Robust Bayesian Portfolio Choices

Evan W. Anderson
Northern Illinois University

Ai-Ru (Meg) Cheng
Northern Illinois University

We propose a Bayesian-averaging portfolio choice strategy with excellent out-of-sample performance. Every period a new model is born that assumes means and covariances are constant over time. Each period we estimate model parameters, update model probabilities, and compute robust portfolio choices by taking into account model uncertainty, parameter uncertainty, and non-stationarity. The portfolio choices achieve higher out-of-sample Sharpe ratios and certainty equivalents than rolling window schemes, the 1/N approach, and other leading strategies do on a majority of 24 datasets. (*JEL* G11, C11, D81)

Received September 8, 2012; accepted October 18, 2015 by Editor Pietro Veronesi.

We propose a robust Bayesian-averaging (BA) portfolio choice rule that achieves higher out-of-sample certainty equivalents and Sharpe ratios than many other popular rules do on a majority of daily, weekly, monthly, and artificial data sets. Our portfolio choice rule assumes an ever growing number of models that possibly describe asset returns. Each model is simple, and each specifies that asset returns have constant means and covariances over time. Every period in our sample, a new model is born. For example, when 20,000 periods of historical returns are observed, there are 20,000 possible models for asset returns. Old models never die, but can become irrelevant over time if their predictions are unsatisfactory. Every period, when new information becomes available, we estimate parameters for each model and compute statistically optimal probabilities that each model is correct. Because the future may be different from the past, we adopt portfolio choice rules that are robust to model misspecification.

We evaluate daily, weekly, and monthly out-of-sample performance on 24 well-known data sets from the Center for Research in Security Prices (CRSP)

We especially thank Pietro Veronesi for numerous detailed and valuable comments that greatly improved the paper. We thank Ren Cheng, Ron Gallant, Lorenzo Garlappi, Lars Peter Hansen, Yulei Luo, Thomas J. Sargent, George Slotsve, Ruey S. Tsay, and Yan Xu; seminar participants at the University of Hong Kong and Northern Illinois University; conference participants at the 2014 Joint Statistical Meetings; and an anonymous referee for helpful comments. We also benefited from comments at the poster session for the conference on Ambiguity and Robustness in Macroeconomics and Finance at the Becker-Friedman Institute. Send correspondence to Evan W. Anderson, 515 Zulauf, Department of Economics, Northern Illinois University, DeKalb, IL 60115, USA; telephone: (630) 450-0533. E-mail: ewanderson@niu.edu.

doi:10.1093/rfs/hhw001 Advance Access publication January 27, 2016

and Ken French's Web site. We show that the robust BA portfolio choice rule statistically outperforms rolling window approaches, 1/N, Jorion's Bayes-Stein procedure, Kan and Zhou's three-fund rule, and other leading rules. To compare performance, we derive asymptotic standard errors for Sharpe ratios and certainty equivalents that are applicable under serial correlation by applying generalized method of moments (GMM). The standard errors improve on the asymptotic standard errors usually reported in the literature, which assume returns are independent and identically distributed over time. We also provide results for simulated data sets and show the superior performance of the robust BA strategy when there is a possibility of regime switches.

A key motivation for the robust BA strategy is that historical means often predict future quantities better than other more complicated forecasts.[1] However, in standard forecasting methods, such as rolling window approaches, the amount of data used to predict future means is *a-priori* determined. In our approach, excess returns are forecasted exclusively by past sample means and variances without preselecting the window size. Investors do not know which window size is best and use Bayesian statistical methods to determine the probabilities over all windows. The probabilities change over time, as new information arrives, and typically are positive for many different window sizes. Our approach is less susceptible to look-ahead bias than non-Bayesian methods are because we use only previously available information to form weak priors over window sizes.[2]

Our robust BA rule builds on a large body of literature. Previous research typically focuses on optimal portfolio choices in a formally specified economic environment. There has been little progress made on portfolio choices that work well out of sample. Although, so called optimal choices are by construction the best choices in a given sample, they tend to perform poorly out of sample.

Since Markowitz's seminal work on optimal portfolio choices, the dominant paradigm applies and extends his approach. Markowitz assumes investors have preferences over the mean and variance of asset returns. When there exists a risk-free rate, the solution to Markowitz's problem is to invest the fractions

$$\frac{1}{\theta}\Sigma^{-1}\mu \tag{1}$$

in risky assets where θ is a measure of risk aversion, Σ is an $n \times n$ covariance matrix of asset returns, and μ is an $n \times 1$ vector of mean excess returns. The portfolio choice rule in Equation (1) is widely used and also is the optimal choice rule under several other sets of assumptions. Merton (1973) shows in

---

[1] Welch and Goyal (2008) demonstrate that historical means better predict the equity premium than other forecasts.

