A consistent and feasible estimator of optimal prior beliefs replaces the population moments in the equation above with sample moments. The consistency of this estimator for the beliefs follows immediately from the Continuous Mapping Theorem. Consistency of the posterior covariance matrix under optimal priors follows from the fact that the optimal shrinkage places all weight on the sample as the sample estimator becomes arbitrarily precise. The only free parameter remaining to be chosen is the bandwidth parameter  $h$ , which can be selected via a simulated optimization procedure described below in footnote 6.

### 5.3 Optimal Priors for Principal Factors

When the prior expected factor loadings are centered at zero and all factors correspond to principal components, the formula for optimal prior variances simplifies further. The zero prior allows rearranging the summation and multiplication in the definition of  $\mathcal{B}$  in equation 20. By the orthonormality of principal component factor weights, the terms characterizing estimator bias simplify:

$$\hat{\mathcal{B}}_{k,l}^* = \left( \sum_{q=1}^N \beta_{q,k} \beta_{q,l} \right) \left( \sum_{r=1}^N \beta_{r,k} \beta_{r,l} \right) = 1 \{k = l\}$$

Then the optimal shrinkage coefficients can be computed as the solution to the system of equations 19 with:

$$\xi_k^* = \sigma_{F_k}^2, \text{ and, } \Psi_{\{k,k\}}^* = \frac{(N^2 + 2N)h^4}{T^3\sigma_{F_k}^4} + \sigma_{F_k}^2$$

So that:

$$\delta_k^* = \frac{\xi_k^*}{\Psi_{\{k,k\}}^*} = \frac{T^3\sigma_{F_k}^6}{(N^2 + 2N)h^2 + T^3\sigma_{F_k}^6} \quad (23)$$

Equation 13 can then invert the shrinkage coefficients to solve for the implied optimal prior beliefs in the following result.

**Proposition 4** *Suppose the likelihood of the data is given by equation 6 and that prior expected factor loadings are fixed at zero in equation 4 with prior variance-covariance matrix,  $\underline{\Omega}$  so that:*

$$\beta_{i,k} = 0 \quad \underline{\Omega}_{i,\{k,k\}} = \frac{T^2\sigma_{F_k}^4}{(N^2 + 2N)h^2} \quad (24)$$

*If, in addition, the prior standard deviation parameters set posterior variances for each security to the sample variance, then the posterior expected covariance matrix minimizes finite sample mean square error in the class of all priors with expected factor loadings fixed at zero.*

This result characterizes the relationship between the number of observations, the number of securities, and the bandwidth of the estimator in determining the optimal shrinkage for each factor’s contribution to the covariance matrix. Note that the MSE optimal prior beliefs diverge at a rate faster than  $T$ , with the important property that beliefs become diffuse as the ratio  $T/N$  becomes large. This rate of convergence, which is an effect of the bandwidth specification, is compatible with the optimal convergence rates presented in Cai *et al.* (2010).

## 6 Monte Carlo Tests: Goodness of Fit and Portfolio Allocation

This section presents a battery of simulation tests that evaluate the finite sample performance of the proposed covariance matrix estimators. Table 1 summarizes the asset universes and estimators considered in the simulation. These tests calculate the sample covariance matrix from historical data on the returns for a number of securities, which serves as the “reference” covariance matrix for that asset universe. The simulation exercises generate a sample of mean-zero returns from these covariance matrices and then fits the different estimators to the simulated sample, allowing for direct comparison of the performance between these fitted estimates and the “true” covariance matrix.

### 6.1 Reference Data and Estimators

The first simulation test forms a set of reference covariance matrices corresponding to the sample covariance matrix estimated from 14 country portfolios, 25 Value-Size sorted portfolios, and 49 industry portfolios, where the return series are taken from Ken French’s website. These universes characterize how the estimators perform in asset allocation exercises at an asset class level in different contexts. Beyond these three universes, I consider a number of other security universes listed in table 1, the detailed results for which are available in an online appendix.

The second simulation test forms a random reference covariance matrix by selecting  $N$  stocks from the CRSP database. Specifically, for each year, as of January 1 of that year, I filter for all stocks in the CRSP database with complete 10 year histories of monthly returns. From these stocks, I randomly (and uniformly) select  $N$  stocks without replacement. Calculating the reference covariance matrix as the sample covariance matrix for these  $N$  stocks, I generate a single time series of normally distributed returns. As such, each simulation randomly selects a set of stocks to define the reference covariance matrix, and then performs a single monte carlo test with that

reference covariance matrix to evaluate the sampled loss. This test mirrors the treatment in Fan *et al.* (2012) and helps to characterize how the estimators perform in asset allocation exercises at the individual security level. I also consider generating random reference covariance matrices using a sample of European stock returns from DataStream and European Mutual Funds from Lipper, again reporting the detailed results in an online appendix.

Two non-Bayesian estimators provide reference points: the unrestricted sample covariance matrix and a single-factor model of covariance with an equal-weighted factor estimated using OLS. For Bayesian shrinkage implementations, Ledoit and Wolf have presented several priors representing different shrinkage targets. The shrinkage model for these simulation tests comes from Ledoit & Wolf (2004a), which shrinks the sample covariance matrix toward the Single Factor covariance matrix. I also fit several other Ledoit and Wolf estimators with different shrinkage targets, the detailed results for which are available in the web appendix.

In addition to the Stein-optimal posterior (SOP) covariance matrix estimator from equation 23, the simulations include estimators based on a Benchmark Driven Correlation prior (BDC) with a single factor and a mean-reverting (MR) prior specifications from section 4 equations 15 and 16. These priors are diffuse for idiosyncratic noise, so that the variance of the error term has prior degree of freedoms  $\nu_0 = 0$  and scale parameter  $s_0^2 = 0.01$ . The standard deviation of the prior is selected using a simulation technique to select the bandwidth for the Stein-optimal posterior covariance matrix, resulting in a fully-automated estimator.<sup>6</sup>

A brief summary of the tournament between all estimators with  $T = 25$  observations used to fit the estimator appears in Table 2. The best estimator is often one of Ledoit and Wolf’s specifications, justifying their widespread adoption. To focus on the performance of the SOP estimator, the central columns represent the potential improvement upon the SOP estimator by using the ex-post best alternative. In the samples for which the SOP estimator performs least well, an alternative can substantially reduce mean square error, though the typical improvement gain is often small. These models include four different shrinkage targets, with different specifications underperforming the Stein-optimal posterior covariance matrix in different investment universes. In terms of minimizing portfolio volatility, the improvements to using the ex-post best alternative estimator rarely exceed 25 basis points, an improvement that is usually less when the portfolio weights are constrained to be non-negative. Importantly, this comparison is the best *ex-post* improvement by choosing the best estimator *after* observing the simulation, not an *a priori* measure.

The rightmost columns of Table 2 compares the SOP estimator’s performance with the average

---

<sup>6</sup>Specifically, the algorithm pre-estimates the sample covariance matrix for the simulated data. Using that pre-estimate, the algorithm simulates 1,000 samples of returns. The bandwidth is then selected to minimize mean square error loss between the Bayesian Posterior estimators and the pre-estimated sample covariance matrix within this secondary simulation sample. As such, computing the Bayesian Posterior estimators required simulating 1,000 sets of returns to compute the bandwidth for each of the 1,200 simulations in the Monte Carlo study.

performance of the four Ledoit & Wolf estimators, giving a more balanced perspective of the SOP estimator’s relative performance. In some samples, the average performance of Ledoit & Wolf estimators still deliver some improvement in mean square error. However, with the exception of the Lipper sample of European Mutual Funds, the SOP’s estimated minimum variance portfolio delivers at least 45 basis points lower volatility than the average volatility from the minimum variance portfolios calculated using the Ledoit & Wolf estimators. For portfolio weights calculated with non-negativity constraints, the SOP’s minimum variance portfolio uniformly dominates the corresponding average Ledoit & Wolf estimator.

## 6.2 Finite Sample Goodness of Fit

To provide more details on estimator performance across sample sizes, Table 3 presents the finite-sample mean square error for each of the estimators at several horizons. The Stein-optimal posterior expectation and Ledoit and Wolf shrinkage estimators are consistently among the three best performing estimators in minimizing square error. The only other estimator that competes with these two is the posterior expected covariance matrix with a Mean-Reverting prior specification.

