A Sharpe Ratio Neutral Prior for Bayesian Portfolio Selection

Roman Croessmann
Institute for Finance & Banking
Ludwig-Maximilians-University Munich
Ludwigstrasse 28 Backbuilding, D-80539 Munich

Email: croessmann@bwl.lmu.de
Phone: +49 (0)89 2180 6995

January 15, 2016

A Sharpe Ratio Neutral Prior for Bayesian Portfolio Selection

ABSTRACT

The standard noninformative prior for Bayesian portfolio selection implies strong and unreasonable prior information about the achievable Sharpe ratio. This has critical implications for portfolio selection. We develop a reparametrization that allows to specify a prior which is flat in the achievable Sharpe ratio. Applications suggest that Bayesian portfolio selection with the Sharpe ratio neutral prior does not encounter the usual pathologies of unconstrained mean-variance optimization.

JEL classification: C11, C58, G11

Estimation error makes it hard to benefit from portfolio theory. Many researchers even argue that, in the presence of parameter uncertainty, investors are better off if they completely disregard sophisticated portfolio selection approaches and stick to ad-hoc but robust allocation rules like naive diversification (see e.g. DeMiguel, Garlappi, and Uppal (2009)). This is of course intellectually unsatisfactory – how can ignoring all the available data be optimal?

In theory, the Bayesian framework introduced by Zellner and Chetty (1965) allows to derive optimal portfolio decisions under parameter uncertainty; optimal with respect to the data, the chosen likelihood and the chosen prior distribution of the parameters. But which likelihood should be chosen? And which prior specification is reasonable in this context? The resulting portfolio performance directly depends on those choices and a reasonable specification, especially of the prior, is nontrivial.

The usual pathologies of unconstrained mean-variance optimization, i.e. extreme and unstable portfolio weights, as well as a large discrepancy between in-sample and out-of-sample performance, seem to remain when the standard noninformative prior, i.e. the Jeffreys prior for the multivariate normal likelihood, is used. It is thus often argued that accounting for parameter uncertainty with noninformative priors does not lead to substantial improvements in comparison to a mean-variance optimization with unconstrained maximum-likelihood parameter estimates.[1] This conclusion however is only valid if the standard noninformative prior is reasonably noninformative in the portfolio selection context. As discussed in Box and Tiao (1973), the informational content of a prior must be judged in the context of its application. Consider a prior which is noninformative about some parameter but which implies strong and unreasonable prior information about a specific transformation of that parameter. Such a prior should not be used in a model in which this transformation is of central importance.

This article shows that the standard noninformative prior is not a reasonable choice in the portfolio selection context because it effectively rules out all parameter combinations which imply a reasonable achievable Sharpe ratio. This seemingly innocuous prior has critical implications for portfolio optimization as it suggests that high expected returns can be obtained cheaply, i.e. that they do not come with a reasonable amount of risk. We develop a reparametrization that allows to specify a prior which is flat and thus neutral in the achievable Sharpe ratio. Recent advances in Markov Chain Monte Carlo (MCMC) methods enable us to use this prior for large-scale portfolio decisions. Applications with simulated and empirical data suggest that Bayesian portfolio selection with the Sharpe ratio neutral prior does not encounter the usual pathologies of unconstrained mean-variance optimization: It leads to well-diversified and stable portfolios with in-sample Sharpe ratio performances which are usually close to their out-of-sample counterparts. Accounting for parameter uncertainty in the Bayesian framework might meaningfully improve portfolio decisions after all. It is just not trivial to specify a reasonable prior in this context.

The article proceeds as follows. Section 1 gives a short introduction to the Bayesian portfolio selection paradigm. Section 2 shows that the standard noninformative prior implies strong and unreasonable prior information about the achievable Sharpe ratio. Section 3 develops the Sharpe ratio neutral prior. The simulation study in Section 4 and the empirical implementation in Section 5 investigate the usefulness of this novel approach. Section 6 concludes the article.

[1] See e.g. Avramov and Zhou (2010) who write "Indeed, to exhibit the decisive advantage of the Bayesian portfolio analysis, it is generally necessary to elicit informative priors that account for events, macro conditions, asset pricing theories, as well as any other insights relevant to the evolution of stock prices."

1 The Bayesian portfolio selection framework

For the rest of the paper consider the following standard setup. Assume that there exists a risk-free rate r_f and that the excess returns R_t of N investable assets have an iid multivariate normal distribution with unknown mean μ and unknown covariance matrix Σ

$$R_t \sim \mathcal{N}(\mu, \Sigma)\,. \tag{1}$$

Further assume that a matrix R^T with T historical excess return observations per asset is accessible. In the Bayesian portfolio optimization paradigm, the optimal portfolio weights are given by an expected utility maximization under the predictive distribution of the returns

$$\max_w \int U(w' R_{T+1} + r_f)\, p(R_{T+1})\ \mathrm{d}R_{T+1}. \tag{2}$$

The predictive distribution

$$p(R_{T+1}) = \iint p(R_{T+1}|\mu, \Sigma)\, p\Big(\mu, \Sigma | R^T\Big)\, \mathrm{d}\mu\ \mathrm{d}\Sigma \tag{3}$$

depends on the posterior distribution $p\Big(\mu, \Sigma | R^T\Big)$ which in turn is proportional to the product of the likelihood and the prior

$$p\Big(\mu, \Sigma | R^T\Big) \propto p\Big(R^T | \mu, \Sigma\Big)\, p(\mu, \Sigma)\,. \tag{4}$$

Mean-variance optimization assumes that the preferences of an investor are such that his expected utility maximization is equivalent to a maximization of a quadratic objective function

$$\max_w \left(w'\mu - \frac{\gamma}{2} w'\Sigma w\right) \tag{5}$$

where γ gives the risk aversion of the investor. If the parameters μ and Σ are known, the optimal portfolio weights directly follow as

$$w^* = \frac{1}{\gamma}\Sigma^{-1}\mu. \tag{6}$$

Problems in mean-variance optimization arise from the fact that μ and Σ are generally unknown and have to be inferred from available data. In the Bayesian framework,

(5) is maximized under the predictive distribution and the ex-ante optimal portfolio weights follow as

$$w_{Bayes} = \frac{1}{\gamma}\Sigma_{T+1}^{-1}\mu_{T+1} \tag{7}$$

where μ_{T+1} denotes the vector of expected excess returns implied by the predictive distribution and Σ_{T+1} denotes the covariance matrix of the excess returns implied by the predictive distribution.

The specification of a reasonable prior $p(\mu, \Sigma)$ is nontrivial in this framework. Priors which seem quite noninformative at first glance may imply strong and unreasonable prior information about economically relevant parameter transformations. The next section shows that the standard noninformative prior implies strong and unreasonable prior information about the achievable Sharpe ratio and is thus not a reasonable choice in the portfolio selection context.

2 The Jeffreys prior and the achievable Sharpe ratio

The standard noninformative prior for Bayesian portfolio selection is the Jeffreys prior for the multivariate normal likelihood

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{N+1}{2}}. \tag{8}$$

It is used in many influential articles in the field of Bayesian portfolio selection, see e.g. Klein and Bawa (1976), Bawa, Brown, and Klein (1979), Kandel, McCulloch, and Stambaugh (1995) and Stambaugh (1997). The mean-variance optimal portfolio weights when this prior is used are known to be[2]

$$w_{Jeff} = \frac{1}{\gamma}\frac{T-N-2}{T+1}\hat{\Sigma}^{-1}\hat{\mu}. \tag{9}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean and the sample covariance matrix

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t \tag{10}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})'. \tag{11}$$

[2] See Appendix A for some background on this result.

For comparison, consider the weights that result when parameter uncertainty is not accounted for and the sample mean and the sample covariance matrix are used as plug-in estimates

$$w_S = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}. \tag{12}$$

The weight vectors w_{Jeff} and w_S are just scaled by a different constant. Thus, accounting for parameter uncertainty in the Bayesian framework under the standard noninformative prior leads to the same relative weights in the risky assets as when standard sample estimates are used. That also implies that the tangency portfolio weight estimates exactly coincide for every data set.[3] This is a classical result but it is somewhat surprising. Accounting for parameter uncertainty in the Bayesian framework seems to make little difference.[4] Due to this result, it is often argued that informative priors are needed to obtain meaningfully improved portfolio weights in the Bayesian framework.

Several authors already questioned the suitability of the Jeffreys prior in the portfolio choice context. Tu and Zhou (2010) for example recognize that the standard noninformative prior implies strong information about the cross-sectional variation of the portfolio weights and Kandel, McCulloch, and Stambaugh (1995) question its noninformativity in the context of testing portfolio efficiency. The next section follows in their footsteps and shows that the Jeffreys prior should not be used in the mean-variance optimization context because it implies strong and unreasonable prior information about the achievable Sharpe ratio.