[2] In the last decade, there has been a resurgence of interest in using distributed lag approaches to forecast means and variances where often parameters in an *a-priori* specified flexible functional form are estimated to determine the weights on past observations. For example, see Ghysels, Santa-Clara, and Valkanov (2005). Our approach is similar, except we do not impose a functional form and instead use Bayesian statistical methods to determine the weights.

continuous time, with power utility, constant means, and constant covariances, that this portfolio choice rule is optimal. It is also well known that the same portfolio choice rule is optimal when agents have exponential utility in discrete time and excess asset returns are normally distributed. Samuelson (1970) shows that Equation (1) is an approximation for the optimal portfolio choice rule for much more general utility functions and assumptions about the stochastic processes governing asset returns. Campbell and Viceira (2002) apply a similar approximation to a wide range of examples. See Meucci (2005) for extensions and applications of Markowitz's problem.[3]

Applications of Markowitz's approach often estimate constant means and covariances for asset returns from historical data. The estimates are typically assumed to be exactly correct and then plugged into the portfolio choice rule in Equation (1) to obtain both in-sample and out-of-sample portfolio choices. In sample, if means and covariances are constant then this strategy is by design optimal (for mean-variance preferences) and cannot be improved upon. Out of sample, if means and covariances do not change from their in-sample values then it again is optimal. However, in practice, means and covariances typically change through time, and this strategy usually performs extremely poorly out of sample.

Several studies have shown that simple naive portfolio strategies perform better than Markowitz's rule and other so called optimal choices, out of sample. DeMiguel, Garlappi, and Uppal (2009) show on several monthly datasets that the naive strategy, which invests an equal amount in each risky asset, called 1/N, generally performs better out of sample than the leading sophisticated models of optimal choice do. The 1/N rule does not involve optimization and instead invests the fraction 1/N in each available risky asset, every period regardless of future expectations. Tu and Zhou (2011) demonstrate that combining 1/N with leading models leads to additional gains.[4]

The poor out-of-sample performance of Markowitz's and related approaches happens for several reasons. One reason is that estimates of means, variances, and other parameters are often imprecise, but the imprecision is usually ignored when computing optimal portfolio choices. A second reason is that although economic models are only an approximation to reality, investors' doubts about their models are usually ignored when making portfolio choices. For example, it is often assumed that asset returns have constant variances over time but it is possible that this assumption is false and variances actually follow a more complicated process such as a GARCH (Generalized Autoregressive Conditional Heteroskedasticity) process. A third reason is that the world may be always changing in new and unpredictable ways, and it is difficult to forecast future asset returns using past information.

[3] See Brandt (2010) for a more general survey of portfolio choice problems.

[4] See Duchin and Levy (2009) for a succinct overview of the 1/N strategy.

We propose methods that improve out-of-sample performance by taking into account parameter uncertainty, model uncertainty, and non-stationarity. We deviate from the standard approach by assuming there are large number of possible models of the world, and adopt a robust BA approach where investors constantly update the probabilities of each model. Investors are fully aware that their models may be misspecified despite the richness of their models.

Many other applications of Bayesian methods have been proposed. See, for example, Jorion (1986), Kandel and Stambaugh (1996), Xia (2001), Kan and Zhou (2007), and Tu and Zhou (2010), and for important contributions where there is one model but parameters are uncertain and updated over time. See Avramov (2002), Cremers (2002), and Tu and Zhou (2004) for contributions where there is more than one model.[5] In our approach there are a large number of possible models, whereas most previous applications of Bayesian averaging have used a small number of models.

Following Hansen and Sargent (1995); Hansen, Sargent, and Tallarini (1999); Anderson, Hansen, and Sargent (2003); Hansen and others (2006);and Hansen and Sargent (2007b), we allow agents to worry about misspecification by considering perturbations to the probability density of asset returns that decrease utility. Maenhout (2004) and Uppal and Wang (2003) apply this approach to portfolio choice problems. Dow and Werlang (1992); Garlappi, Uppal, and Wang (2007); and Epstein and Schneider (2010) have considered portfolio choice models with alternative formulations of model uncertainty.

1. Overview

This section informally provides an overview of nonrobust and robust BA portfolio choices.

1.1 The models

We begin by describing the information and models available at an hypothetical date $t-1$. At date $t-1$, there are $t-1$ possible models for the excess returns of risky assets over the risk-free asset. The mth model (where $1 \leq m \leq t-1$) assumes excess returns have constant means and covariances between dates m and $t-1$. For example, Model 1 has a constant mean and covariance between dates 1 and $t-1$; and Model 2 has a constant mean and covariance between dates 2 and $t-1$. The last model is model $t-1$ with a constant mean and covariance that applies only at date $t-1$. The models use weak prior information from earlier asset returns.

Investors are uncertain if any of the models are correct. All models could be correct, some models could be correct, or no models could be correct. To use standard statistical inference, we initially make the assumption that at least one

[5] For a summary of Bayesian portfolio choice studies, see Avramov and Zhou (2010) and Pastor and Veronesi (2009).

of the available models is correct. However, when making optimal portfolio choices, we consider the possibility that none of the existing models are correct. We use the following notation for model probabilities:

Definition: When there are s possible models, $P_s(m|\mathcal{F}_{t-1})$ is the probability that model m is correct, and no older model $q<m$ is correct, conditioned on information available at time $t-1$ and the assumption that at least one of the s models is correct.

In our setup, the s subscript is important because the probability of model m depends upon how many other possible models exist. When there are currently $t-1$ possible models, $s=t-1$.