In these simulations, the bandwidth parameter for the Stein-optimal posterior expectation is adaptively determined using the simulation procedure described in footnote 6. The effectiveness of this approach in selecting the bandwidth is demonstrated in Table 4. While the estimator’s performance is stable for nearby bandwidth specifications, the simulation-optimized bandwidth performs better than any fixed model. Noting that the bandwidth represents the average idiosyncratic monthly volatility of a security in a very large factor model, the simulation-optimized average bandwidth around 1% is a fairly reasonable setting a priori. For very large bandwidths that assign almost all the variance of a security to idiosyncratic factors, all covariances converge to zero and, consequently, the Stein-optimal posterior expectation’s mean square error degrades.

As with the bandwidth in the Stein-optimal posterior expectation, the prior variance for the Factor Model and Mean Reverting models can also be adaptively determined using the proposed simulation algorithm. Table 5 evaluates the extent to which this tuning affects estimator performance for the Mean Reverting model finding that, while the estimator performs well across a variety of prior specifications, the simulation-optimizing approach sets the prior variance effectively. The simulation-optimized optimal prior behaves as expected with the number of observations and dimension of the covariance matrix, tightening when  $N/T$  is large and becoming diffuse as the sample size grows.



### 6.3 Performance in Optimal Portfolio Diversification

In financial applications, an estimator’s most relevant performance measure is the out-of-sample performance of the statistically estimated optimal portfolios. The simulation exercise evaluates this performance by calculating the minimum variance portfolio weights for the estimated covariance matrices and computing the volatility of that portfolio with the “true” reference covariance matrix. While this analysis doesn’t exactly match the dynamic features captured in portfolio backtesting, it does characterize the performance of the estimator in a static, myopic portfolio allocation setting. Here, the simulation approach provides a richer sampling environment and has been used in a number of portfolio evaluation studies, including Markowitz & Usmen (2003), Harvey *et al.* (2008), and Fan *et al.* (2008). To fix the problem, given a population of  $N$  securities with normally distributed returns having mean  $\mu$  and covariance matrix  $\Sigma$ , the objective is to select the portfolio weights  $w$  that maximize utility for an investor with risk aversion parameter  $\gamma$ :

$$U = w' \mu - \gamma w' \Sigma w \tag{25}$$

subject to  $\sum_{i=1}^N w_i = 1$

When Markowitz (1952) first proposed this problem, he recognized the problems of simply inputting sample estimates  $\hat{\mu}$  and  $\hat{\Sigma}$  into the optimization problem, suggesting additional constraints to help control portfolio exposures. Frost & Savarino (1988) illustrate well the benefits of hard constraints, while Jagannathan & Ma (2003) relate such non-negativity constraints to a shrinkage of the covariance matrix estimate.<sup>7</sup> To focus on the quality of the covariance matrix estimate, I focus on the sampling properties of the global minimum-variance portfolio weights, effectively maximizing 25 for an arbitrarily large risk aversion parameter  $\gamma$ . This exercise concentrates on the accuracy of the covariance matrix and its role in asset allocation.

As illustrated in the Table 2, it is not uncommon for another estimator to deliver portfolios with lower volatility than the Stein-optimal posterior expectation. However, the gains from selecting

---

<sup>7</sup>As illustrated by DeMiguel *et al.* (2007), naïve diversification rules often outperform statistically-optimal diversification. Britten-Jones (1999) and Okhrin & Schmid (2006) analytically solve for the distributional properties of optimal portfolio weights, underscoring their instable sampling properties. Other researchers have sought to perturb the decision problem itself for more stable sampling properties in the optimized weights. For example, Michaud (1998) proposes resampling the weights and Goldfarb & Iyengar (2003) considers robust optimization models. Several researchers have also considered incorporating Bayesian prior beliefs for the weights in estimating the inputs to the portfolio allocation process, an approach canonized by Black & Litterman (1992) and developed further by Kan & Zhou (2007), Chevrier & McCulloch (2008), Tu & Zhou (2010), and Avramov & Zhou (2010). Golosnoy & Okhrin (2009), Frahm & Memmel (2010), and Carrasco & Noumon (2012) present regularization techniques portfolio weights based on their sampling properties. DeMiguel *et al.* (2009) propose quadratic constraints on portfolio weights, which Fan *et al.* (2012) relate to a model of optimized portfolio weights in a statistically sparse risk model. I do not analyze these efforts separately, as the covariance matrix estimator proposed here could be incorporated into many of the algorithms.

the ex-post optimal covariance estimator are typically small, exceeding 60 basis points in only two samples. Table 6 provides a more detailed perspective of this property, reporting the volatility of estimated unconstrained minimum variance portfolios. Here, the Stein-optimal posterior continues to perform well, although not as uniformly as in terms of mean square error. In these samples, none of the other estimators deliver portfolios with more than 50 basis points of improvement in annualized volatility.

Interestingly, with the exception of portfolios based on the sample covariance matrix, almost every statistically-estimated portfolio outperforms both the completely naïve  $1/N$  portfolio diversification rule as well that weights all securities equally as the zero-correlation naïve  $1/V$  diversification rule that weights assets proportionally to the inverse of their variance. In some cases, notably those settings with a large number of securities, the difference can be over 10% in annualized volatility. As such, while naïve diversification may be particularly useful when evaluated using measures that incorporate portfolio average returns in addition to volatility, the benefits to optimal diversification do appear substantial and can be realized even with extremely small sample sizes.

Looking at minimum-variance portfolio weights restricted to long-only positions gives a similar, albeit slightly muted differential in portfolio performance. As seen in table 7, the maximal difference between models in annualized portfolio volatility for the asset-class universes is never greater than 55 basis points. Among the Bayesian models, the Ledoit and Wolf shrinkage model often performs best, although the differential in performance between this and the Stein-optimal posterior is never greater than 20 basis points. Overall, the Stein-optimal posterior delivers stable and effective low-volatility portfolio weights.

## 7 Conclusion

A Bayesian perspective of covariance matrix estimation provides a flexible mechanism for introducing structure into the estimation problem. The simple Stein-optimal posterior expectation proposed here is easily implemented, fully automated, and performs well in a variety of asset allocation problems while allowing a completely empirical specification of prior beliefs. The sampling properties of the covariance matrix estimate itself are remarkably stable across different environments. The optimized minimum variance portfolios dominate naïve diversification rules even in small samples and perform quite well compared to any other estimated covariance matrix in portfolio diversification exercises. As with shrinkage estimators, the Stein-optimal posterior expectation can be applied not only directly to the portfolio optimization problem, but also as a part of more technical approaches that still rely on the estimated covariances of asset returns.

## References

- Aguilar, O., & West, M. 2000. Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, **18**, 338–357.
- Alexander, Carol. 2001. Orthogonal GARCH. *In: Mastering Risk - Financial Times*, vol. 2. Prentice Hall.
- Avramov, Doran, & Zhou, Guofu. 2010. Bayesian Portfolio Analysis. *Annual Review of Financial Economics*, **2**, 25–47.
- Avramov, Doran, & Wermers, Russ. 2006. Investing in Mutual Funds when Returns are Predictable. *Journal of Financial Economics*, **81**, 339–377.
- Bai, Jushan, & Liao, Yuan. 2012. Efficient Estimation of Approximate Factor Models via Regularized Maximum Likelihood. *Mimeo*, January.
- Bai, Jushan, & Ng, Serena. 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica*, **70**(1), 191–221.
- Banegas, Ayelen, Gillen, Benjamin, Timmermann, Allan, & Wermers, Russ. 2013. The Cross-section of Conditional Mutual Fund Performance in European Stock Markets. *Journal of Financial Economics*, **108**(3), 699–726.
- Barnard, John, McCulloch, Robert, & Meng, Xiao-Li. 2000. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, **10**, 1281–1311.
- Bensmail, Halima, & Celeux, Gilles. 1996. Regularized Gaussian Discriminant Analysis through Eigenvalue Decomposition. *Journal of the American Statistical Association*, **91**, 1743–1748.
- Bickel, Peter, & Levina, Elizaveta. 2008a. Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics*, **36**, 199–227.
- Bickel, Peter J, & Levina, Elizaveta. 2008b. Covariance Regularization by Thresholding. *The Annals of Statistics*, **36**, 2577–2804.
- Black, Fischer, & Litterman, Robert. 1992. Global Portfolio Optimization. *Financial Analysts Journal*, **48**, 28.
- Blume, Marshall. 1971. On the Assessment of Risk. *The Journal of Finance*, **26**, 1–10.
- Blume, Marshall. 1975. Betas and Their Regression Tendencies. *Journal of Finance*, **30**(3), 785–795.
- Britten-Jones, Mark. 1999. The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. *Journal of Finance*, **54**, 655–671.
- Burda, Martin. 2014. Parallel Constrained Hamiltonian Monte Carlo for BEKK Model Comparison. *Advances in Econometrics*, **34**, Forthcoming.