2.1 The prior on the achievable Sharpe ratio

The achievable Sharpe ratio, i.e. the Sharpe ratio of a mean-variance efficient portfolio follows as

$$SR_{max} = \frac{w^{*\prime}\mu}{\sqrt{w^{*\prime}\Sigma w^{*}}} = \frac{\left(\frac{1}{\gamma}\Sigma^{-1}\mu\right)'\mu}{\sqrt{\left(\frac{1}{\gamma}\Sigma^{-1}\mu\right)'\Sigma\left(\frac{1}{\gamma}\Sigma^{-1}\mu\right)}} = \sqrt{\mu'\Sigma^{-1}\mu}. \tag{13}$$

[3] The tangency portfolio weights under the Jeffreys prior are

$$w^{TP}_{Jeff} = \frac{1}{\iota'\frac{1}{\gamma}\frac{T-N-2}{T+1}\hat{\Sigma}^{-1}\hat{\mu}}\frac{1}{\gamma}\frac{T-N-2}{T+1}\hat{\Sigma}^{-1}\hat{\mu} = \frac{1}{\iota'\hat{\Sigma}^{-1}\hat{\mu}}\hat{\Sigma}^{-1}\hat{\mu}$$

where ι is a $N \times 1$ vector of ones. Those weights correspond to the tangency portfolio weights when $\hat{\Sigma}$ and $\hat{\mu}$ are used as plug-in estimates.

[4] In the comparative study of DeMiguel, Garlappi, and Uppal (2009), the performance of the Bayesian approach with a noninformative prior is not even reported because "the performance of the Bayesian diffuse-prior portfolio is virtually indistinguishable from that of the sample-based mean-variance portfolio."

The achievable Sharpe ratio states how much expected return is received for one unit of volatility in mean-variance efficient portfolios. As (13) shows, estimates of μ and Σ directly imply an estimate of the achievable Sharpe ratio. If the achievable Sharpe ratio is overestimated, it appears that high expected returns can be obtained cheaply, i.e. that they do not come with a reasonable amount of risk. Intuitively, this can destabilize the portfolio weight estimates as they now have the tendency to chase small differences in expected returns. The following analysis shows that the Jeffreys prior leads to such an overestimation.

The Jeffreys prior is improper which complicates its analysis. To investigate what the Jeffreys prior implies for the achievable Sharpe ratio, we use the fact that the Jeffreys prior can be obtained in the limit of the conjugate prior of a multivariate normal distribution. We proceed as follows: We first derive what the conjugate prior implies for the achievable Sharpe ratio and then investigate what happens in the limit at which this proper prior approaches the Jeffreys prior.

The conjugate prior for the multivariate normal distribution is the normal-inverse-Wishart prior

$$\Sigma \sim \mathcal{IW}_{\nu_0}\Big(\Lambda_0^{-1}\Big)$$
$$\mu|\Sigma \sim \mathcal{N}\Big(\mu_0, \frac{1}{\kappa_0}\Sigma\Big) \tag{14}$$

which corresponds to the the following density

$$p(\mu,\Sigma) \propto |\Sigma|^{-((\nu_0+d)/2+1)} \exp\Big(-\frac{1}{2}\mathrm{tr}\Big(\Lambda_0\Sigma^{-1}\Big) - \frac{\kappa_0}{2}\left(\mu-\mu_0\right)'\Sigma^{-1}\left(\mu-\mu_0\right)\Big). \tag{15}$$

The Jeffreys prior is obtained in the limit of this prior for $\kappa_0 \to 0$, $v_0 \to -1$ and $|\Lambda_0| \to 0$, irrespective of the choice of μ_0 (see Gelman et al. (2014), p. 73). To simplify the analysis, we choose $\mu_0 = 0$ without loss of generality for our results about the Jeffreys prior.

Appendix B shows that the normal-inverse-Wishart prior with $\mu_0 = 0$ implies that the prior distribution of the achievable Sharpe ratio SR_{max} is

$$SR_{max} \sim \mathrm{Nakagami}\Big(\frac{N}{2}, \frac{N}{\kappa_0}\Big). \tag{16}$$

The density implied by (16) is of the following form (up to proportionality)

$$p(SR_{max}) \propto SR_{max}^{N-1} \times \exp\Big(-\frac{\kappa_0}{2}SR_{max}^2\Big). \tag{17}$$

Note that the prior distribution of SR_{max} is independent of the choice of ν_0 and Λ_0.[5] To investigate what the Jeffreys prior implies for the achievable Sharpe ratio, we can thus simply investigate what happens to (17) when κ_0 approaches zero.

Taking the limit $\kappa_0 \to 0$ leads to the central result of this section, i.e. the improper prior on SR_{max} implied by the Jeffreys prior

$$p(SR_{max}) \propto SR_{max}^{N-1}. \tag{18}$$

As (18) shows, the prior grows monotonically in SR_{max} and it grows faster and faster as N increases. Thus, when the allegedly noninformative Jeffreys prior is employed, (μ, Σ)-combinations that lead to large achievable Sharpe ratios are favored a priori and this effect increases in the number of investable assets. Stated differently, the Jeffreys prior strongly suggests that large expected returns can be obtained cheaply, i.e. that the do not come with a reasonable amount of risk.

To get an intuition about the magnitude of prior information specified by the Jeffreys prior, consider a portfolio of 10 assets and monthly return observations. A reasonable achievable Sharpe ratio on a monthly basis would be for example $SR_{max} = 0.1$. As (18) shows, the Jeffreys prior implies that a Sharpe ratio of 1.0 on a monthly basis is $1.0^9/0.1^9 = 10^9$ times as likely a priori as a Sharpe ratio of 0.1. This seems to be extremely informative.

Why does the Jeffreys prior imply such extreme information about the achievable Sharpe ratio? Recall that μ and Σ are modeled as independent by the Jeffreys prior

$$\begin{aligned} p(\mu, \Sigma) &= p(\mu) \times p(\Sigma) \\ &\propto c \times |\Sigma|^{-\frac{N+1}{2}} \\ &\propto |\Sigma|^{-\frac{N+1}{2}}. \end{aligned}$$

This is problematic because, roughly speaking, the subspace of (μ, Σ)-combinations in which expected returns are closely linked to the covariance matrix by a set of common risk factors is small compared to the whole space of (μ, Σ)-combinations. Therefore, a marginally noninformative prior which models μ and Σ independently specifies the prior information that expected returns are very likely not a compensation for common risk. When expected returns are not a compensation for common risk, then the achievable Sharpe ratio grows without bounds in the number of investable assets (see MacKinlay (1995)). Thus, a valid intuitive interpretation is that the Jeffreys

[5] The derivation in Appendix B also shows that the distribution of the covariance matrix Σ does not matter for the distribution of SR_{max} as long as the conditional distribution of the means has the form depicted in (14) with $\mu_0 = 0$.

prior implies large achievable Sharpe ratios because it models the possibility of priced common risk as very unlikely.

This section showed that the Jeffreys prior implies strong and unreasonable prior information about the achievable Sharpe ratio. The next section investigates whether the data can be assumed informative enough about the achievable Sharpe ratio to diminish the strong information specified by the Jeffreys prior.

2.2 The posterior distribution of the achievable Sharpe ratio

This section investigates how strongly the posterior distribution of the achievable Sharpe ratio is effected by the Jeffreys prior. We proceed as follows: First, we generate a large number of matrices with return observations R^T given some true parameters μ and Σ. We follow DeMiguel, Garlappi, and Uppal (2009) and choose true parameters such that the implied ex-ante achievable Sharpe ratios are of reasonable size. Second, we obtain a large number of draws from the joint posterior distribution of μ and Σ under the Jeffreys prior for each sample matrix R^T and compute the achievable Sharpe ratio for each pair of draws. We then investigate the posterior distribution of the achievable Sharpe ratio under the Jeffreys prior.

1. Data generation: We adapt the data generating process of DeMiguel, Garlappi, and Uppal (2009) and assume that the monthly excess returns of N risky assets follow a single-factor structure

$$R_t = \beta f_t + \epsilon_t \tag{19}$$

with

$$f_t \sim \mathcal{N}\left(\mu_f, \sigma_f^2\right), \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon), \quad \mu_f = \frac{0.08}{12}, \quad \sigma_f = \frac{0.16}{\sqrt{12}}$$

where Σ_ϵ is diagonal with volatilities drawn from a uniform distribution with support $\left[\frac{0.1}{\sqrt{12}}, \frac{0.3}{\sqrt{12}}\right]$. The elements of β are spread evenly between 0.5 and 1.5. This data generating process implies a Sharpe ratio of the factor of $SR_f = \frac{\mu_f}{\sigma_f} = 0.1443$ monthly and 0.5 annualized. Due to the factor structure, this Sharpe ratio serves as upper bound for the achievable Sharpe ratio of portfolios constructed of the N investable assets. To be able to investigate the effect of the sample length and the number of investable assets on the posterior of the achievable Sharpe ratio, we generate data sets for all combinations of $T = \{120, 360\}$ months of data and $N = \{5, 10, 25, 50, 100\}$ investable assets.