The $t-1$ models have an important logical structure: if a model is correct, then all younger models also are correct. For example, if Model 2 is correct, then Models $3,4,\ldots t-1$ also must be correct because if means and covariances are constant at dates $2,3,4,\ldots t-1$, then they also must be constant at dates $3,4,5,\ldots t-1$ (so that Model 3 is right). It further follows that means and covariances are constant at dates $4,5,6\ldots t-1$ (so that Model 4 is right) and so on.

The following example illustrates the overlapping probabilities. Lets assume $t-1=4$ and Model 1 is correct with probability 10%, Model 2 is correct with probability 50%, Model 3 is correct with probability 80%, and Model 4 is correct with probability 100%. Then the probability that Model 2 is correct and Model 1 is not correct is 40%; the probability that Model 3 is correct and Models 1 and 2 are not correct is 30%; and the probability that Model 4 is correct and Models 1, 2, and 3 are not correct is 20%. In this case we set

$$P_4(1|\mathcal{F}_4)=0.10, \quad P_4(2|\mathcal{F}_4)=0.40, \quad P_4(3|\mathcal{F}_4)=0.30, \quad \text{and} \quad P_4(4|\mathcal{F}_4)=0.20.$$

To conserve words we often say $P_4(3|\mathcal{F}_4)$ is the probability that Model 3 is correct (when further clarification is not needed), but it is important to remember that this really is the probability that Model 3 is correct and Models 1 and 2 are not correct; when it is assumed that at least one of the models $\{1,2,3,4\}$ is correct.

1.2 Updating probabilities and parameters

At date t, investors update probabilities and parameters in two stages. First, a new model, model t, is born and the probabilities of all models are adjusted to $P_t(m|\mathcal{F}_{t-1})$ for its inclusion. The new model has weak prior information that says the mean and variance of each asset's returns are equal to the common past mean and the common past variance of all assets. The adjusted probabilities take into account that the new model has some positive probability, and thus the probabilities of older models are reduced. See Sections 3.1 and 8 for discussions of several ways to adjust probabilities.

Second, investors observe excess returns at date t, denoted R_t, and update the parameters and probabilities of all models using Bayes rules based on this new information. The probability of each model, m, is updated from $P_t(m|\mathcal{F}_{t-1})$ to $P_t(m|\mathcal{F}_t)$. The available models are now: Model 1 with a constant mean and covariance between dates 1 and t, Model 2 with a constant mean and covariance between dates 2 and t, and so on. The last model is now model t with a constant mean and covariance that applies only at date t.

1.3 Expectations and optimal portfolio choices

Investors form expectations of future returns using the probabilities $P_t(m|\mathcal{F}_t)$ over models and estimates of means and covariances in each model. Optimal portfolio choices, ϕ_t, at time t are calculated when investors have standard mean-variance preferences and robust mean-variance preferences. Out-of-sample portfolio returns are observed at time $t+1$. See Sections 4 and 5 for the details.

1.4 Timeline

In summary, the timeline of events is as follows:

Date		**Events**
$t-1$	▹	There are $t-1$ existing models that possibly describe past asset returns. Each model m has probability $P_{t-1}(m|\mathcal{F}_{t-1})$.
t	▹	A new model, model t, is born and probabilities for all models are adjusted for its inclusion. Each model m now has probability $P_t(m|\mathcal{F}_{t-1})$.
	▹	Excess returns at time t, R_t, are observed. Parameters for each model are updated. Probabilities are updated so that each model m now has probability $P_t(m|\mathcal{F}_t)$.
	▹	Expectations of excess returns at time $t+1$, R_{t+1}, are formed.
	▹	Optimal portfolio choices, ϕ_t, at time t are computed.
$t+1$	▹	Excess returns at time $t+1$, R_{t+1}, are observed and excess portfolio returns $\phi_t' R_{t+1}$ are realized.

2. Models

We now formally describe the available models. At period t, there are $t-1$ existing models (that were also alive at time $t-1$) and one new model of an n dimensional vector of excess returns.

2.1 Existing models

Let M_{t-1} be the collection of all possible models at time $t-1$. Each possible model $m \in M_{t-1}$ assumes the n dimensional vector of available excess returns follows a multivariate normal distribution with a constant mean μ and a constant covariance matrix Σ:

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

The constant mean and covariance differ across models and are not known perfectly by investors. Over time, agents learn about the mean and covariance by observing actual returns. At time $t-1$, for model $m \in M_{t-1}$, beliefs about mean returns, μ, and the covariance of returns, Σ, are:

$$\Sigma \sim \mathcal{IW}(\Lambda_{m,t-1}, \nu_{m,t-1}), \tag{2}$$

$$\mu | \Sigma \sim \mathcal{N}(\mu_{m,t-1}, \Sigma / \kappa_{m,t-1}). \tag{3}$$

The notation $\mathcal{IW}(\Lambda_{m,t-1}, \nu_{m,t-1})$ denotes the inverse-Wishart distribution with the inverse scale matrix $\Lambda_{m,t-1}$ and degrees of freedom $\nu_{m,t-1}$. The subscript "$m, t-1$" indicates these are the beliefs in the particular model m at time $t-1$. The variables $\mu_{m,t-1}$ and $\kappa_{m,t-1}$ give the mean and precision of beliefs about μ. It follows that the joint prior beliefs of μ and Σ are normal-inverse-Wishart:

$$\mu, \Sigma \sim \mathcal{NIW}(\mu_{m,t-1}, \kappa_{m,t-1}, \Lambda_{m,t-1}, \nu_{m,t-1}).$$