- Cai, Tony, & Liu, Weidong. 2011. Adaptive Thresholding for Sparse Covariance Matrix Estimation. *Journal of the American Statistical Association*, **106**, 672–684.
- Cai, Tony, Harrison Zhang, Cun-Hui, & Zhou, Harrison. 2010. Optimal Rates of Convergence for Sparse Covariance Matrix Estimation. *Annals of Statistics*, **38**, 2118–2144.
- Carrasco, Marine, & Noumon, Nérée. 2012. Optimal Portfolio Selection using Regularization. *Mimeo*.
- Chevrier, Thomas, & McCulloch, Robert E. 2008. Using Economic Theory to Build Optimal Portfolios. *Mimeo*. Available at SSRN: <http://ssrn.com/abstract=1126596>.
- Connor, Gregory, & Korajczyk, Robert A. 1993. A Test for the Number of Factors in an Approximate Factor Model. *Journal of Finance*, **48**(4), 1263–91.
- Daniels, Michael J, & Kass, Robert E. 1999. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, **94**, 1254–1265.
- Daniels, Michael J, & Kass, Robert E. 2001. Shrinkage Estimators for Covariance Matrices. *Biometrics*, **57**, 1173–1184.
- DeMiguel, Victor, Garlappi, Lorenzo, & Uppal, Raman. 2007. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review Financial Studies*.
- DeMiguel, Victor, Garlappi, Lorenzo, Nogales, Francisco, & Uppal, Raman. 2009. A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *Management Science*, **55**, 798–812.
- Ebner, Markus, & Neumann, Thorsten. 2008. Time-varying factor models for equity portfolio construction. *European Journal of Finance*, **14**(5), 381–395.
- Engle, Robert. 2002. Dynamic Conditional Correlation. *Journal of Business and Economic Statistics*, **20**, 339–350.
- Fan, Jianqing, Fan, Yingying, & Lv, Jinchi. 2008. High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics*, **247**, 186–197.
- Fan, Jianqing, Liau, Yuan, & Mincheva, Martina. 2011. High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, **39**, 3320–3356.
- Fan, Jianqing, Zhang, Jingjin, & Yu, Ke. 2012. Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, **107**, 592–606.
- Frahm, Gabriel, & Memmel, Christoph. 2010. Dominating Estimators for Minimum-Variance Portfolios. *Journal of Econometrics*, **159**, 289–302.
- Frost, Peter A, & Savarino, James E. 1986. An Empirical Bayes Approach to Efficient Portfolio Selection. *Journal of Financial and Quantitative Analysis*, **21**(3), 293–305.

- Frost, Peter A, & Savarino, James E. 1988. For Better Performance - Constrain Portfolio Weights. *Journal of Portfolio Management*, **15**(1), 29–34.
- Geweke, John. 2005. *Contemporary Bayesian Econometrics and Statistics*. Wiley-Interscience.
- Goldfarb, D, & Iyengar, G. 2003. Robust portfolio selection problems. *Mathematics of Operations Research*, **28**(1), 1–38.
- Golosnoy, Vasyl, & Okhrin, Yarema. 2009. Flexible Shrinkage in Portfolio Selection. *Journal of Economic Dynamics and Control*, **33**, 317–328.
- Harrison, Jeff, & West, Mike. 1989. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag.
- Harvey, Campbell, Liechty, John, & Liechty, Merrill. 2008. Bayes vs. Resampling: A Rematch. *Journal of Investment Management*, **6**, 1–17.
- Jagannathan, R, & Ma, TS. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, **58**(4), 1651–1683.
- Jones, Christopher, & Shanken, Jay. 2005. Mutual fund performance with learning across funds. *Journal of Financial Economics*, **78**(3), 507–552.
- Kan, Raymond, & Zhou, Guofu. 2007. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, **42**(3), 621–656.
- Koop, Gary. 2003. *Bayesian Econometrics*. Wiley-Interscience.
- Lam, Clifford, & Fan, Jianqing. 2009. Sparsity and Rates of Convergence in Large Covariance Matrix Estimation. *Annals of Statistics*, **37**, 4254–4278.
- Ledoit, Olivier, & Wolf, Michael. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, **10**, 603–621.
- Ledoit, Olivier, & Wolf, Michael. 2004a. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, **88**(2), 365–411.
- Ledoit, Olivier, & Wolf, Michael. 2004b. Honey, I shrunk the sample covariance matrix - Problems in mean-variance optimization. *Journal of Portfolio Management*, **30**(4), 110–119.
- Ledoit, Olivier, & Wolf, Michael. 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, **40**(2), 1024–1060.
- Ledoit, Olivier, & Wolf, Michael. 2013. Optimal Estimation of a Large-Dimensional Covariance Matrix under Stein’s Loss. *Mimeo*.
- Leonard, Tom, & Hsu, John S.J. 1992. Bayesian Inference for a Covariance Matrix. *The Annals of Statistics*, **20**, 1669–1696.
- Liechty, John C, Liechty, Merrill W, & Muller, Peter. 2004. Bayesian Correlation Estimation. *Biometrika*, **91**, 1–14.

- Liu, Chuanhai. 1993. Bartlett's decomposition of the posterior distribution of the covariance for normal monotone ignorable missing data. *Journal of Multivariate Analysis*, **46**, 198–206.
- Markowitz, Harry. 1952. Portfolio Selection. *The Journal of Finance*, **7**(1), 77–91.
- Markowitz, Harry, & Usmen, Nilufer. 2003. Resampled Frontiers vs Diffuse Bayes: An Experiment. *Journal Of Investment Management*, **1**, 9–25.
- Michaud, Richard. 1998. *Efficient Assset Management: A Practial Guide to Stock Portfolio Optimization*. Oxford University Press.
- Okhrin, Yarema, & Schmid, Wolfgang. 2006. Distributional Properties of Portfolio Weights. *Journal of Econometrics*, **134**, 235–256.
- Pastor, Lubos. 2000. Portfolio selection and asset pricing models. *Journal of Finance*, **55**(1), 179–223.
- Pastor, Lubos, & Stambaugh, Robert. 2002. Investing in equity mutual funds. *Journal of Financial Economics*, **63**(3), 351–380.
- Pinheiro, Jose C, & Bates, Douglas M. 1996. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, **6**, 289–296.
- Pourahmadi, Mohsen. 1999. Joint mean-covariance models iwth applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, **86**, 677–690.
- Pourahmadi, Mohsen. 2000. Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, **87**, 425–435.
- Sharpe, William F. 1963. A Simplified Model for Portfolio Analysis. *Management Science*, **9**(2), 277–293.
- Stein, Charles. 1955. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*.
- Tu, Jun, & Zhou, Guofu. 2010. Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*, **45**, 959–986.
- Voev, Valeri. 2008. Dynamic Modelling of Large-Dimensional Covariance Matrices. *In: Bauwens, L., Pohlmeier, W., & Veredas, D. (eds), High Frequency Financial Econometrics: Recent Developments*. Physica-Verlag Rudolf Liebig GmbH.
- Yang, Rouoyong, & Berger, James O. 1994. Estimation of a Covariance Matrix Using the Reference Prior. *The Annals of Statistics*, **22**, 1195–1211.
- Zhou, X., Nakajima, J., & West, M. 2014. Dynamic dependent factor models: Improving forecasts and portfolio decisions in financial time series. *International Journal of Forecasting*, **30**(2012-09), 963–980. Under review at: International Journal of Forecasting.