2. Sampling from posterior: The posterior distribution of the primitive parameters given a matrix R^T of sample observations under the Jeffreys prior is[6]

$$\Sigma \sim \mathcal{IW}_{T-1}\left(\frac{1}{T}\hat{\Sigma}^{-1}\right)$$

$$\mu|\Sigma \sim \mathcal{N}\left(\hat{\mu}, \frac{1}{T}\Sigma\right). \tag{20}$$

The posterior distribution of SR_{max} that follows from (20) does not resemble any known distribution.[7] One can however easily obtain samples from this posterior by sampling Σ and $\mu|\Sigma$ from (20) and subsequently computing $SR_{max} = \sqrt{\mu'\Sigma^{-1}\mu}$ for each sampled (μ, Σ)-pair.

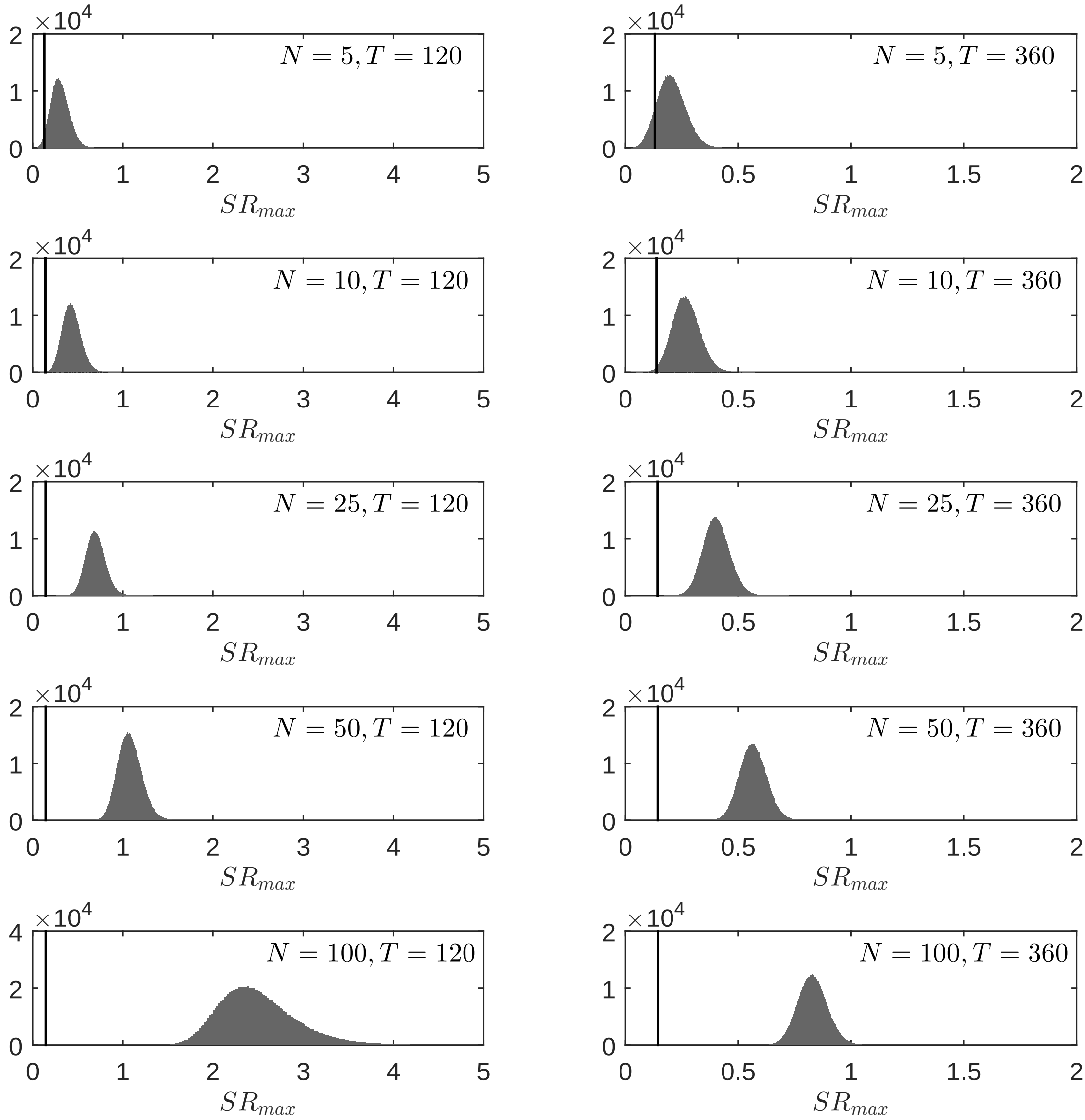
Figure 1 depicts histograms for a large number of such draws from the posterior distributions of SR_{max} for different T and N combinations. The vertical line gives the true achievable Sharpe ratio in each case. Figure 1 shows that the Jeffreys prior leads to posterior distributions that heavily overstate the achievable Sharpe ratio. As expected from the analysis of the prior, the effect increases with the size of the cross-section N and decreases with the sample length T. In all investigated $N \geq 10$ cases, there is basically no posterior mass in the region of the true achievable Sharpe ratio. Even with a small cross-section of $N = 5$, the achievable Sharpe ratio is strongly overestimated.

To put this result in the familiar mean-variance efficient frontier context, Figure 2 translates the case with $N = 25$ and $T = 120$ into the (μ, σ)-space. The achievable Sharpe ratio is the slope of the mean-variance efficient frontier in the presence of a risk-free rate. Thus, the posterior distribution of the achievable Sharpe ratio directly translates into a posterior confidence region for the mean-variance efficient frontier. As Figure 2 shows, the true mean-variance efficient frontier is far outside the 95% posterior confidence region that follows under the Jeffreys prior. Bayesian portfolio selection with the Jeffreys prior suggests that the achievable Sharpe ratio is big and thus rules out all efficient frontiers with a reasonable risk-return tradeoff.

[6] See Appendix C.

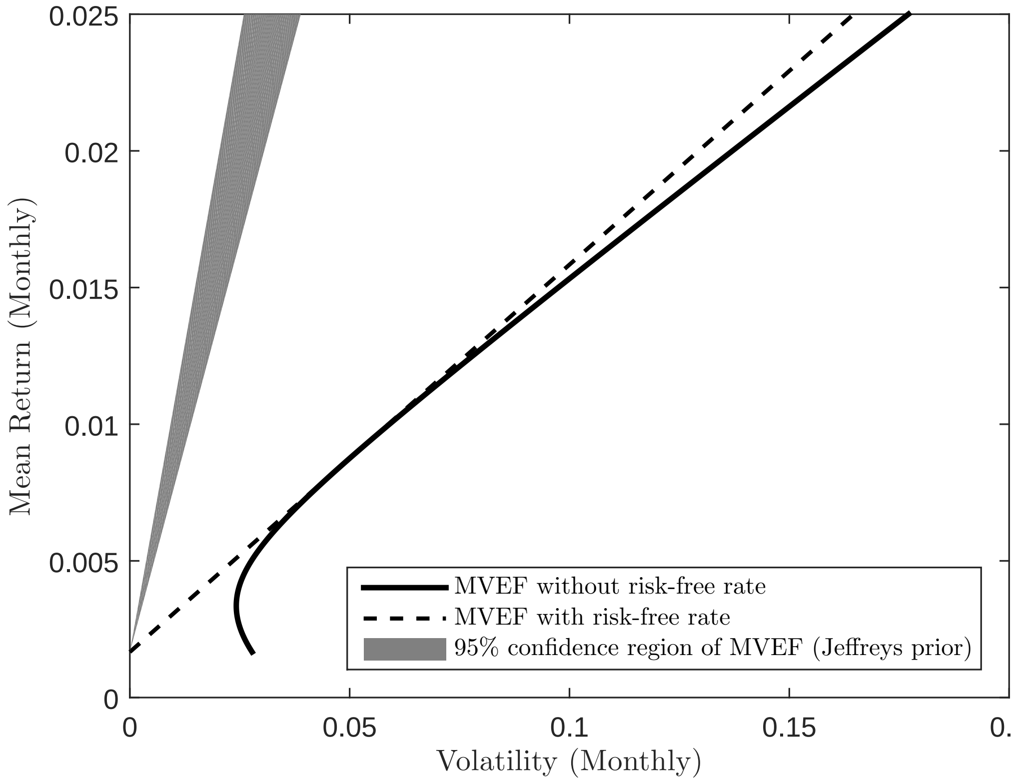
[7] In contrast to Section 2.1, the achievable Sharpe ratio is not Nakagami distributed because $\hat{\mu}$ is generally not a zero vector.

Figure 1: Posterior distributions of SR_{max}.



This Figure shows histograms for draws from the posterior distributions of the achievable Sharpe ratio when the Jeffreys prior is employed. The vertical line gives the true achievable Sharpe ratio in each case. The Jeffreys prior leads to posterior distributions that heavily overstate the achievable Sharpe ratio. This effect is getting stronger as the size of the cross-section increases and slowly decreases in the sample length.

Figure 2: Posterior distribution of SR_{max} and the mean-variance efficient frontier



This Figure shows the 95% posterior confidence region of the mean-variance efficient frontier (with risk-free rate) when the Jeffreys prior is employed. The data generating process follows DeMiguel, Garlappi, and Uppal (2009) and the results are depicted for $N = 25$ investable assets and $T = 120$ monthly observations per asset. The true mean-variance efficient frontiers (with and without risk-free rate) are depicted for comparison. The Jeffreys prior leads to posterior distributions in which all efficient frontiers with a reasonable risk return tradeoff are ruled out. The true mean variance efficient frontier is far outside the 95% confidence region.