As discussed in later sections, beliefs evolve over time but maintain the same functional form. Although it is standard to capture the evolution of beliefs by tracking the parameters $(\mu_m, \kappa_m, \Lambda_m, \nu_m)$ over time, this is inconvenient for economic interpretations. It is more convenient to define

$$\Sigma_{m,t-1} = \left(\frac{1}{\delta_{m,t-1}} \right) \Lambda_{m,t-1}, \qquad \delta_{m,t-1} = \nu_{m,t-1} - n - 1,$$

and track the evolution of the parameters $(\mu_m, \kappa_m, \Sigma_m, \delta_m)$ because the mean of an $\mathcal{IW}(\Lambda_{m,t-1}, \nu_{m,t-1})$ random variable is $\Sigma_{m,t-1}$ and the value of $\delta_{m,t-1}$ is easily comparable across datasets.[6]

2.2 New models

In our BA method, each period a new model is born and endowed with a prespecified prior. This section describes the prior beliefs of the new model born at time t, before time t returns, R_t, are observed. The new model is labeled model t and assumes the $n \times 1$ vector of available excess returns follows a multivariate normal distribution with a constant mean μ and a constant covariance matrix $\Sigma: R_t \sim \mathcal{N}(\mu, \Sigma)$.

[6] For the inverse-Wishart distribution to have a finite mean, it is necessary for $\delta_{m,t-1}$ to be positive or, equivalently, it is necessary for $\nu_{m,t-1}$ to be greater than $n+1$. Because the later requirement depends on the number of assets, it is not easily comparable across datasets.

Prior beliefs of the agents about the mean and covariance are normal-inverse-Wishart:

$$\mu, \Sigma \sim \mathcal{NIW}(\mu_{t,t-1}, \kappa_{t,t-1}, \Lambda_{t,t-1}, \nu_{t,t-1})$$

with $\Lambda_{t,t-1} = \delta_{t,t-1}\Sigma_{t,t-1}$ and $\nu_{t,t-1} = \delta_{t,t-1} + n + 1$. The subscript $t, t-1$ indicates these are the beliefs in the particular model t at time $t-1$. $\mu_{t,t-1}$ gives the best guess of mean returns; $\Sigma_{t,t-1}$ gives the best guess of covariances; $\delta_{t,t-1}$ captures the confidence of beliefs about covariances; and $\kappa_{t,t-1}$ captures the confidence of beliefs about means.

In our applications, we set the priors for model t as:

$$\mu_{t,t-1} = \bar{\mu}_{t-1}\mathbf{1}_n, \qquad \kappa_{t,t-1} = 1, \qquad \Sigma_{t,t-1} = \bar{\lambda}_{t-1}\mathbf{I}_n, \qquad \delta_{t,t-1} = 1,$$

where the scalar parameters $\bar{\mu}_{t-1}$ and $\bar{\lambda}_{t-1}$ are time-varying and are the common past sample means and variances across all assets. We set

$$\bar{\mu}_{t-1} = \frac{1}{n}\sum_{i=1}^{n}\bar{\mu}_{i,t-1}, \qquad \bar{\lambda}_{t-1} = \frac{1}{n}\sum_{i=1}^{n}\bar{\lambda}_{i,t-1},$$

where

$$\bar{\mu}_{i,t-1} = \frac{1}{t-1}\sum_{s=1}^{t-1}R_{i,s}, \qquad \bar{\lambda}_{i,t-1} = \frac{1}{t-2}\sum_{s=1}^{t-1}(R_{i,s} - \bar{\mu}_{i,t-1})^2$$

and $R_{i,s}$ is the excess return on asset i between periods $s-1$ and s. Thus the prior beliefs about mean excess returns and variances are identical for each asset. The prior beliefs about the covariance between any two assets is zero. Because $\kappa_{t,t-1} = 1$ and $\delta_{t,t-1} = 1$, the priors for means and variances are very weak.[7]

3. Updating Models and Probabilities

In this section, we update model probabilities and parameters when a new model is born and when new information arrives.

3.1 Updating model probabilities when a new model is born

We describe several methods for updating the probabilities of models after model t is born but before time t information is observed. By updating probabilities, we are choosing priors for next period. In the basic version of the BA algorithm, described in Sections 2 through 5, investors *a priori* choose one particular prior rule and use that rule at all dates. In Section 8, we formally

[7] Because for Model 1 at time one there is no history of data, we set $\bar{\mu}_0 = 0$. To use historical data to estimate variances, we need at least two periods of data, so we set both $\bar{\lambda}_0$ and $\bar{\lambda}_1$ to 0.0001 with daily data, 0.0005 with weekly data, and 0.002 with monthly data. In later sections, we use a burn-in period, so the prespecified values for $\bar{\mu}_0$, $\bar{\lambda}_0$, and $\bar{\lambda}_1$ have a negligible effect on our results.

embed prior selection inside optimization problems and allow investors to use different priors at different dates.

We begin by discussing two simple priors, the $1/t$ and equal-weighted priors. We then describe the minimal restrictions for appropriate priors and explore the sharing and power priors that satisfy the minimal restrictions and build on the simple priors.