## Tables

Table 1: Reference Return Universes and Estimators

This table lists the data samples used for the simulation samples as well as the estimators fitted to the simulated data. Data on returns for the reference securities were used to calculate sample covariance matrices that served as the “reference” covariance matrix. The simulation exercises generated a sample of mean-zero returns from these reference covariance matrices and then fit the estimators to the simulated sample, allowing comparison between these fitted estimates and the objective reference covariance matrix. Sample factor models are estimated using OLS with factors extracted via principal components analysis. Shrinkage estimators are all computed using the asymptotically optimal shrinkage intensity. Prior and bandwidth parameters for the Bayesian covariance matrices are determined using a simulated optimization procedure described in footnote 6.

<b>Reference Portfolios</b>	<b>N</b>	<b>Sample Estimators:</b>
Country	20	Sample Covariance Matrix (S)
Size & Book-to-Market	25	Single-Factor Covariance Matrix (1F)
Size & Momentum*	25	Three-Factor Covariance Matrix (3F)*
Size & Reversal*	25	Five-Factor Covariance Matrix (5F)*
Size & Long-Term Reversal*	25	
Global Size & Book to Market*	25	<b>Ledoit and Wolf Shrinkage Estimators:</b>
Global Size & Momentum*	25	Single-Factor (LWSF)
Industry*	30	Constant Correlation (LWCC)*
Industry	49	1 Parameter (LW1P)*
Size & Book-to-Market*	100	2 Parameter (LW2P)*
<b>Random Security Universe Samples</b>		<b>Bayesian Posterior Estimators:</b>
US Stocks (CRSP)		Stein-Optimal Posterior (SOP)
European Stocks (DataStream)*		Benchmark-Driven Correlation (BDC)
European Mutual Funds (Lipper)*		Mean-Reverting (MR)

(All Reference Portfolio Returns from Ken French’s Website.)

\* Detailed Results Reported in Appendix

Table 2: Summary of Simulation Tournament Results

The tournament generates 1,200 simulated data samples of  $T = 25$  normally-distributed, mean-zero returns with the reference covariance matrix, defined as the sample covariance matrix computed from historical return data for the security universes in table 1. The estimators are then fit to the simulated data. Mean Sq Error reports the mean square error between the fitted estimate and the reference covariance matrix. To evaluate portfolio selection, the Minimum Volatility exercise computes the variance-minimizing portfolio weights from the fitted covariance matrix and evaluates the true variance for this portfolio using the reference covariance matrix. The Min Constr Volatility follows the same approach, but imposes a non-negativity constraint when computing variance-minimizing portfolio weights. The Best Estimator when  $T = 25$  reports the estimator that “won” the tournament by minimizing loss, whether in terms of the Mean Sq Error or the estimated portfolio’s volatility, identified by the abbreviations in Table 1. The “Potential Improvement on SOP Loss” for mean square error reports the percent decrease in mean square error by utilizing the *ex-post* best estimator instead of the Stein-optimal posterior expectation (23). For the minimum variance measures, the “Potential Improvement on SOP Loss” reports the absolute decrease in annualized portfolio volatility by utilizing the ex-post best estimator instead of the Stein-optimal posterior expectation. The “SOP Improvement on Average LW Loss” provides a similar comparison of the performance of the Stein-optimal posterior expectation against the average performance of the four Ledoit & Wolf shrinkage estimators, reflecting the impact of uncertainty in defining the prior target for the shrinkage estimator.

Reference Portfolios	N	Best Estimator when $T = 25$			Potential Improvement on SOP Loss when $T = 25$			SOP Improvement on Average LW Loss when $T = 25$		
		Mean Sq Error	Minimum Volatility	Min Constr Volatility	Mean Sq Error	Minimum Volatility	Min Constr Volatility	Mean Sq Error	Minimum Volatility	Min Constr Volatility
Country	20	MR	SOP	MR	11%	-	-	4%	0.70	0.33
Size & Book-to-Market	25	LWCC	LW1P	1F	4%	0.25	0.18	-6%	1.24	0.75
Industry	49	MR	SOP	MR	9%	-	0.12	2%	0.89	0.48
Size & Momentum*	25	LW1P	LW1P	1F	5%	0.54	0.09	-6%	1.29	0.65
Size & Reversal*	25	LW1P	LW1P	1F	9%	0.42	0.15	-2%	0.72	0.46
Size & Long-Term Reversal*	25	LW1P	LW1P	1F	4%	-	0.06	-5%	1.19	0.62
Industry*	30	MR	SOP	LWCC	10%	0.55	0.14	0%	1.20	0.44
Global Size & Book-to-Market*	25	LW1P	LW1P	LW1P	4%	0.49	0.06	-5%	0.53	0.29
Global Size & Momentum*	25	LW1P	LW1P	3F	3%	0.51	0.04	1%	1.66	0.38
Size & Book-to-Market 10x10*	100	LWCC	SOP	MR	8%	-	0.24	1%	1.37	0.65
Individual Security Universe										
US Stocks (CRSP)	25	LWCC	SOP	SOP	34%	-	-	10%	0.52	0.29
	50	LWCC	SOP	LWCC	37%	-	0.01	12%	0.54	0.26
	100	LWCC	LWSF	LWCC	35%	0.09	0.04	8%	0.47	0.28
European Stocks* (DataStream)	25	LWCC	SOP	LWCC	47%	-	0.10	10%	0.55	0.41
	50	LWCC	SOP	LWCC	53%	-	0.29	12%	0.65	0.46
	100	LWCC	SOP	LWCC	58%	-	0.63	17%	0.65	0.39
European Mutual Funds* (Lipper)	25	LWCC	LW1P	LWSF	3%	0.98	0.01	-1%	(0.40)	0.24
	50	LWCC	LW1P	LWSF	5%	0.90	0.01	0%	(0.38)	0.28
	100	LW1P	LWSF	LWSF	6%	0.47	0.08	0%	0.27	0.28

\* Detailed Results Reported in Appendix



Table 3: Simulated Estimator Finite-Sample Mean Squared Error

This table presents the simulated mean squared error in estimating covariance matrices for several reference asset universes. Panel A reports results for 20 country portfolios, 25 sorted portfolios (on size and book-to-market), and 49 industry portfolios from Ken French’s website. Panel B reports results for a randomly drawn set of  $N$  stocks from within the CRSP database. The “reference” covariance matrix is defined as the sample covariance matrix computed from return data for the respective asset universe. For these reference covariance matrices, the simulation generates 1,200 simulated data samples of normally-distributed, mean-zero returns with a variable number of observations,  $T$ . The columns report the mean square error for the respective estimator models. The Single Factor covariance matrix is fitted using an equal-weighted factor. The Ledoit and Wolf Shrinkage matrix estimator uses a single factor target with the asymptotically optimal shrinkage intensity. The bandwidth parameter for the Stein-Optimal Posterior covariance matrix (23) and the prior parameters for the Benchmark Driven Correlation and Mean Reverting (Section 4) covariance matrices are chosen by using the simulated optimization procedure described in footnote 6.