As Figure 1 and Figure 2 show, the unreasonable prior information specified by the Jeffreys prior is not overruled by the data for reasonable sample sizes. When this prior is used in the Bayesian framework, the posterior suggests that the achievable Sharpe ratio is large even if the true achievable Sharpe ratio is small. From an economic point of view, the Jeffreys prior is not a reasonable choice in the portfolio selection context. It effectively rules out the set of reasonable solutions, i.e. the set of solutions where expected returns are an adequate compensation for common risk. The next sections develops a prior which is flat in the achievable Sharpe ratio and thus does not rule out (μ, σ)-combinations which imply a reasonable risk-return tradeoff.

3 A Sharpe ratio neutral prior

The last sections showed that the Jeffreys prior implies very strong and unreasonable prior information about the achievable Sharpe ratio. The basic idea of this section is to find a reparametrization of the multivariate normal likelihood

$$R_t \sim \mathcal{N}(\mu, \Sigma) \tag{21}$$

that allows to easily control the prior of the achievable Sharpe ratio. Section 3.1 derives this reparametrization and Section 3.2 shows how a Sharpe ratio neutral prior can be specified under this reparametrization. Our analysis is related to Kandel, McCulloch, and Stambaugh (1995) which questions the noninformativity of the Jeffreys prior in the context of testing portfolio efficiency and develops an informative prior which is less unreasonable about their measure of portfolio efficiency.

3.1 A Useful Reparametrization

To obtain a reparametrization which allows to easily put a prior on the achievable Sharpe ratio, we use the fact that the Sharpe ratio of the factor in a single-factor representation gives the upper bound for the achievable Sharpe ratio of portfolios build from the assets that are priced by this factor. This holds because a factor prices all asset if and only if it lies on the mean-variance efficient frontier (see Roll (1977)). From asset pricing theory it is well known that there always exists such a single-factor representation that prices all assets if the law of one price holds (see for example Cochrane (2001)). Thus, by assuming that the law of one price holds we can write

$$E(R_t) = \beta\lambda \quad \forall t. \tag{22}$$

Because β follows from time-series regressions of returns on the pricing factor we can write

$$R_t = \beta f_t + \epsilon_t \quad \forall t, \tag{23}$$

where f_t and ϵ_t are orthogonal and

$$E(f_t) = \lambda, \quad E(\epsilon_t) = 0 \quad \forall t. \tag{24}$$

This rerepresentation helps, because the Sharpe ratio of the factor now gives the upper bound for the achievable Sharpe of all portfolios build from the assets that are priced by the factor.[8]

Because, in contrast to Pástor and Stambaugh (2000) and Pástor (2000), we do not want to model a belief in a specific asset pricing model, we consider the true pricing factor in (23) unknown and model it as a normal random variable

$$f_t \sim \mathcal{N}\Big(\lambda, \sigma_f^2\Big) \quad \forall t. \tag{25}$$

The error terms in (23) are modeled as multivariate normal

$$\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon) \quad \forall t \tag{26}$$

and as in Tipping and Bishop (1999), the factor is integrated out which leads to the following likelihood

$$R_t \sim \mathcal{N}\Big(\beta\lambda, \beta\beta'\sigma_f^2 + \Sigma_\epsilon\Big) . \tag{27}$$

There is one subtlety that is important for the later application: Model (23) has the following well known rotational indeterminacy

$$\begin{aligned} R_t &= \beta f_t + \epsilon_t \\ \Leftrightarrow R_t &= (\beta q)\left(\frac{1}{q} f_t\right) + \epsilon_t \end{aligned} \tag{28}$$

for every nonzero scalar q, because for an arbitrary nonzero q, one obtains exactly the same likelihood.[9] This complicates estimation and we adapt the approach of Geweke and Zhou (1996) to deal with this rotational indeterminacy. Without loss of generality, we fix the rotation by fixing $\sigma_f^2 = 1$ which fixes the absolute value of q and constrain the first entry in β to be positive which fixes the sign of q. The likelihood becomes

$$R_t \sim \mathcal{N}(\beta\lambda, \beta\beta' + \Sigma_\epsilon) . \tag{29}$$

where β_1, i.e. the first element of β, is now restricted to be positive and the upper bound for the Sharpe ratio is now given by λ. This likelihood will be used in the rest of this article.

---

[8] The achievable Sharpe ratio equals the Sharpe ratio of the factor if the factor can be replicated by a portfolio of the investable assets. Otherwise, the Sharpe ratio of the factor acts only as upper bound for the achievable Sharpe ratio.

[9] For a discussion of the rotational indeterminacy in such factor models see e.g. Geweke and Zhou (1996) and Connor, Goldberg, and Korajczyk (2010).

The likelihood in (29) is just a useful reparametrization of the likelihood in (21). It does not assume that a single-factor structure holds because it does not restrict the covariance matrix of the error terms. This reparametrization also does not rule out that the assets are priced by a multi-factor model because for every multi-factor model, there always exist an equivalent single-factor representation that prices all assets.[10] Because (29) is just a reparametrization of (21), the maximum-likelihood estimates of the expected returns and the covariance matrix under (29) are identical to the maximum-likelihood estimates under the usual (μ, Σ)-parametrization of the multivariate normal likelihood. What we achieved with (29) is a nonrestrictive reparametrization that enables us to easily specify a prior on the expected returns $\mu = \beta\lambda$, the covariance matrix $\Sigma = \beta\beta' + \Sigma_\epsilon$ and the Sharpe ratio of the factor $SR_f = \lambda$, i.e. the upper bound for the achievable Sharpe ratio.

3.2 A Sharpe ratio neutral prior specification

Consider the reparameterized likelihood (29). To be reasonably noninformative about the expected returns and the achievable Sharpe ratio, we choose flat priors on β and λ. This results in a flat prior on the expected returns $\mu = \beta\lambda$ and in a flat prior on the Sharpe ratio of the factor $SR_f = \lambda$.[11] To specify a reasonably noninformative prior on Σ_ϵ we adapt the approach of Barnard, McCulloch, and Meng (2000) and decompose Σ_ϵ into the volatilities of the error terms τ and the correlation matrix of the error terms Ω

$$\Sigma_\epsilon = \text{diag}(\tau)\, \Omega \, \text{diag}(\tau) \tag{30}$$

where diag(z) gives a diagonal matrix with the vector z as diagonal. We choose a flat prior on positive values of the volatilities τ. For the correlation matrix Ω we use the prior suggested by Barnard, McCulloch, and Meng (2000)

$$p(\Omega) \propto |\Omega|^{\frac{N(N-1)}{2}-1} \left(\prod_i |\Omega_{ii}| \right)^{-\frac{N+1}{2}} \tag{31}$$

where Ω_{ii} is the matrix that results when the ith row and the ith column are removed from Ω. This prior implies a marginally uniform distribution of all pairwise error

[10] See e.g. Cochrane (2001).

[11] Recall that the volatility of the factor has been fixed to one.

term correlations which seems reasonably noninformative in the portfolio selection context.[12]

To sum up, our prior distribution is

$$\begin{aligned} p(\beta, \lambda, \tau, \Omega) &\propto p(\Omega) \\ &\propto |\Omega|^{\frac{N(N-1)}{2}-1} \left(\prod_i |\Omega_{ii}|\right)^{-\frac{N+1}{2}} \end{aligned} \tag{32}$$

which leads to the following posterior

$$\begin{aligned} p\left(\beta, \lambda, \tau, \Omega | R^T\right) &\propto p\left(R^T | \beta, \lambda, \tau, \Omega\right) p(\beta, \lambda, \tau, \Omega) \\ &\propto p\left(R^T | \beta, \lambda, \tau, \Omega\right) p(\Omega) \\ &\propto \mathcal{N}\left(R^T | \beta\lambda, \beta\beta' + \mathrm{diag}(\tau)\, \Omega\ \mathrm{diag}(\tau)\right) \times |\Omega|^{\frac{N(N-1)}{2}-1} \left(\prod_i |\Omega_{ii}|\right)^{-\frac{N+1}{2}} \\ &\propto |\beta\beta' + \mathrm{diag}(\tau)\, \Omega\ \mathrm{diag}(\tau)|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} tr\left(\left(\beta\beta' + \mathrm{diag}(\tau)\, \Omega\ \mathrm{diag}(\tau)\right)^{-1} S\right)\right) \\ &\quad \times |\Omega|^{\frac{N(N-1)}{2}-1} \left(\prod_i |\Omega_{ii}|\right)^{-\frac{N+1}{2}} \end{aligned} \tag{33}$$

where S is the matrix of sum of squares

$$S = \sum_{t=1}^{T} \left(R_t - \bar{R}^T\right) \left(R_t - \bar{R}^T\right)'$$

and $\bar{R}^T$ is the vector of sample averages. The parameters have the following support

$$\begin{aligned} \beta_1 &: (0, \infty) \\ \beta_{-1} &: (-\infty, \infty) \\ \lambda &: (-\infty, \infty) \\ \tau &: (0, \infty) \\ \Omega &: \text{all positive definite matrices with unit diagonal elements.} \end{aligned}$$

The posterior distribution in (33) does not resemble any known distribution and first and second moments of the predictive distribution are not readily available in

[12] A more standard alternative would be to use an inverse-Wishart prior for Σ_ϵ. It is however well known that the inverse-Wishart prior is quite restrictive about the volatilities and also implies some unintended dependency between the correlations and the volatilities. We thus prefer to combine the prior of Barnard, McCulloch, and Meng (2000) for the correlation matrix with a flat prior on the volatilities.

closed-form. One can however use MCMC algorithms to obtain samples from this posterior. In this article we choose Stan, a software which employs Hamiltonian Monte Carlo, for the posterior sampling under the Sharpe ratio neutral prior.[13] The draws generated through Stan are then used to compute first and second moments of the predictive distribution of returns. Subsequently, fully Bayesian optimal portfolio weights are computed given those moments as depicted in Section 1. The next section discusses the sampling procedure in more detail.