3.1.1 The $1/t$ prior. One simple updating method, which we call $1/t$, sets the probability of model t at time t to be $1/t$ and proportionally discounts the probability of all other models:

$$P_t(m|\mathcal{F}_{t-1})=\begin{cases}\left(1-\frac{1}{t}\right)P_{t-1}(m|\mathcal{F}_{t-1}) & \text{if } m<t\\ \frac{1}{t} & \text{if } m=t.\end{cases}$$

This puts the probability of the new model on equal footing with the average probability of the older models.

3.1.2 Equal-weighted prior. Another simple updating method assigns the probability $1/t$ to all models at time t:

$$P_t(m|\mathcal{F}_{t-1})=\frac{1}{t}$$

for $m \leq t$. Unlike other priors, the equal-weighted prior ignores all observed past information on excess returns.

3.1.3 Minimum restrictions on priors. The $1/t$ and equal-weighted priors are highly restrictive and do not allow for many other possible prior selection rules. To understand reasonable minimal restrictions on priors, it is useful to recall if model q is correct then all younger models m, such that $m>q$, also must be correct. This suggests that we can set the priors for new models by transferring probabilities from older models to newer models so that the sum of probabilities of all models born at date m and earlier does not exceed the sum of their previous probabilities. More formally,

$$\sum_{q=1}^{m}P_t(q|\mathcal{F}_{t-1})\leq\sum_{q=1}^{m}P_{t-1}(q|\mathcal{F}_{t-1}) \tag{4}$$

for all $m<t$. No restrictions are placed on the probability of the newest model, model t, other than probabilities must sum to one. The oldest model, Model 1, can not gain probability but all other older models, $m<t$, may lose or gain probability as long as Equation (4) holds.

In Sections 3.1.4 and 3.1.5, we parameterize two different classes of fixed prior rules that satisfy the minimal restrictions.

3.1.4 Sharing prior. The sharing prior satisfies the minimal restrictions and takes a concrete stand on the amount of probability transferred to newer models. Let α denote the fraction of its previous probability that each model shares equally with itself and newer models. Each model also retains the fraction $(1-\alpha)$ of its previous probability. More formally for all models $m \in M_t$, we set:

$$P_t(m|\mathcal{F}_{t-1}) = \begin{cases} (1-\alpha)P_{t-1}(m|\mathcal{F}_{t-1}) + \alpha\left[\sum_{q=1}^{m}\left(\frac{1}{t-q+1}\right)P_{t-1}(q|\mathcal{F}_{t-1})\right] & \text{if } m<t \\ \alpha\left[\sum_{q=1}^{t-1}\left(\frac{1}{t-q+1}\right)P_{t-1}(q|\mathcal{F}_{t-1})\right] & \text{if } m=t \end{cases}$$

for $0<\alpha\leq 1$. When $\alpha=1$, each older model transfers all of its probability equally to itself and all newer models. In this case, the newest model, model t, and the next newest model, model $t-1$, have the same probability; and newer models have higher probabilities than older models. When $\alpha=0$, the newest model receives zero probability and all older models retain all of their previous probability. Because this would cause the newest model to have zero probability throughout its life, we assume α is strictly greater than zero.

Many of the examples in this study use the sharing prior with $\alpha=1$. We call this the *perfect sharing prior* because older models share probabilities equally with newer models. The perfect sharing prior is similar to the equal-weighted prior but previous period posterior probabilities are shared equally with only newer models. If perfect sharing also shared equally with older models, then it would be identical to the equal-weighted prior.

3.1.5 Power prior. The power prior satisfies the minimal restrictions and generalizes the $1/t$ rule presented in Section 3.1.1. Let β, such that $0<\beta\leq 1$, be a parameter that determines the discounting of older models:

$$P_t(m|\mathcal{F}_{t-1}) = \begin{cases} \left[\frac{\beta^{t-m}P_{t-1}(m|\mathcal{F}_{t-1})}{\sum_{q=1}^{t-1}\beta^{t-q}P_{t-1}(q|\mathcal{F}_{t-1})}\right]\left(\frac{t-1}{t}\right) & \text{if } m<t \\ \frac{1}{t} & \text{if } m=t. \end{cases} \tag{5}$$

When $\beta=1$, this rule is equivalent to the $1/t$ rule. As β decreases, the probabilities of older models fall faster, in percentage terms, than the probabilities of newer models:

$$\frac{P_t(m-1|\mathcal{F}_{t-1})}{P_{t-1}(m-1|\mathcal{F}_{t-1})} = \beta\left[\frac{P_t(m|\mathcal{F}_{t-1})}{P_{t-1}(m|\mathcal{F}_{t-1})}\right] \leq \frac{P_t(m|\mathcal{F}_{t-1})}{P_{t-1}(m|\mathcal{F}_{t-1})},$$

when $m<t$. The newest model always receives $1/t$ probability, regardless of the value of β. The summation in the denominator and the

$(t-1)/t$ coefficient in Equation (5) normalize probabilities so that they sum to one.[8]