Panel A: Portfolio Universe Reference Covariance Matrices							
$N$	$T$	Sample	Single Factor	Ledoit & Wolf Shrinkage	Stein Optimal Posterior	Bmk Driven Correlation	Mean Reverting
Panel A.1: 20 Country Portfolios							
20	25	43.42	43.86	41.46	41.40*	43.46	37.20**
	50	20.41	21.92	19.70*	19.94	20.52	19.69**
	100	10.54	12.93	10.36*	10.28**	10.54	10.58
	250	4.15	6.83	4.12*	4.10**	4.15	4.15
	500	2.02	4.71	2.01**	2.01*	2.02	2.02
Panel A.2: 25 Size & Value Sorted Portfolios							
25	25	148.26	151.08	147.62	147.21*	148.65	143.56**
	50	73.57	77.19	73.39*	73.37**	73.76	73.82
	100	36.83	40.46	36.75**	36.78*	36.87	36.89
	250	16.22	20.15	16.21**	16.22*	16.23	16.23
	500	7.39*	11.39	7.39**	7.40	7.40	7.40
Panel A.3: 49 Industry Portfolios							
49	25	303.13	289.70	275.76*	276.13	285.29	253.50**
	50	155.11	158.36	144.82*	145.94	149.79	142.07**
	100	75.56	86.69	72.04**	72.86*	74.39	73.49
	250	30.06	46.80	29.38**	29.43*	30.06	30.00
	500	15.18	33.31	15.00**	15.03*	15.18	15.18
Panel B: Individual Stock Reference Covariance Matrices							
$N$	$T$	Sample	Single Factor	Ledoit & Wolf Shrinkage	Stein Optimal Posterior	Bmk Driven Correlation	Mean Reverting
25	25	18.07	17.43	14.46	14.07*	17.99	11.81**
	50	8.89	10.01	7.56	7.52**	8.90	7.55*
	100	4.29	6.36	3.85**	3.90*	4.29	4.14
	250	1.70	4.25	1.61**	1.64*	1.70	1.70
	500	0.86	3.65	0.83**	0.85*	0.86	0.86
50	25	70.38	62.52	54.79	53.37*	60.38	44.65**
	50	33.80	34.89	27.65*	28.68	30.34	27.13**
	100	16.60	22.55	14.56**	15.02*	15.97	15.41
	250	6.61	15.36	6.19**	6.36*	6.61	6.59
	500	3.34	13.21	3.22**	3.29*	3.34	3.34
100	25	271.94	229.57	206.32	195.15*	222.16	170.76**
	50	133.64	133.48	108.47*	111.05	115.97	106.13**
	100	65.40	86.34	56.81**	59.63	61.41	58.77*
	250	26.61	60.65	24.89**	25.98*	26.46	26.26
	500	13.33	51.42	12.85**	13.23*	13.33	13.33

\*\* , \* Denote the best and second-best fitting models in a sample, respectively

Table 4: Bandwidth Sensitivity for Stein-Optimal Posterior Expectation

This table presents the simulated mean squared error in estimating covariance matrices using the Stein-Optimal Posterior expectation for variable bandwidths. Panel A reports results for 20 country portfolios, 25 sorted portfolios (on size and book-to-market), and 49 industry portfolios from Ken French’s website. Panel B reports results for a randomly drawn set of  $N$  stocks from within the CRSP database. The “reference” covariance matrix is defined as the sample covariance matrix computed from return data for the respective asset universe. For these reference covariance matrices, the simulation generates 1,200 simulated data samples of normally-distributed, mean-zero returns with a variable number of observations,  $T$ . The columns report the mean square error for the Stein-optimal posterior expectation defined in equation (23) using different bandwidths. When the bandwidth is determined by using the simulated optimization procedure described in footnote 6, the rightmost columns report the mean square error and average bandwidth ( $h$ ) for each sample.

Panel A: Portfolio Universe Reference Covariance Matrices								
$N$	$T$	Fixed Bandwidth Parameter					Optimized Bandwidth	
		0.05	0.25	0.50	1.00	2.00	Mean Sq Err	Avg $h$
Panel A.1: 20 Country Portfolios								
20	25	43.42	43.27	42.68	41.34	44.47	41.40	1.10
	50	20.41	20.39	20.27	19.92	21.27	19.94	1.13
	100	10.54	10.54	10.51	10.36	10.37	10.28	1.25
	250	4.15	4.15	4.14	4.13	4.11	4.10	1.46
	500	2.02	2.02	2.02	2.02	2.02	2.01	1.66
Panel A.2: 25 Size & Value Sorted Portfolios								
25	25	148.26	148.08	147.63	147.65	164.44	147.21	0.63
	50	73.57	73.53	73.40	73.41	79.01	73.37	0.60
	100	36.83	36.83	36.82	36.98	39.47	36.78	0.62
	250	16.22	16.22	16.22	16.24	16.70	16.22	0.69
	500	7.39	7.39	7.39	7.40	7.53	7.40	0.77
Panel A.3: 49 Industry Portfolios								
49	25	302.82	296.06	283.40	277.05	400.35	276.13	0.79
	50	155.08	153.39	149.12	146.09	200.97	145.94	0.79
	100	75.56	75.26	74.07	72.97	93.51	72.86	0.80
	250	30.06	30.03	29.84	29.35	33.01	29.43	0.89
	500	15.18	15.18	15.15	15.03	16.17	15.03	0.99
Panel B: Individual Stock Reference Covariance Matrices								
$N$	$T$	Fixed Bandwidth Parameter					Optimized Bandwidth	
		0.01	0.25	0.50	1.00	2.00	Mean Sq Err	Avg $h$
25	25	18.06	17.96	17.51	15.98	13.76	14.07	1.88
	50	8.89	8.88	8.79	8.35	7.45	7.52	1.91
	100	4.29	4.29	4.28	4.18	3.90	3.90	2.03
	250	1.70	1.70	1.70	1.69	1.65	1.64	2.30
	500	0.86	0.86	0.86	0.86	0.85	0.85	2.52
50	25	70.06	68.38	63.50	54.12	54.28	53.37	1.18
	50	33.76	33.41	32.04	28.88	30.39	28.68	1.11
	100	16.60	16.55	16.23	15.22	15.66	15.02	1.15
	250	6.61	6.61	6.58	6.42	6.56	6.36	1.28
	500	3.34	3.34	3.33	3.30	3.36	3.29	1.36
100	25	268.04	251.56	218.68	190.53	273.03	195.15	0.96
	50	132.72	127.91	116.60	106.75	160.16	111.05	0.80
	100	65.28	64.20	60.91	58.18	86.15	59.63	0.68
	250	26.61	26.50	26.02	25.68	34.89	25.98	0.52
	500	13.33	13.31	13.23	13.32	17.61	13.23	0.50

Table 5: Mean-Reverting Prior Sensitivity for Posterior Expectation

This table presents the simulated mean squared error in estimating covariance matrices using the posterior expected covariance matrix using the Mean Reverting Prior presented in section 4.2 under different prior variances. Panel A reports results for 20 country portfolios, 25 sorted portfolios (on size and book-to-market), and 49 industry portfolios from Ken French’s website. Panel B reports results for a randomly drawn set of  $N$  stocks from within the CRSP database. The “reference” covariance matrix is defined as the sample covariance matrix computed from return data for the respective asset universe. For these reference covariance matrices, the simulation generates 1,200 simulated data samples of normally-distributed, mean-zero returns with a variable number of observations,  $T$ . The columns report the mean square error for the Mean Reverting prior with different prior variances. When the prior variance is determined using the simulated optimization procedure described in footnote 6, the rightmost columns report the mean square error and average optimal prior variance for each portfolio universe.

