Journal of Empirical Finance 29 (2014) 402–420

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

An empirical Bayesian approach to stein-optimal covariance matrix estimation☆

Benjamin J. Gillen *

Division of the Humanities and Social Sciences, California Institute of Technology, MC 228-77, Pasadena, CA 91125, United States

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form 20 August 2014
Accepted 15 September 2014
Available online 28 September 2014

JEL Classification:
C11
C55
C58
G11

Keywords:
Bayesian estimation
Covariance matrix shrinkage
Asset allocation

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.

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 and Korajczyk (1993) and Bai and Ng (2002).[1] A slightly different approach focuses on minimizing the finite-sample Stein (1955) mean square error, with a series of papers by Ledoit and Wolf (2003, 2004a,b) proposing shrinkage estimators that form a linear combination of the sample covariance matrix with a more

☆ 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.
* Tel.: +1 626 395 4061; fax: +1 626 405 9841.
E-mail address: bgillen@caltech.edu.
[1] Fan et al. (2008) provide a theoretical foundation for establishing consistency of these estimators in sparse statistical models. Recent work, including Bickel and Levina (2008a,b), Lam and Fan (2009), Cai et al. (2010), Cai and Liu (2011), and Fan et al. (2011), extends the application of sparsity to derive regularization strategies for covariance matrix estimators.

http://dx.doi.org/10.1016/j.jempfin.2014.09.006
0927-5398/

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.[2] While Yang and Berger (1994) present reference priors for the problem, a number of other researchers including Leonard and Hsu (1992) and Daniels and 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 and 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, with a security's factor loading 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 and 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 and Wolf (2012) and Ledoit and Wolf (2013) have analyzed the nonlinear regularization of the eigenvalues for covariance matrices under different loss functions. Further, Bai and 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 and Wolf (2004a) in a more flexible class or 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 and 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 estimators 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, $r_{\cdot,t} = [r_{1,t}, \ldots, r_{N,t}]'$, each of which are normally distributed with known means $\mu = [\mu_1, \ldots, \mu_N]'$ and an unknown covariance matrix Σ. To represent these returns in a linear model, assume that there are K observed factors $F_{1,t}, \ldots, F_{K,t}$ that represent all sources of variance across the securities and that these factors have known covariance matrix Γ. 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) $r_{\cdot,t} \sim N(\mu, \Sigma)$, where μ is known but Σ is unknown.
(b) $F_{\cdot,t} = [F_{1,t}, \ldots, F_{K,t}]' \sim N(\mu_F, \Gamma)$, with both μ_F and Γ known.
(c) $r_{i,\cdot} = [r_{i,1}, \ldots, r_{i,T}]' \in \mathcal{S}(F)$, the column space of the matrix $F = [F_{\cdot,1}', \ldots, F_{\cdot,T}']$ for all i.

[2] For examples, see Yang and Berger (1994) and Bensmail and 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 and 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.