3.2 Updating parameters when new information arrives

At every period t, investors observe excess returns R_t but no other new information useful for predicting future returns. Before observing R_t, investors' beliefs about μ and Σ, in model m, are normal-inverse-Wishart:

$$\mu, \Sigma \sim \mathcal{NIW}(\mu_{m,t-1}, \kappa_{m,t-1}, \delta_{m,t-1}\Sigma_{m,t-1}, \delta_{m,t-1}+n+1).$$

Investors update their beliefs about μ and Σ after seeing R_t so that

$$\mu, \Sigma \sim \mathcal{NIW}(\mu_{m,t}, \kappa_{m,t}, \delta_{m,t}\Sigma_{m,t}, \delta_{m,t}+n+1), \tag{6}$$

where

$$\mu_{m,t} = \frac{\kappa_{m,t-1}\mu_{m,t-1}+R_t}{\kappa_{m,t}}, \tag{7a}$$

$$\Sigma_{m,t} = \frac{\delta_{m,t-1}\kappa_{m,t}\Sigma_{m,t-1}+\kappa_{m,t-1}(R_t-\mu_{m,t-1})(R_t-\mu_{m,t-1})'}{\delta_{m,t}\kappa_{m,t}}, \tag{7b}$$

$\kappa_{m,t}=\kappa_{m,t-1}+1$, and $\delta_{m,t}=\delta_{m,t-1}+1$. In Equation (7a), the mean beliefs in model m at time t are a weighted average of the prior mean (the mean at time $t-1$) and the new observation R_t. More weight is placed on the prior mean as $\kappa_{m,t-1}$ increases. In Equation (7b), the mean beliefs about Σ are approximately a weighed average of prior beliefs and the squared difference of returns from prior means. Larger values of $\delta_{m,t-1}$ increase the weight on prior beliefs about Σ.[9]

Because beliefs at time t have the same functional form as beliefs at time $t-1$, a normal-inverse-Wishart prior is a conjugate prior when means and covariances are constant over time. Because in each model initial beliefs about the mean and variances of returns are normal-inverse-Wishart, the beliefs at every future date also will be normal-inverse-Wishart.

3.3 Updating model probabilities when new information arrives

Before observing excess returns, R_t, investors believe model m is correct with probability $P_t(m|\mathcal{F}_{t-1})$. After seeing R_t, investors update the probability that model m is correct, $P_t(m|\mathcal{F}_t)$, using Bayes rule:[10]

$$P_t(m|\mathcal{F}_t) = \frac{L(R_t|m,\mathcal{F}_{t-1})\,P_t(m|\mathcal{F}_{t-1})}{\sum_{m\in M_t} L(R_t|m,\mathcal{F}_{t-1})\,P_t(m|\mathcal{F}_{t-1})}.$$

---

8 Its straightforward to verify that this rule satisfies the minimal restrictions stated in Section 3.1.3.

9 The updating formulas in this section are well known. For example, see Section 10.2 of Bishop (2006).

10 As discussed earlier, $P_t\left(m|\mathcal{F}_{t-1}\right)$ is the probability that model m is correct and no other model $q, q<m$ is correct. Because under this interpretation, the probabilities of the models are disjoint, Bayes rule is the statistically optimal way to update probabilities when new information is observed.

This can be easily computed because the likelihood of observing R_t, given information at time $t-1$, in model m is:

$$L(R_t|m,\mathcal{F}_{t-1}) = \frac{\kappa_{m,t-1}^{n/2}|\Lambda_{m,t-1}|^{\nu_{m,t-1}/2}\Gamma_n(\nu_{m,t}/2)}{\pi^{n/2}\kappa_{m,t}^{n/2}|\Lambda_{m,t}|^{\nu_{m,t}/2}\Gamma_n(\nu_{m,t-1}/2)}, \tag{8}$$

where $\Lambda_{m,t} = \delta_{m,t}\Sigma_{m,t}$, $\nu_{m,t} = \delta_{m,t} + n + 1$, and $\Gamma_n(x)$ denotes the multivariate gamma function.[11]

4. Nonrobust Portfolio Choices

Assume there exists a risk-free rate between times t and $t+1$, denoted R_{ft+1}, whose value is known at time t. As in earlier sections, let R_{t+1} be the n-dimensional vector of excess returns on the risky assets that are realized at time $t+1$ and unknown at time t. The excess returns are formed by subtracting the nominal risk-free rate from the nominal return on the risky assets. During each period, t, investors compute optimal portfolio weights, ϕ_t, on all available assets and earn excess returns $\phi_t' R_{t+1}$ the following period.

4.1 Mean-variance portfolio choices

As discussed in Markowitz (1952) and many subsequent papers, an investor with mean-variance preferences maximizes

$$E\left(\phi_t' R_{t+1} + R_{ft+1}|\mathcal{F}_t\right) - \frac{\theta}{2} V\left(\phi_t' R_{t+1} + R_{ft+1}|\mathcal{F}_t\right) \tag{9}$$

by choice of portfolio weights ϕ_t, where E denotes expectation, V denotes variance, and θ is a measure of risk aversion. We can write Objective (9) as

$$\phi_t'\hat{\mu}_t + R_{ft+1} - \frac{\theta}{2}\phi_t'\hat{\Sigma}_t\phi_t,$$

where $\hat{\mu}_t$ and $\hat{\Sigma}_t$ are the investor's beliefs about conditional means and variances. The optimal portfolio weights are

$$\phi_t = \frac{1}{\theta}\hat{\Sigma}_t^{-1}\hat{\mu}_t.$$

4.2 Bayesian-averaging portfolio choices

We compute mean-variance portfolio choices when expectations are formed using Bayesian averaging. At time t, investors who use the BA algorithm believe the distribution of excess returns is $R_{t+1} \sim \mathcal{N}(\mu, \Sigma)$ where there are