3.3 Posterior sampling with Stan

We use Stan to draw samples from the posterior distribution that follows under the Sharpe ratio neutral prior. The dimensionality of the posterior is $D = \frac{N^2+3N+2}{2}$ which means ones has to sample $D = 1{,}326$ parameters for a portfolio choice problem with $N = 50$ investable assets. Efficient sampling from such high-dimensional (and rather complicated) posteriors is basically unachievable with standard MCMC algorithms. Hamiltonian Monte Carlo with Stan leads to much less autocorrelated samples than standard MCMC algorithms which allows to use a rather small number of iterations and makes the approximation of the posterior computationally feasible.

For each evaluation of the posterior, four parallel Markov chains are run with 4,000 iterations per chain. Whenever the Stan convergence diagnostic $\hat{R}$ signals convergence difficulties by exceeding values of 1.1 for a least one parameter, the sampling procedure is repeated and the number of iterations doubled. The chains are initialized at values that roughly capture the scale of the parameters. Given the S draws from the posterior in each simulation repetition, the "exact" mean and the "exact" covariance matrix of the predictive distribution, denoted as μ_{T+1} and Σ_{T+1}, are calculated. This can be achieved by using the properties of mixture distributions made up from convex combinations, i.e.[14]

$$\mu_{T+1} = \frac{1}{S}\sum_{s=1}^{S}\mu_s \tag{34}$$

[13] We are very thankful to the developers of Stan (Stan Developement Team (2015)), for providing this great software, and to Brian Lau, for providing the MATLAB interface (Lau (2015)).

[14] Usually, draws from the posterior are used to generate draws from the predictive distribution. This however adds another layer of possible sampling inaccuracy. To avoid this, we calculate the "exact" mean and the "exact" covariance matrix of the predictive distribution given the draws from the posterior. Both procedures lead to the same moments of the predictive distribution asymptotically but the mixture distribution method is computationally much less intensive and maximally precise.

and

$$\Sigma_{T+1} = \frac{1}{S}\sum_{s=1}^{S}\Big(\left(\mu_s - \mu_{T+1}\right)\left(\mu_s - \mu_{T+1}\right)' + \Sigma_s\Big). \tag{35}$$

where μ_s and Σ_s are the individual draws from the posterior. The tangency portfolio weight estimates under the proposed Sharpe ratio neutral prior follow as

$$w_{SRN} = \frac{1}{\iota'\Sigma_{T+1}^{-1}\mu_{T+1}}\Sigma_{T+1}^{-1}\mu_{T+1}. \tag{36}$$

4 Simulation study

This section uses a simulation study to analyze the portfolios that results under the Sharpe ratio neutral prior. The in-sample and out-of-sample Sharpe ratio performances as well as the portfolio weight stability are investigated for two different data generating processes. First, the data generating process of DeMiguel, Garlappi, and Uppal (2009) which we already used in section 2.2 and under which returns are generated by a single-factor structure. Second, a three-factor structure which is calibrated to historical return data of the Fama and French (1993) three-factor model – a setup suggested by Tu and Zhou (2011). Data sets are generated for all combinations of $T = \{120, 360, 1200\}$ months of data and $N = \{10, 25, 50\}$ investable assets. The Sharpe ratio performance and portfolio weight stability of the tangency portfolio weights under the Sharpe ratio neutral prior is compared with the performance and weight stability of a naive diversification strategy and the performance and weight stability of the tangency portfolio weights under the Jeffreys prior.

4.1 DeMiguel, Garlappi, and Uppal (2009) one-factor structure

It is assumed that the monthly excess returns of N risky investable assets follow a single-factor structure

$$R_t = \beta f_t + \epsilon_t \tag{37}$$

with

$$f_t \sim \mathcal{N}\Big(\mu_f, \sigma_f^2\Big), \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon), \quad \mu_f = \frac{0.08}{12}, \quad \sigma_f = \frac{0.16}{\sqrt{12}}$$

where Σ_ϵ is diagonal with volatilities drawn from a uniform distribution with support $\left[\frac{0.1}{\sqrt{12}}, \frac{0.3}{\sqrt{12}}\right]$. The elements of β are spread evenly between 0.5 and 1.5. This data generating process is a slight variation of the factor structure in MacKinlay and Pastor (2000) – the only difference is that MacKinlay and Pastor (2000) assume homoscedastic error terms.

DeMiguel, Garlappi, and Uppal (2009) showed that the naive diversification strategy consequently outperforms a large variety of sophisticated approaches from the literature under this data generating process. It is thus interesting to see whether accounting for parameter uncertainty under the Sharpe ratio neutral prior leads to a good performance relative to a naive diversification strategy in this setup.

This data generating process implies that the naive diversification strategy is close to optimal in terms of Sharpe ratio performance and thus hard to beat by construction. Nevertheless, as the first set of rows of Table 1 shows, Bayesian portfolio selection under the Sharpe ratio neutral prior outperforms the naive diversification strategy for $N = 25$ and $N = 50$ in terms of average Sharpe ratio. For $N = 10$, the Sharpe ratio neutral prior performs slightly worse for 120 months of data and 360 months of data. These results suggest that, in contrast to the Jeffreys prior, large cross-sections can be beneficial when the Sharpe ratio neutral prior is employed. The Sharpe ratio neutral prior strongly outperforms the Jeffreys prior for all investigated (N, T)-combinations.

A closer investigation of the influence of the size of the cross-section shows that the posterior distribution of the β's is very wide for small N and becomes more concentrated when N increases. Intuitively, this is likely due to the fact that it is hard to infer a common source of price risk from just a small number of assets. Large cross-sections reveal more information about which variation is priced and thus can benefit Bayesian portfolio optimization under the Sharpe ratio neutral prior.

The second set of rows shows the relative frequencies at which the Sharpe ratio neutral prior outperformed the other strategies in terms of Sharpe ratio. The Sharpe ratio neutral prior outperformed naive diversification in almost every simulation run for $N = 25$ and $N = 50$ investable assets. With $N = 10$ investable assets, the Sharpe ratio neutral prior outperformed naive diversification in sightly less than 50% of the simulation runs with $T = 120$ months of data and in slightly more than 50% of the simulation runs with $T = 360$ and $T = 1{,}200$ months of data. The Sharpe ratio neutral prior outperformed the Jeffreys prior in 100% of the simulations runs for all investigated (N, T)-combinations except for the $N = 10$ and $T = 120$ case where it outperformed with a relative frequency of 98.60%.

The third set of rows gives the average in-sample Sharpe ratios under the Sharpe ratio neutral prior and the Jeffreys prior. As can be expected from the analysis in Section 2, the achievable Sharpe ratio is strongly overestimated under the Jeffreys prior, especially for large cross-sections. The direct comparsison shows that the in-sample Sharpe ratios under the Sharpe ratio neutral prior are slightly underestimated under this data generating process but much closer to the true achievable Sharpe ratios.

Table 1: Simulation results: One-factor model

	$N = 10$			$N = 25$			$N = 50$		
	$T = 120$	$T = 360$	$T = 1{,}200$	$T = 120$	$T = 360$	$T = 1{,}200$	$T = 120$	$T = 360$	$T = 1{,}200$
SR_{True}	13.69	13.69	13.67	14.13	14.12	14.13	14.28	14.28	14.28
$SR_{1/N}$	13.36	13.35	13.35	13.97	13.97	13.97	14.19	14.20	14.20
SR_{SRN}	12.42	13.26	13.30	14.08	14.11	14.10	14.26	14.28	14.28
SR_{Jeff}	5.07	8.32	11.43	3.03	5.75	9.80	1.67	4.42	7.48
$SR_{SRN} > SR_{1/N}$	46.00	53.00	55.33	98.40	99.40	98.00	98.00	100.00	100.00
$SR_{SRN} > SR_{Jeff}$	98.60	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
IS SR_{SRN}	12.42	13.10	13.36	12.98	13.30	13.67	10.21	12.78	13.40
IS SR_{Jeff}	32.66	21.50	16.42	54.11	30.58	20.04	86.87	42.54	24.86
$\bar{\sigma}(w_{True})$	7.24	7.35	7.12	2.97	2.94	2.95	1.50	1.53	1.50
$\bar{\sigma}(w_{SRN})$	19.43	11.49	11.04	3.37	3.10	3.25	1.65	1.56	1.53
$\bar{\sigma}(w_{Jeff})$	388.66	131.24	24.90	181.69	302.21	28.39	247.50	94.65	83.13
Runs	500	500	300	500	500	100	100	100	100

This table shows simulation results for all combinations of $N = \{10, 25, 50\}$ investable assets and $T = \{120, 360, 1200\}$ months of data. The data generating process is adapted from DeMiguel, Garlappi, and Uppal (2009). All values are given in percentages. The first set of rows shows the average monthly out-of-sample Sharpe ratio performance of the different strategies. The second set of rows shows the relative frequencies in which the Sharpe ratio neutral prior outperformed the other strategies in terms of out-of-sample Sharpe ratio. The third set of rows shows the in-sample Sharpe ratios of the data based strategies. The fourth set of rows shows the average cross-sectional standard deviations of the respective portfolio weights as a measure of diversification. The last row shows the number of simulation runs for each (N, T)-combination.