Panel A: Portfolio Universe Reference Covariance Matrices								
$N$	$T$	Fixed Prior Standard Deviation					Optimized Prior	
		1.0%	5%	10%	25%	100%	Mean Sq Err	Avg $\sigma$
Panel A.1: 20 Country Portfolios								
20	25	39.85	34.60	37.14	40.95	43.22	37.20	10%
	50	25.36	18.50	19.29	19.85	20.39	19.69	18%
	100	19.00	11.44	11.05	10.56	10.55	10.58	31%
	250	12.39	5.69	4.66	4.20	4.15	4.15	150%
	500	8.50	3.17	2.29	2.04	2.02	2.02	149%
Panel A.2: 25 Size & Value Sorted Portfolios								
25	25	221.58	148.99	143.59	147.25	148.18	143.56	12%
	50	154.58	76.84	74.20	73.72	73.58	73.82	25%
	100	109.72	40.09	38.28	36.94	36.84	36.89	85%
	250	63.96	18.99	17.13	16.28	16.23	16.23	150%
	500	34.93	9.34	7.78	7.41	7.39	7.40	149%
Panel A.3: 49 Industry Portfolios								
49	25	288.27	227.82	253.50	280.49	300.74	253.50	10%
	50	209.15	136.56	142.07	148.74	154.72	142.07	10%
	100	148.96	76.75	73.44	74.12	75.53	73.49	10%
	250	87.80	36.05	30.75	30.01	30.06	30.00	21%
	500	54.65	19.42	15.78	15.22	15.18	15.18	141%
Panel B: Individual Stock Reference Covariance Matrices								
$N$	$T$	Fixed Prior Standard Deviation					Optimized Prior	
		1.0%	5%	10%	25%	100%	Mean Sq Err	Avg $\sigma$
25	25	14.94	11.19	11.65	15.02	17.85	11.81	10%
	50	12.40	7.44	7.27	8.18	8.87	7.55	14%
	100	10.61	4.95	4.25	4.20	4.29	4.14	18%
	250	8.30	2.80	1.94	1.71	1.70	1.70	1000%
	500	6.60	1.66	1.01	0.87	0.86	0.86	1000%
50	25	56.00	41.30	46.92	59.77	69.42	44.65	9%
	50	45.77	26.89	27.00	30.93	33.66	27.13	10%
	100	37.95	17.30	15.30	16.05	16.59	15.41	15%
	250	27.67	9.04	6.82	6.59	6.61	6.59	115%
	500	20.27	5.06	3.55	3.34	3.34	3.34	1000%
100	25	205.98	161.50	189.32	238.50	268.95	170.76	7%
	50	165.27	103.88	107.49	124.05	132.97	106.13	9%
	100	129.64	63.46	58.77	63.29	65.33	58.77	10%
	250	87.24	31.75	26.45	26.47	26.61	26.26	15%
	500	62.17	17.10	13.64	13.32	13.33	13.33	999%

Table 6: Out of Sample Volatility of Simulated Minimum Variance Portfolios

This table evaluates the performance of different estimators in a variance minimization exercise. Panel A reports results for 20 country portfolios, 25 sorted portfolios (on size and book-to-market), and 49 industry portfolios from Ken French’s website. Panel B reports results for a randomly drawn set of  $N$  stocks from within the CRSP database. The “reference” covariance matrix is defined as the sample covariance matrix computed from return data for the respective asset universe. For these reference covariance matrices, the simulation generates 1,200 simulated data samples of normally-distributed, mean-zero returns with a variable number of observations,  $T$ . The minimum volatility exercise computes the variance-minimizing portfolio weights from the fitted covariance matrix. The columns report the “true” volatility of these portfolios under the reference covariance matrix for the respective estimators. The Single Factor covariance matrix is fitted using an equal-weighted factor. The Ledoit and Wolf Shrinkage estimator uses a single factor prior with the asymptotically optimal shrinkage intensity. The bandwidth parameter for the Stein-Optimal Posterior covariance matrix (23) and the prior parameters for the Benchmark Driven Correlation and Mean Reverting (Section 4) covariance matrices are chosen by using the simulated optimization procedure described in footnote 6. The Benchmark Portfolios include the  $1/N$  portfolio, which equally weights all securities in the asset universe, and the  $1/V$  portfolio, which weights all securities proportionally to the inverse of their variance.

Panel A: Portfolio Universe Reference Covariance Matrices								
Benchmark Portfolios	$T$	Sample	Single Factor	Ledoit & Wolf Shrinkage	Stein Optimal Posterior	Bmk Driven Correlation	Mean Reverting	
<i>Panel A.1: 20 Country Portfolios</i>								
$1/N$	17.92	25	29.61	16.14	15.89*	15.70**	19.01	16.16
$1/V$	17.33	50	16.61	15.35	14.74**	14.74*	16.76	15.05
		100	14.47	14.86	14.01**	14.03*	14.53	14.30
		250	13.50	14.53	13.43**	13.46*	13.51	13.53
		500	13.23	14.38	13.21**	13.23*	13.23	13.24
<i>Panel A.2: 25 Size &amp; Value Sorted Portfolios</i>								
$1/N$	24.08	25	1363.46	19.56	17.53*	17.15**	21.00	19.77
$1/V$	22.36	50	19.86	19.85	16.34*	16.11**	19.25	16.40
		100	16.29	19.97	15.50*	15.36**	16.46	15.89
		250	14.88	20.12	14.76*	14.72**	14.93	14.94
		500	14.50	20.16	14.47*	14.46**	14.53	14.54
<i>Panel A.3: 49 Industry Portfolios</i>								
$1/N$	16.64	25	239.13	14.02	13.93*	13.65**	14.00	14.01
$1/V$	15.94	50	89.44	13.76	12.98*	12.87**	13.25	13.18
		100	14.90	13.78	12.23*	12.19**	12.62	12.44
		250	11.87	13.91	11.48**	11.49*	11.89	11.82
		500	11.21	13.95	11.11**	11.13*	11.22	11.23
Panel B: Individual Stock Reference Covariance Matrices								
Benchmark Portfolios	$T$	Sample	Single Factor	Ledoit & Wolf Shrinkage	Stein Optimal Posterior	Bmk Driven Correlation	Mean Reverting	
<i><math>N = 25</math></i>								
$1/N$	17.49	50	15.91	13.91	13.15*	13.08**	15.55	13.50
$1/V$	15.37	100	12.98	13.66	12.39**	12.39*	12.99	12.69
		250	11.94	13.57	11.86**	11.89*	11.95	11.95
		500	11.62	13.57	11.60**	11.62*	11.62	11.62
<i><math>N = 50</math></i>								
$1/N$	17.15	50	357.39	12.98	11.50**	11.56*	11.70	11.89
$1/V$	15.03	100	12.15	12.94	10.47**	10.61*	10.73	10.76
		250	9.59*	12.98	9.48**	9.66	9.59	9.63
		500	9.09**	13.07	9.10	9.24	9.09*	9.09
<i><math>N = 100</math></i>								
$1/N$	16.98	50	283.33	12.39	10.00**	10.24*	10.29	10.41
$1/V$	14.83	100	262.33	12.58	8.52**	8.87*	8.93	9.07
		250	5.85**	12.80	6.77*	7.02	6.99	7.00
		500	5.10**	12.94	5.87	6.10	5.10*	5.10

\*\* , \* Denote the best and second-best portfolio estimators in a sample, respectively

Table 7: Out of Sample Volatility of Minimum Variance Portfolios with Non-negativity Constraints

This table evaluates the performance of different estimators in a constrained variance minimization exercise. Panel A reports results for 20 country portfolios, 25 sorted portfolios (on size and book-to-market), and 49 industry portfolios from Ken French’s website. Panel B reports results for a randomly drawn set of  $N$  stocks from within the CRSP database. The “reference” covariance matrix is defined as the sample covariance matrix computed from return data for the respective asset universe. For these reference covariance matrices, the simulation generates 1,200 simulated data samples of normally-distributed, mean-zero returns with a variable number of observations,  $T$ . The constrained minimum volatility exercise computes the variance-minimizing portfolio weights from the fitted covariance matrix subject to non-negativity constraints on the portfolio weights. The columns report the “true” volatility of these portfolios under the reference covariance matrix for the respective estimators. The Single Factor covariance matrix is fitted using an equal-weighted factor. The Ledoit and Wolf Shrinkage estimator uses a single factor prior with the asymptotically optimal shrinkage intensity. The bandwidth parameter for the Stein-Optimal Posterior covariance matrix (23) and the prior parameters for the Benchmark Driven Correlation and Mean Reverting (Section 4) covariance matrices are chosen by using the simulated optimization procedure described in footnote 6. The Benchmark Portfolios include the  $1/N$  portfolio, which equally weights all securities in the asset universe, and the  $1/V$  portfolio, which weights all securities proportionally to the inverse of their variance.