[11] The likelihood formula Equation (8) is well known. See Murphy (2012).

t possible models of μ and Σ. If model $m \in M_t$ is correct, then investors' best approximation to μ and Σ is:

$$\mu, \Sigma \sim \mathcal{NIW}(\mu_{m,t}, \kappa_{m,t}, \delta_{m,t}\Sigma_{m,t}, \delta_{m,t}+n+1).$$

It follows that investors believe the conditional expectation of excess returns is:

$$\hat{\mu}_t = E(R_{t+1}|\mathcal{F}_t) = \sum_{m \in M_t} \mu_{m,t} P_t(m|\mathcal{F}_t), \tag{10}$$

and the conditional variance of excess returns is:

$$\hat{\Sigma}_t = V(R_{t+1}|\mathcal{F}_t) = \sum_{m \in M_t} \left(\bar{\Sigma}_{m,t} + \mu_{m,t}\mu'_{m,t}\right) P_t(m|\mathcal{F}_t) - \hat{\mu}_t \hat{\mu}'_t, \tag{11}$$

where

$$\bar{\Sigma}_{m,t} = V(R_{t+1}|m, \mathcal{F}_t) = \left(\frac{1+\kappa_{m,t}}{\kappa_{m,t}}\right) \Sigma_{m,t} \tag{12}$$

is the variance of R_{t+1} conditioned on model m. Note that $\bar{\Sigma}_{m,t}$ includes an adjustment for parameter uncertainty in mean returns.

We assume BA investors have mean-variance preferences. Their optimal portfolio weights are

$$\phi_t = \frac{1}{\theta} \hat{\Sigma}_t^{-1} \hat{\mu}_t, \tag{13}$$

where $\hat{\mu}_t$ and $\hat{\Sigma}_t$ are the investor's beliefs about conditional means and variances given in Equations (10) and (11). We call the trading rule in Equation (13), the (nonrobust) BA portfolio strategy.

5. Robust Portfolio Choices

Robust investors believe the world may be nonstationary, are unsure of the future, worry that their best available approximation specification of asset returns is wrong, and consider decision-making procedures that take into account model misspecification.

We begin by writing standard mean-variance preferences as

$$\int \left[\phi'_t \hat{\mu}_t + \phi'_t z_{t+1} + R_{f\,t+1} - \frac{\theta}{2}\left(\phi'_t z_{t+1}\right)^2\right] f(z_{t+1}|\mathcal{F}_t) dz_{t+1}, \tag{14}$$

where expected values are taken with respect to the conditional density of z_{t+1} and where z_{t+1} is defined to be the deviation of R_{t+1} from its conditional mean: $z_{t+1} = R_{t+1} - \hat{\mu}_t$. Objective (14) is equivalent to Objective (9). In standard mean-variance problems, investors assume $f(z_{t+1}|\mathcal{F}_t)$ is the correct probability density function for z_{t+1} even though f could be misspecified in unimaginable ways.

5.1 Robust mean-variance optimization

We formulate a robust version of a mean-variance optimization problem and allow agents to have doubts about future expectations. We formally incorporate concerns, into the investor's problem, that the best approximating density, $f(z_{t+1}|\mathcal{F}_t)$, is misspecified. The investor wants to choose portfolio allocations that will do well even when the approximation is wrong. However, because the approximation is reasonable, the investor focuses on specifications that are close to the approximating specification, and constructs portfolio choices to maximize utility on the worst specification that is close to the approximating specification.

Following Hansen, Sargent, and Tallarini (1999) and Anderson, Hansen, and Sargent (2003), we incorporate concerns for model misspecifications by letting investors solve the following robust problem at time t:

$$\max_{\phi_t}\min_{\varrho_t}\int \varrho_t(z_{t+1})\left[\phi_t'(\hat{\mu}_t+z_{t+1})+R_{ft+1}-\frac{\theta}{2}\left(\phi_t'z_{t+1}\right)^2\right]f(z_{t+1}|\mathcal{F}_t)dz_{t+1}$$
$$+\frac{1}{\tau}\int \varrho_t(z_{t+1})\log\varrho_t(z_{t+1})f(z_{t+1}|\mathcal{F}_t)dz_{t+1} \quad (15)$$

subject to the constraint

$$\int \varrho_t(z_{t+1})f(z_{t+1}|\mathcal{F}_t)dz_{t+1}=1, \quad (16)$$

where $z_{t+1}=R_{t+1}-\hat{\mu}_t$. The robust problem introduces the function ϱ, which alters the investor's perceived density of asset returns. We interpret $f(z_{t+1}|\mathcal{F}_t)$ as the agent's best approximation for the distribution of z_{t+1} and $f(z_{t+1}|\mathcal{F}_t)\varrho_t.(z_{t+1})$ as an alternative distribution for z_{t+1}. Constraint (16) requires that $f(z_{t+1}|\mathcal{F}_t)\varrho_t.(z_{t+1})$ be a probability density function.