The fourth set of rows shows the average cross-sectional standard deviations of the respective portfolio weights in percentages as a measure of diversification. The ex-ante optimal portfolio weights are well-diversified under this data generating process. Nevertheless, the Jeffreys prior leads to very extreme portfolio weights that counter the intuition of diversification. The Sharpe ratio neutral prior resolves this problem and produces well-diversified portfolios with cross-sectional standard deviations of the portfolio weights only slightly above the standard deviation of the ex-ante optimal portfolio weights. The cross-sectional standard deviation of the weights in a naive diversification strategy is zero and hence omitted from the table.

Our simulation results under the data generating process of DeMiguel, Garlappi, and Uppal (2009) suggest that Bayesian portfolio selection with the Sharpe ratio neutral prior does not encounter the usual pathologies of mean-variance optimization. The resulting portfolios are well-diversified and the in-sample Sharpe ratio performance is close to what this approach achieves out-of-sample. This is in direct contrast to the Jeffreys prior which produces very extreme portfolio weights and strongly overestimates the achievable Sharpe ratio. Additionally, the Sharpe ratio neutral prior outperforms the Jeffreys prior in terms of Sharpe ratio in close to 100% of the simulation runs and performs well in comparison to a naive diversification strategy.

Results might be qualitatively different under other data generating processes. The next section investigates whether the results change qualitatively when a three-factor model, calibrated to historical data of the Fama and French (1993) three-factor model, is used to generate the returns.

4.2 Tu and Zhou (2011) three-factor structure

The data generating process used in this section is taken from Tu and Zhou (2011). It is assumed that the monthly excess returns of N risky investable assets follow a three-factor structure

$$R_t = \beta_1 f_{1,t} + \beta_2 f_{2,t} + \beta_3 f_{3,t} + \epsilon_t. \tag{38}$$

The means and the covariance matrix of the factors are calibrated to monthly return data of the Fama and French (1993) three-factor model. We use a slightly longer calibration period than Tu and Zhou (2011), i.e. from July 1963 until September 2015 instead of July 1963 until August 2007. As in Tu and Zhou (2011), the β's are evenly spread between 0.9 and 1.2 for the market factor, between -0.3 and 1.4 for the small-minus-big factor and between -0.5 and 0.9 for the high-minus-low factor and all β's are randomly assigned to the assets.[15] The noise is assumed to be multivariate normal

$$\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$$

where Σ_ϵ is diagonal with volatilities drawn from a uniform distribution with support $\left[\frac{0.1}{\sqrt{12}}, \frac{0.3}{\sqrt{12}}\right]$. The Fama and French (1993) model has proven to explain the cross-section of expected returns and the common variation of returns quite well in empirical data. It is thus interesting to see how the Sharpe ration neutral prior performs in this simulation environment.

The first set of rows of Table 2 shows that Bayesian portfolio selection under the Sharpe ratio neutral prior outperforms naive diversification in terms of average Sharpe ratio for $T = 360$ and $T = 1{,}200$ months of data and slightly underperforms naive diversification for $T = 120$ months of data for all sizes of the cross-section. As under the data generating process of DeMiguel, Garlappi, and Uppal (2009), the Sharpe ratio neutral prior strongly outperforms the Jeffreys prior for all investigated (N, T)-combinations.

[15] This roughly captures the observed ranges of factor loadings of the Fama-French 25 size and book-to-market portfolios.

As the second set of rows shows, the Sharpe ratio neutral prior outperforms naive diversification in more than 50% of the simulation runs for every (N, T)-combination except for the $(N = 50, T = 120)$-case were it outperforms in only 35% of the simulation runs. The Sharpe ratio neutral prior outperforms the Jeffreys prior in over 95% of the simulation runs for $T = 120$ and $T = 360$ months of data.

Table 2: Simulation results: Three-factor model

	$N = 10$			$N = 25$			$N = 50$		
	$T = 120$	$T = 360$	$T = 1{,}200$	$T = 120$	$T = 360$	$T = 1{,}200$	$T = 120$	$T = 360$	$T = 1{,}200$
SR_{True}	15.80	15.80	15.79	17.62	17.63	17.65	18.76	18.79	18.74
$SR_{1/N}$	12.72	12.71	12.72	13.18	13.18	13.17	13.34	13.34	13.34
SR_{SRN}	12.12	13.56	14.55	13.14	13.44	14.39	13.20	13.37	13.60
SR_{Jeff}	6.92	10.70	13.78	5.64	9.26	13.56	3.31	7.64	12.68
$SR_{SRN} > SR_{1/N}$	56.80	83.20	99.00	51.20	69.60	96.00	35.00	51.00	72.00
$SR_{SRN} > SR_{Jeff}$	96.40	95.20	88.00	99.60	98.00	77.00	100.00	100.00	84.00
IS SR_{SRN}	12.78	13.75	14.55	12.06	12.57	14.13	9.52	12.81	13.35
IS SR_{Jeff}	33.63	23.19	18.33	55.00	32.72	22.69	87.55	44.61	28.27
$\bar{\sigma}(w_{True})$	21.93	21.74	21.40	13.62	13.23	13.34	8.27	8.16	8.28
$\bar{\sigma}(w_{SRN})$	43.08	13.81	17.00	3.44	2.86	3.65	3.45	1.56	1.43
$\bar{\sigma}(w_{Jeff})$	209.88	88.17	31.54	189.08	132.40	25.72	145.40	59.53	22.66
Runs	500	500	100	250	250	100	100	100	50

This table shows simulation results for all combinations of $N = \{10, 25, 50\}$ investable assets and $T = \{120, 360, 1200\}$ months of data. The data generating process is a three-factor structure which is calibrated to historical return data of the Fama and French (1993) three-factor model. All values are given in percentages. The first set of rows shows the average monthly out-of-sample Sharpe ratio performance of the different strategies. The second set of rows shows the relative frequencies in which the Sharpe ratio neutral prior outperformed the other strategies in terms of out-of-sample Sharpe ratio. The third set of rows shows the in-sample Sharpe ratios of the data based strategies. The fourth set of rows shows the average cross-sectional standard deviations of the respective portfolio weights as a measure of diversification. The last row shows the number of simulation runs for each (N, T)-combination.

The third set of rows shows that the achievable Sharpe ratio is again strongly overestimated under the Jeffreys prior, especially for large cross-sections. The Sharpe ratio neutral prior on the other hand leads to a slight underestimation of the achievable Sharpe ratio. The in-sample Sharpe ratios under the Sharpe ratio neutral prior are however much closer to the true achievable Sharpe ratios then under the Jeffreys prior.

The fourth set of rows shows that the three-factor structure leads to much larger average cross-sectional standard deviations of the optimal portfolio weights then the data generating process of DeMiguel, Garlappi, and Uppal (2009). The Sharpe ratio neutral prior produces rather well-diversified portfolios whereas the Jeffreys prior leads to extreme allocations.

Overall, the simulation under the Tu and Zhou (2011) three-factor structure leads to qualitatively similar results as under the data generating process of DeMiguel, Garlappi, and Uppal (2009). The resulting portfolios under the Sharpe ratio neutral prior are much more diversified and the in-sample Sharpe ratio performances are much closer to their out-of-sample counterparts then under the Jeffreys prior. The Sharpe ratio neutral prior performs reasonably well in terms of average Sharpe ratio when compared to a naive diversification strategy and consequently outperforms Bayesian portfolio selection with the Jeffreys prior.