Panel A: Portfolio Universe Reference Covariance Matrices								
Reference Portfolios	Fitted $T$	Sample	Single Factor	Ledoit & Wolf Shrinkage	Stein Optimal Posterior	Bmk Driven Correlation	Mean Reverting	
Panel A.1: 20 Country Portfolios								
$1/N$	17.92	25	16.32	16.17	16.18	16.15*	16.33	16.11**
$1/V$	17.33	50	15.70	15.61*	15.62	15.61**	15.71	15.67
		100	15.34	15.27**	15.30	15.29*	15.34	15.35
		250	15.11	15.09**	15.10	15.10*	15.11	15.11
		500	15.04	15.04	15.04*	15.04**	15.04	15.04
Panel A.2: 25 Size & Value Sorted Portfolios								
$1/N$	24.08	25	17.93	17.64**	17.79	17.79	17.90	17.65*
$1/V$	22.36	50	17.55	17.45*	17.50	17.49	17.54	17.45**
		100	17.39	17.38	17.38	17.37**	17.39	17.37*
		250	17.30	17.33	17.30*	17.30**	17.30	17.30
		500	17.28	17.32	17.28*	17.28**	17.28	17.28
Panel A.3: 49 Industry Portfolios								
$1/N$	16.64	25	14.42	13.89**	14.01	13.99	14.12	13.94*
$1/V$	15.94	50	13.55	13.35**	13.36	13.35	13.43	13.35*
		100	13.11	13.10	13.03*	13.02**	13.07	13.05
		250	12.82	12.96	12.81*	12.80**	12.82	12.83
		500	12.73	12.91	12.72*	12.72**	12.73	12.73
Panel B: Individual Stock Reference Covariance Matrices								
Reference Portfolios	Fitted $T$	Sample	Single Factor	Ledoit & Wolf Shrinkage	Stein Optimal Posterior	Bmk Driven Correlation	Mean Reverting	
$N = 25$		25	14.68	14.31	14.22*	14.16**	14.64	14.24
$1/N$	17.49	50	13.75	13.87	13.58*	13.57**	13.75	13.69
$1/V$	15.37	100	13.17	13.60	13.12**	13.12*	13.17	13.21
		250	12.90	13.52	12.88**	12.89*	12.90	12.90
		500	12.77*	13.51	12.77**	12.77	12.77	12.77
$N = 50$		25	13.97	13.53	13.37*	13.35**	13.55	13.47
$1/N$	17.15	50	12.88	13.18	12.67**	12.69*	12.80	12.81
$1/V$	15.03	100	12.22	13.00	12.16**	12.18*	12.24	12.24
		250	11.78*	12.94	11.77**	11.79	11.78	11.81
		500	11.66*	12.94	11.66**	11.67	11.66	11.66
$N = 100$		25	13.51	13.10	12.82*	12.77**	12.99	12.99
$1/N$	16.98	50	12.14	12.71	11.90**	11.95*	12.04	12.05
$1/V$	14.83	100	11.41	12.59	11.36**	11.40*	11.45	11.46
		250	10.93*	12.63	10.92**	10.94	10.96	10.96
		500	10.74*	12.62	10.74**	10.74	10.74	10.74

\*\* , \* Denote the best and second-best portfolio estimators in a sample, respectively

## A1: Proofs

### A1.1 Proof of Proposition 2

Suppose the likelihood of the data is given by equation 6 and an investor's prior belief is given by equation 4 with parameters:

$$\underline{\beta}_{i,k} = \begin{cases} \hat{\beta}_{SF}, & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases}, \underline{\Omega}_{i,\{j,k\}} = \begin{cases} \frac{1-\delta}{\delta} T \hat{\sigma}_{F_k}^2, & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

Then the posterior covariance matrix is given by the Ledoit and Wolf estimator in equation 14.

**Proof.** The proof for off-diagonal entries in the posterior covariance matrix follows directly from equation 13, which simplifies so that the weight assigned to the prior expected factor loadings is constant across factors and assets:

$$\bar{\Sigma} = \delta \sum_{k=1}^N \sigma_{F_k}^2 \hat{B}_k \hat{B}_k' + (1 - \delta) \sum_{k=1}^N \sigma_{F_k}^2 \underline{B}_k \underline{B}_k' + \Lambda$$

The proper specifications for  $\underline{s}$  and  $\underline{v}$  will set the matrix  $\Lambda = \delta \underline{\Lambda} + (1 - \delta) \hat{\Lambda}$  where  $\underline{\Lambda}$  is the diagonal matrix with  $(k, k)$  entry equal to the idiosyncratic variance estimated in the restricted single factor model and  $\hat{\Lambda}$  is the idiosyncratic variance in the unrestricted covariance matrix. This can be done by setting idiosyncratic beliefs so that:

$$\underline{v} = T\delta, \text{ and, } \underline{s}_i^2 = \hat{\sigma}_{\epsilon,i,SF}^2 - \frac{1}{T\delta} \left( \hat{\beta}_i - \underline{\beta}_i \right)' \left( \underline{\Omega}_i + (F'F)^{-1} \right)^{-1} \left( \hat{\beta}_i - \underline{\beta}_i \right)$$

This specification establishes the result:

$$\bar{\Sigma} = (1 - \delta) \left( \sigma_{F_1}^2 \hat{B}_1 \hat{B}_1' + \underline{\Lambda} \right) + \delta \left( \sum_{k=1}^N \sigma_{F_k}^2 \hat{B}_k \hat{B}_k' + \hat{\Lambda} \right) = (1 - \delta) \Sigma_{SF} + \delta \Sigma_S = \Sigma_{LW}^*$$

■

### A1.2 Proof of Proposition 3

The risk function in equation 18 is minimized when  $\delta_1, \dots, \delta_N$  are chosen to equal the solution to the following set of  $N$  linear equations:

$$\Psi \delta = \xi$$

where:

$$\xi_k = \sum_{q=1}^N \sigma_{F,q}^2 \mathcal{B}_{k,q}$$

$$\Psi_{\{k,l\}} = \sigma_{F,l}^2 (\mathcal{B}_{k,l} + \mathcal{V}_{k,l})$$

**Proof.** The mechanical details are somewhat tedious, but they simply involve taking the derivative of the risk function and quite a bit of rudimentary algebra pushing around the orders of summation and simplifying. First, note that the risk function can be written as:

$$\mathcal{R}(\delta_1, \delta_2, \dots, \delta_N) = E \left[ \sum_{i=1}^N \sum_{j=1}^N \left( \sum_{k=1}^N \sigma_{F_k}^2 \left( (\beta_{i,k} \beta_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) - \delta_k (\hat{\beta}_{i,k} \hat{\beta}_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) \right) \right)^2 \right] \quad (26)$$

Exchanging the order of differentiation and expectation defines the first order conditions for  $\delta_l$ :

$$\frac{d}{d\delta_l} \mathcal{R}(\delta_1, \delta_2, \dots, \delta_N) = E \left[ \sum_{i=1}^N \sum_{j=1}^N 2 \left( \sum_{k=1}^N \sigma_{F_k}^2 \left( (\beta_{i,k} \beta_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) - \delta_k (\hat{\beta}_{i,k} \hat{\beta}_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) \right) \right) \left( -(\hat{\beta}_{i,l} \hat{\beta}_{j,l} - \underline{\beta}_{i,l} \underline{\beta}_{j,l}) \right) \right] = 0 \quad (27)$$

With a bit of rearrangement, exchanging summation and expectation and moving the bias terms to the right hand side gives:

$$\sum_{k=1}^N \sigma_{F_k}^2 \sum_{i=1}^N \sum_{j=1}^N E \left[ (\beta_{i,k} \beta_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) (\hat{\beta}_{i,l} \hat{\beta}_{j,l} - \underline{\beta}_{i,l} \underline{\beta}_{j,l}) \right] \quad (28)$$

$$= \sum_{k=1}^N \delta_k \sigma_{F_k}^2 \sum_{i=1}^N \sum_{j=1}^N E \left[ (\hat{\beta}_{i,k} \hat{\beta}_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) (\hat{\beta}_{i,l} \hat{\beta}_{j,l} - \underline{\beta}_{i,l} \underline{\beta}_{j,l}) \right]$$