5 Empirical analysis

This section investigates with empirical data whether the usual pathologies of unconstrained mean-variance optimization remain when the Sharpe ratio neutral prior is used. As in our simulation study, the performance of the Sharpe ratio neutral prior is compared to the performance of Bayesian portfolio selection with the Jeffreys prior and to a naive diversification strategy. Because our approach is computationally very demanding, we concentrate on two standard data sets from the literature. First, monthly return data of the Fama-French 25 portfolios, sorted by book-to-market and size, equally-weighted and from July 1926 until July 2015. Second, monthly return data of the Fama-French 49 industry portfolios, equally-weighted and from July 1969 until July 2015.[16]

A rolling estimation windows with $T = \{120, 240, 360\}$ months of data is used to estimate the portfolio weights. Reestimation takes place every twelve months and portfolio weights are applied to the subsequent twelve months.[17] In line with our simulation study, we compare out-of-sample Sharpe ratios, in-sample estimates of the achievable Sharpe ratio and the average cross-sectional standard deviations of the portfolio weights as a measure of portfolio diversification. Additionally, we investigate the average absolute change in the portfolio weights from one estimation period to the next as a measure of weight stability over time.

$$\overline{Abs}(w) = \frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} \left| w_{i,t} - w_{i,t-1} \right| . \tag{39}$$

The first set of rows of Table 3 shows the out-of-sample Sharpe ratios for the three different strategies. The p-values of the differences between the Sharpe ratio under the Sharpe ratio neutral prior and the other two approaches are given in parentheses. To compute these p-values, we follow DeMiguel, Garlappi, and Uppal (2009) and use

[16] Data prior to July 1969 is ignored to avoid missing values.

[17] A monthly reestimation could favor our approach but would increase the already extensive computing time by the factor of twelve and is therefore not conducted.

the Jobson and Korkie (1981) methodology with the correction given in Memmel (2003). Four Sharpe ratios are statistically different from each other at the five percent level: The Sharpe ratio neutral prior significantly outperforms the Jeffreys prior in the Fama-French 25 data set with an estimation window of $T = 120$ months and outperforms naive diversification with $T = 240$ and $T = 360$ months estimation window. With $T = 360$ months estimation window, the Jeffreys prior outperforms the Sharpe ratio neutral prior in the Fama-French 25 data set. These performance differences are also in an economically meaningful order of magnitude. All other performance differences are not statistically significant at the five percent level.

The second set of rows gives the in-sample estimates of the achievable Sharpe ratio. The Sharpe ratio neutral prior leads to in-sample estimates that are quite close to the out-of-sample performance achieved by this strategy. The Jeffreys prior on the other hand leads to highly overestimated achievable Sharpe ratio. This overestimation increases in the size of the cross-section and decreases in the length of the estimation window which is in line with the results of our analysis of the Jeffreys prior in the previous sections.

Table 3: Empirical results

	FF 25 Portfolios			FF 49 Industry Portfolios		
	$T = 120$	$T = 240$	$T = 360$	$T = 120$	$T = 240$	$T = 360$
SR_{SRN}	22.17	27.18	27.30	20.79	19.23	16.94
SR_{Jeff}	3.72	27.90	37.75	17.42	20.23	20.87
	(0.00)	(0.43)	(0.01)	(0.32)	(0.45)	(0.37)
$SR_{1/N}$	21.15	22.62	22.45	21.40	19.59	18.19
	(0.32)	(0.00)	(0.02)	(0.43)	(0.46)	(0.42)
IS SR_{SRN}	21.21	29.58	32.14	16.72	19.53	20.18
IS SR_{Jeff}	82.98	66.08	58.47	109.87	75.24	64.09
$\bar{\sigma}(w_{SRN})$	2.34	7.64	12.87	1.92	1.70	1.61
$\bar{\sigma}(w_{Jeff})$	192.95	88.79	68.10	48.38	29.83	23.59
$\overline{Abs}\,(w_{SRN})$	1.13	2.11	2.75	0.90	0.43	0.33
$\overline{Abs}\,(w_{Jeff})$	168.74	28.06	12.83	22.00	7.84	5.38

This table shows empirical results for the Fama-French 25 portfolios and the Fama-French 49 industry portfolios with rolling estimation windows of $T = \{120, 240, 360\}$ months of data. The first set of rows shows the average monthly Sharpe ratio performance of the different strategies. The p-values of the differences between the Sharpe ratio under the Sharpe ratio neutral prior and the other two approaches are given in parentheses. The second set of rows gives the in-sample estimates of the achievable Sharpe ratios. The third set of rows shows the average cross-sectional standard deviations of the respective portfolio weights as a measure of diversification. The last set of rows gives the average absolute change in the portfolio weights from one estimation period to the next as a measure of weight stability over time. All values, except for the p-values, are given in percentages.

The third set of rows shows the average cross-sectional standard deviations of the respective portfolio weights in percentages as a measure of diversification. As with simulated data, the Jeffreys prior leads to very extreme portfolio weights which would not be implementable. The Sharpe ratio neutral prior on the other hand resolves this problem and produces well diversified portfolios.

The last set of rows gives the average absolute change in the portfolio weights from one estimation period to the next as a measure of weight stability over time. The Jeffreys prior leads to portfolio weights which often change drastically from one estimation period to the next. Even with an estimation period of thirty years, the weights under the Jeffreys prior are very unstable and show average absolute changes of 12.83% for the Fama-French 25 portfolios and of 5.38% for the Fama-French 49 portfolios. Under the Sharpe ratio neutral prior, the portfolio weights are much more stable over time with an average absolute change of 2.75% and 0.33% respectively.

All results in Table 3 are in line with the results from our analysis in Section 2 and the results from the simulation study in Section 4. The Jeffreys prior leads to overestimated achievable Sharpe ratios and to extreme and unstable portfolio weights. The Sharpe ratio neutral prior leads to estimates of the achievable Sharpe ratio that are close to what is achieved out-of-sample and to well-diversified portfolios weights which are much more stable over time. Judging the relative performance of the strategies is difficult with empirical data due to the large uncertainty in the out-of-sample Sharpe ratios. Nevertheless, three of the four performance differences in Table 3 which are significant at the five percent level are in favor of the Sharpe ratio neutral prior which is a first indication that this approach performs reasonably well.

6 Conclusion

It is often argued that mean-variance optimizers act as error maximizers: the optimization chases small differences in expected returns, even though these differences are likely due to estimation error. The usual pathologies of mean-variance optimization, i.e. extreme and unstable portfolio weights as well as a large discrepancy between in-sample and out-of-sample performance, are said to follow from this error-maximizing property. If this is a proper explanation, then one would expect Bayesian approaches, which account for uncertainty in the parameter estimates, to resolve those pathologies. The Bayesian literature however concluded that accounting for parameter uncertainty with the standard noninformative prior does not meaningfully improve portfolio decisions.

This paper elaborates on the fact that the specification of a reasonable prior for Bayesian portfolio selection is a nontrivial task. Priors which seem innocuous at first glance can readily imply strong and unreasonable prior information about economically relevant parameter transformations. We show that the standard noninformative prior for Bayesian portfolio selection effectively rules out all parameter combinations which lead to a reasonable risk-return tradeoff. It does so by implying very strong and unreasonable prior information about the achievable Sharpe ratio. This has critical implications for mean-variance optimization as it suggests that high expected returns can be obtained cheaply, i.e. that they do not come with a reasonable amount of risk.

The reparametrization developed in this article allows us to specify a prior which is flat in the achievable Sharpe ratio. We do not argue that this is, in any sense, the "best" possible specification of a prior for portfolio selection or that this prior specification should be considered noninformative. It however resolves a major problem of the standard noninformative prior by being reasonably noninformative about the achievable Sharpe ratio. Recent advances in MCMC methods allow us to investigate the implications of this prior for large-scale portfolio decisions. The results from a simulation study and from two empirical data sets suggest that Bayesian portfolio selection with the Sharpe ratio neutral prior does not encounter the usual pathologies of unconstrained mean-variance optimization. It leads to stable and well-diversified portfolio weights and to in-sample performances which are usually close to their out-of-sample counterparts. Accounting for parameter uncertainty in the Bayesian framework might meaningfully improve portfolio decisions after all. It is just not trivial to specify a reasonable prior in this context.

We hope that this article can serve as a reference point for a literature that closer examines the implications of different priors in the portfolio selection context. In the past decades, priors where often chosen due to their analytical tractability or because they imply full conditionals of the posterior which are easy to sample from. Hamiltonian Monte Carlo does not require simple full conditionals and thus allows for much more flexibility in the specification of the prior. The Hamiltonian Monte Carlo implementation in Stan, the software which was used to produce many results of this article, makes even large-scale Bayesian portfolio selection with nonstandard priors computationally feasible. This enables researchers to design priors which specify adequate prior information in the portfolio selection context and allows them to investigate their implications for portfolio decisions. The results about the Jeffreys prior and the development and analysis of the Sharpe ratio neutral prior in this article will hopefully stimulate further work in this branch of literature.

Appendix A: Optimal portfolio weights under the Jeffreys prior

The conjugate prior for the multivariate normal likelihood is the normal-inverse-Wishart prior

$$\Sigma \sim \mathcal{IW}_{\nu_0}\Big(\Lambda_0^{-1}\Big)$$
$$\mu|\Sigma \sim \mathcal{N}\Big(\mu_0, \frac{1}{\kappa_0}\Sigma\Big). \tag{40}$$

which corresponds to the following density

$$p(\mu,\Sigma) \propto |\Sigma|^{((\nu_0+d)/2+1)} \exp\Big(-\frac{1}{2}\mathrm{tr}\Big(\Lambda_0\Sigma^{-1}\Big) - \frac{\kappa_0}{2}(\mu-\mu_0)'\,\Sigma^{-1}(\mu-\mu_0)\Big).$$

The predictive distribution under this prior follows as

$$R_{T+1} \sim t_{\nu_0+T-N+1}(\mu_c, \Sigma_c) \tag{41}$$

with

$$\mu_c = \frac{\kappa_0\mu_0 + T\hat{\mu}}{\kappa_0 + T} \tag{42}$$

and

$$\Sigma_c = \frac{\kappa_0+T+1}{(\kappa_0+T)(\nu_0+T-N+1)}\Big(\Lambda_0 + C + \frac{\kappa_0 T}{\kappa_0+T}(\hat{\mu}-\mu_0)(\hat{\mu}-\mu_0)'\Big) \tag{43}$$

where $t_{\nu_0+T-N+1}$ denotes a multivariate t-distribution with $\nu_0+T-N+1$ degrees of freedom. From the moments of the multivariate t-distribution, it follows that the predictive distribution has expected returns $\mu_{T+1} = \mu_c$ and covariance matrix $\Sigma_{T+1} = \frac{\nu_0+T-N+1}{\nu_0+T-N-1}\Sigma_c$. Thus, the optimal portfolio weights under the normal-inverse-Wishart prior are

$$\begin{aligned} w_c =& \frac{1}{\gamma}\Sigma_{T+1}^{-1}\mu_{T+1} \\ =& \frac{1}{\gamma}\frac{(\nu_0+T-N-1)(\kappa_0+T)}{\kappa_0+T+1}\Big(\Lambda_0 + T\hat{\Sigma} + \frac{\kappa_0 T}{\kappa_0+T}(\hat{\mu}-\mu_0)(\hat{\mu}-\mu_0)'\Big)^{-1}\hat{\mu} \end{aligned} \tag{44}$$

The Jeffreys prior is obtained in the limit of the normal-inverse-Wishart prior for $\kappa_0 \to 0$, $v_0 \to -1$ and $|\Lambda_0| \to 0$ irrespective of the choice of μ_0 (see Gelman et al.

(2014)). Taking this limit of (44) leads to the optimal portfolio weights under the Jeffreys prior

$$w_{Jeff} = \frac{1}{\gamma}\frac{T-N-2}{T+1}\hat{\Sigma}^{-1}\hat{\mu}. \tag{45}$$

Appendix B: Prior distribution of SR_{max} under the Jeffreys prior

It is known (see e.g. Schott (2005)) that for

$$\mu \sim \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0}\Sigma\right). \tag{46}$$

with $\mu_0 = 0$

$$\kappa_0 \mu'\Sigma^{-1}\mu \sim \chi^2(N). \tag{47}$$

From the relationship between the gamma and the χ^2 distribution it follows that

$$\mu'\Sigma^{-1}\mu \sim \text{gamma}\left(\frac{N}{2}, \frac{2}{\kappa_0}\right). \tag{48}$$

Since $SR_{max} = \sqrt{\mu'\Sigma^{-1}\mu}$ it holds that

$$SR_{max}^2 \sim \text{gamma}\left(\frac{N}{2}, \frac{2}{\kappa_0}\right). \tag{49}$$

Using the fact that square root of a gamma distributed random variable has a Nakagami distribution (see e.g. Zhang (2015)), it follows that the achievable Sharpe ratio is

$$SR_{max} \sim \text{Nakagami}\left(\frac{N}{2}, \frac{N}{\kappa_0}\right) \tag{50}$$

which corresponds to the following density

$$\begin{aligned} p(SR_{max}) &= \Gamma\left(\frac{N}{2}\right)\kappa_0^{\frac{N}{2}} SR_{max}^{N-1}\exp\left(-\frac{\kappa_0}{2}SR_{max}^2\right) \\ &\propto SR_{max}^{N-1}\exp\left(-\frac{\kappa_0}{2}SR_{max}^2\right). \end{aligned} \tag{51}$$

For $\kappa_0 \to 0$ the conditional distribution of μ becomes proportional to a constant. The distribution of the achievable Sharpe ratio given in (51) becomes

$$p(SR_{max}) \propto SR_{max}^{N-1}. \tag{52}$$

Appendix C: Posterior distribution of SR_{max}

The posterior distribution under the normal-inverse-Wishart prior is again normal-inverse-Wishart

$$\Sigma \sim \mathcal{IW}_{\tilde{\nu_0}}\left(\tilde{\Lambda_0}^{-1}\right)$$
$$\mu|\Sigma \sim \mathcal{N}\left(\tilde{\mu_0}, \frac{1}{\tilde{\kappa_0}}\Sigma\right). \tag{53}$$

with

$$\begin{aligned}
\tilde{\mu_0} &= \frac{\kappa_0\mu_0 + T\hat{\mu}}{\kappa_0 + T} \\
\tilde{\kappa_0} &= \kappa_0 + T \\
\tilde{\nu_0} &= \nu_0 + T \\
\tilde{\Lambda_0} &= \Lambda_0 + T\hat{\Sigma} + \frac{\kappa_0 T}{\kappa_0 + T}\left(\hat{\mu} - \mu_0\right)\left(\hat{\mu} - \mu_0\right)'
\end{aligned} \tag{54}$$

and

$$\Sigma_c = \frac{\kappa_0 + T + 1}{\left(\kappa_0 + T\right)\left(T + 2\right)}\left(I + T\hat{\Sigma} + \frac{\kappa_0 T}{\kappa_0 + T}\hat{\mu}\hat{\mu}'\right). \tag{55}$$

The Jeffreys prior is obtained in the limit of this prior for $\kappa_0 \to 0$, $v_0 \to -1$ and $|\Lambda_0| \to 0$ irrespective of the choice of μ_0. Thus, the posterior distribution under the Jeffreys prior is

$$\Sigma \sim \mathcal{IW}_{T-1}\left(\frac{1}{T}\hat{\Sigma}^{-1}\right)$$
$$\mu|\Sigma \sim \mathcal{N}\left(\hat{\mu}, \frac{1}{T}\Sigma\right). \tag{56}$$

These results can also be found in Gelman et al. (2014).

References

Avramov, Doron, and Guofu Zhou, 2010, Bayesian portfolio analysis, *Annual Review of Financial Economics* 2, 25–47.

Barnard, John, Robert McCulloch, and Xiao-Li Meng, 2000, Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage, *Statistica Sinica* 10, 1281–1311.

Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein, 1979, *Estimation risk and optimal portfolio choice.* (North-Holland New York).

Box, George E.P., and George C. Tiao, 1973, *Bayesian inference in statistical analysis.* (John Wiley and sons, inc.).

Cochrane, John H., 2001, *Asset pricing.* (Princeton University Press).

Connor, Gregory, Lisa Goldberg, and Robert Korajczyk, 2010, *Portfolio risk analysis.* (Princeton University Press).

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?, *Review of Financial Studies* 22, 1915–1953.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin, 2014, *Bayesian data analysis (3rd ed.).* (Chapman & Hall/CRC).

Geweke, John, and Guofu Zhou, 1996, Measuring the pricing error of the arbitrage pricing theory, *Review of Financial Studies* 9, 557–587.

Jobson, J.D., and Bob M. Korkie, 1981, Performance hypothesis testing with the Sharpe and Treynor measures, *The Journal of Finance* 36, 889–908.

Kandel, S., Robert McCulloch, and Robert F. Stambaugh, 1995, Bayesian inference and portfolio efficiency, *Review of Financial Studies* 8, 1–53.

Klein, Roger W., and Vijay S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215–231.

Lau, Brian, 2015, MatlabStan: the MATLAB interface to Stan, *http://mc-stan.org/matlab-stan.html.*

MacKinlay, Craig A., 1995, Multifactor models do not explain deviations from the CAPM, *Journal of Financial Economics* 38, 3–28.

MacKinlay, Craig. A., and Lubos Pastor, 2000, Asset pricing models: implications for expected returns and portfolio selection, *Review of Financial Studies* 13, 883–916.

Memmel, Christoph, 2003, Performance hypothesis testing with the Sharpe ratio, *Finance Letters* 1.

Pástor, Lubos, 2000, Portfolio selection and asset pricing models, *The Journal of Finance* 55, 179–223.

Pástor, Lubos, and Robert F. Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.

Roll, Richard, 1977, A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129–176.

Schott, James R., 2005, *Matrix analysis for statistics.* (Wiley).

Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.

Stan Developement Team, 2015, Stan: A C++ library for probability and sampling, Version 2.8.0, *http://mc-stan.org/.*

Tipping, Michael E., and Christopher M. Bishop, 1999, Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B* 61, 611–622.

Tu, Jun, and Guofu Zhou, 2010, Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty, *Journal of Financial and Quantitative Analysis* 45, 959–986.

Tu, Jun, and Guofu Zhou, 2011, Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies, *Journal of Financial Economics* 99, 204–215.

Zellner, Arnold, and Karuppan V. Chetty, 1965, Prediction and decision problems in regression models from the Bayesian point of view, *Journal of the American Statistical Association* 60, 608–616.

Zhang, Keith Q., 2015, *Wireless communications: Principles, theory and methodology.* (John Wiley & Sons).