A zero equality will be helpful in reducing the above conditions:

$$E \left[ (\beta_{i,k} \beta_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) (\hat{\beta}_{i,l} \hat{\beta}_{j,l} - \beta_{i,l} \beta_{j,l}) \right] \quad (29)$$

$$= E \left[ E \left[ (\beta_{i,k} \beta_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) (\hat{\beta}_{i,l} \hat{\beta}_{j,l} - \beta_{i,l} \beta_{j,l}) \mid \beta_{i,k}, \beta_{j,k}, \beta_{i,l}, \beta_{j,l} \right] \right]$$

$$= E \left[ (\beta_{i,k} \beta_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) \left( E \left[ \hat{\beta}_{i,l} \hat{\beta}_{j,l} \mid \beta_{i,k}, \beta_{j,k}, \beta_{i,l}, \beta_{j,l} \right] - \beta_{i,l} \beta_{j,l} \right) \right]$$

$$= E \left[ (\beta_{i,k} \beta_{j,k} - \underline{\beta}_{i,k} \underline{\beta}_{j,k}) (\beta_{i,l} \beta_{j,l} - \beta_{i,l} \beta_{j,l}) \right] = 0$$

The first and second equalities hold by the Law of Iterated Expectations. The last equality is a consequence of the orthogonal factors and the unbiasedness of the OLS estimates. This equality reduces the expectation on the left hand side of the First Order Conditions 28:

$$\begin{aligned}
& E \left[ (\beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k}) \left( \hat{\beta}_{i,l}\hat{\beta}_{j,l} - \underline{\beta}_{i,l}\underline{\beta}_{j,l} \right) \right] \\
&= E \left[ (\beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k}) \left( \hat{\beta}_{i,l}\hat{\beta}_{j,l} - \beta_{i,l}\beta_{j,l} + \beta_{i,l}\beta_{j,lk} - \underline{\beta}_{i,l}\underline{\beta}_{j,l} \right) \right] \\
&= E \left[ (\beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k}) \left( \hat{\beta}_{i,l}\hat{\beta}_{j,l} - \beta_{i,l}\beta_{j,l} \right) \right] + E \left[ (\beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k}) (\beta_{i,l}\beta_{j,l} - \underline{\beta}_{i,l}\underline{\beta}_{j,l}) \right] \\
&= E \left[ (\beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k}) (\beta_{i,l}\beta_{j,l} - \underline{\beta}_{i,l}\underline{\beta}_{j,l}) \right]
\end{aligned} \tag{30}$$

A similar analysis for the expectation on the right hand side of the FOC's:

$$\begin{aligned}
& E \left[ \left( \hat{\beta}_{i,k}\hat{\beta}_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k} \right) \left( \hat{\beta}_{i,l}\hat{\beta}_{j,l} - \underline{\beta}_{i,l}\underline{\beta}_{j,l} \right) \right] \\
&= E \left[ \left( \hat{\beta}_{i,k}\hat{\beta}_{j,k} - \beta_{i,k}\beta_{j,k} + \beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k} \right) \left( \hat{\beta}_{i,l}\hat{\beta}_{j,l} - \beta_{i,l}\beta_{j,l} + \beta_{i,l}\beta_{j,l} - \underline{\beta}_{i,l}\underline{\beta}_{j,l} \right) \right] \\
&= E \left[ \left( \hat{\beta}_{i,k}\hat{\beta}_{j,k} - \beta_{i,k}\beta_{j,k} \right) \left( \hat{\beta}_{i,l}\hat{\beta}_{j,l} - \beta_{i,l}\beta_{j,l} \right) \right] + E \left[ \left( \hat{\beta}_{i,k}\hat{\beta}_{j,k} - \beta_{i,k}\beta_{j,k} \right) (\beta_{i,l}\beta_{j,l} - \underline{\beta}_{i,l}\underline{\beta}_{j,l}) \right] \\
&\quad + E \left[ (\beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k}) \left( \hat{\beta}_{i,l}\hat{\beta}_{j,l} - \beta_{i,l}\beta_{j,l} \right) \right] + E \left[ (\beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k}) (\beta_{i,l}\beta_{j,l} - \underline{\beta}_{i,l}\underline{\beta}_{j,l}) \right] \\
&= E \left[ \left( \hat{\beta}_{i,k}\hat{\beta}_{j,k} - \beta_{i,k}\beta_{j,k} \right) \left( \hat{\beta}_{i,l}\hat{\beta}_{j,l} - \beta_{i,l}\beta_{j,l} \right) \right] + 0 \\
&\quad + 0 + E \left[ (\beta_{i,k}\beta_{j,k} - \underline{\beta}_{i,k}\underline{\beta}_{j,k}) (\beta_{i,l}\beta_{j,l} - \underline{\beta}_{i,l}\underline{\beta}_{j,l}) \right]
\end{aligned} \tag{31}$$

Combining these results and using the definitions of  $\mathcal{B}$  and  $\mathcal{V}$  to represent their summed components, write the first order conditions as:

$$\sum_{k=1}^N \sigma_{F_k}^2 \mathcal{B}_{k,l} = \sum_{k=1}^N \delta_k \sigma_{F_k}^2 (\mathcal{V}_{k,l} + \mathcal{B}_{k,l}) \tag{32}$$

Corresponding to the linear system of equations 18. ■

## A2: Empirical Priors for General Shrinkage Estimators

The result in proposition 2 immediately extends to shrinkage estimators with any prior factor specification, but sometimes the structured shrinkage target lacks an immediate factor representation. To address this setting, denote the eigenvalue/eigenvector decomposition for the shrinkage target (an arbitrary, strictly positive-definite covariance matrix) as  $\Sigma_P = B_P \Gamma_P B_P'$  and the corresponding decomposition for the sample covariance matrix as  $\Sigma_S = B_S \Gamma_S B_S'$ . Let the complete set of orthogonal factors  $F_1, \dots, F_N$  represent this unrestricted return generating process, but rescale these so that the variance of the  $k$ th factor is now equal to the  $k$ th eigenvalue of shrinkage target  $\Sigma_P$ . An



estimator that shrinks the sample covariance matrix towards  $\Sigma_P$  can be represented as:

$$\Sigma_P^* = (1 - \delta) \Sigma_P + \delta \Sigma_S = (1 - \delta) B_P \Gamma_P B_P' + \delta B_S \Gamma_S B_S' \quad (33)$$

Here, the rescaling of the derived factors allows for a uniform shrinkage to apply across all factors. This result is summarized in proposition 5, the proof of which is omitted, as it is almost identical to that for proposition 2.

**Proposition 5** *Suppose the likelihood of the data is given by equation 6 and an investor's prior belief is given by equation 4 with parameters:*

$$\underline{\beta}_{i,k} = B_{P\{i,k\}}, \quad \underline{\Omega}_{i,\{j,k\}} = \begin{cases} \frac{1-\delta}{\delta} T \hat{\sigma}_{F_k}^2, & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

*Then the posterior covariance matrix is given by the shrinkage estimator in equation 33.*

## A2.1 Empirical Bayesian Priors Satisfying Minimal Weight Restrictions

An immediate corollary of proposition 5 relates to a shrinkage technique proposed by Jagannathan & Ma (2003). They show non-negativity constraints on the minimum variance portfolio are equivalent to a shrinkage of the covariance matrix determined by the shadow costs of those constraints. In particular, given covariance matrix  $\Sigma_S$ , a vector shadow costs for each asset's non-negativity constraint  $\lambda$ , and denoting the vector with  $N$  ones by  $1_N$ , the constrained minimum variance portfolio is equivalent to the unconstrained minimum variance portfolio for the shrinkage covariance matrix  $\Sigma_C^*$  defined as:<sup>8</sup>

$$\Sigma_C^* = \Sigma_S - 0.5 (\lambda 1_N' + 1_N \lambda') = 0.5 \Sigma_S + 0.5 (\Sigma_S - \lambda 1_N' - 1_N \lambda') \quad (34)$$

Taking the eigenvalue decomposition, define  $B_C \Gamma_C B_C' = \Sigma_S - \lambda 1_N' - 1_N \lambda'$  and invoking proposition 5 immediately proves the following corollary:

**Corollary 1** *Suppose the likelihood of the data is given by equation 6 and an investor's prior belief is given by equation 4 with parameters:*

$$\underline{\beta}_{i,k} = B_{C\{i,k\}}, \quad \underline{\Omega}_{i,\{j,k\}} = \begin{cases} T \hat{\sigma}_{F_k}^2, & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

*Then the posterior covariance matrix matches the Jagannathan and Ma estimator in equation 34.*

---

<sup>8</sup>Note that the formula presented here is slightly different than that which appears in the original text, which was afflicted by a typographical error.