In the robust problem, investors penalize perturbations to the density of returns by the relative entropy of alternative distributions with respect to the best approximating distribution:

$$\frac{1}{\tau}E\left[\varrho_t(z_{t+1})\log\varrho_t(z_{t+1})|\mathcal{F}_t\right].$$

The scalar parameter τ is a measure of model uncertainty aversion, and larger values correspond to higher levels of aversion. In the limit, as τ approaches zero from the right, the investor's problem becomes identical to the (nonrobust) mean-variance problem described in Section 4. The investor uses relative entropy to measure the closeness of distributions and is worried about receiving less utility if alternative specifications, whose distributions have a small relative entropy with respect to his best approximation, happen to be true. As discussed by Hansen and Sargent (2007b) and Hansen and others (2006), this formulation is equivalent to a robust control problem with constraints on the degree of model misspecification.[12]

The inner minimization, in the robust problem, considers a "constrained" worst-case possibility for asset returns. The solution to the inner minimization is

$$\varrho_t^*(z_{t+1}) = \frac{\exp\left[-\tau\phi_t' z_{t+1} + \frac{\theta\tau}{2}\left(\phi_t' z_{t+1}\right)^2\right]}{E\left(\exp\left[-\tau\phi_t' z_{t+1} + \frac{\theta\tau}{2}(\phi_t' z_{t+1})^2\right]|\mathcal{F}_t\right)}. \tag{17}$$

Thus the agent focuses on the distribution $f(z_{t+1}|\mathcal{F}_t)\varrho_t^*(z_{t+1})$ when solving the outer maximization problem for portfolio choices. Substituting this minimizing choice of ϱ into the robust problem yields:

$$\max_{\phi_t}\Bigg(\phi_t'\hat{\mu}_t + R_{ft+1} - \frac{1}{\tau}\log\int\exp\left[-\tau\phi_t' z_{t+1} + \frac{\theta\tau}{2}\left(\phi_t' z_{t+1}\right)^2\right] f(z_{t+1}|\mathcal{F}_t)dz_{t+1}\Bigg). \tag{18}$$

In the control theory literature, Problem (18) is known as a risk-sensitive optimization problem. Under general conditions, as discussed by Hansen and Sargent (2007b), risk-sensitive optimization problems are observationally equivalent to robust optimization problems. In the risk-sensitive problem, expectations are taken with respect to the best approximating distribution, $f(z_{t+1}|\mathcal{F}_t)$, not the alternative distribution $f(z_{t+1}|\mathcal{F}_t)\varrho_t^*(z_{t+1})$.

5.2 Robust Bayesian-averaging portfolio choices

BA investors believe the distribution of the means and variances of excess returns is a mixture of normal-inverse-Wisharts and fully take into account estimation error or parameter uncertainty. We assume *robust* BA investors ignore parameter uncertainty in variances (within each model), when making portfolio choices, for two reasons. First, because robust agents worry about all aspects of the return distribution it is not necessary for them to fully take into account parameter uncertainty. Second, if robust agents believe that the distribution of the means and variances of excess returns is a mixture of normal-inverse-Wisharts, then there are no reasonable solutions to the robust problem since the integral in Equation (18) is infinite whenever agents invest in risky assets. In this case, for any positive level of aversion to model uncertainty, robust agents only invest in the risk-free asset.[13] Because robust agents ignore parameter uncertainty in variances, when making portfolio choices, their

[12] A somewhat similar approach to mean-variance analysis has been undertaken by Garlappi, Uppal, and Wang (2007). Their approach penalizes deviations from mean returns in a different way.

[13] If we assumed different distributions for excess returns, then it is possible that the robust problem, in the presence of parameter uncertainty, will have interesting solutions.

approximating specification is that the distribution of excess returns is a mixture of normals.

When making portfolio choices, robust BA investors are still uncertain about the correct model, still uncertain about mean returns in each model, and (as all robust investors) are uncertain about the entire specification. At time t, model m is correct with probability $P_t(m|\mathcal{F}_t)$ and conditioned on model m (but not the realization of μ), the distribution of excess returns is

$$R_{t+1} \sim \mathcal{N}\left(\mu_{m,t}, \bar{\Sigma}_{m,t}\right). \tag{19}$$

Parameter uncertainty in mean returns is taken into account and, as in Formula (12), $\bar{\Sigma}_{m,t}$ includes an adjustment to $\Sigma_{m,t}$ because μ is unknown. Views about means, $E(R_{t+1}|\mathcal{F}_t)$, and variances, $V(R_{t+1}|\mathcal{F}_t)$, of excess returns coincide with those who fully take into account of parameter uncertainty. However, views about other aspects of the return distribution are substantially different, and the exclusion of estimation error in variances leads to much different robust portfolio choices.

We assume that robust BA investors ignore parameter uncertainty in variances only when forming their approximating model and making portfolio choices. They do consider parameter uncertainty in variances when updating model probabilities and model parameters. Thus, robust and nonrobust Bayesian investors use the same formulas to update probabilities and parameters.

When excess returns are a mixture of normals, we can write the robust objective more explicitly. Conditioned on model m, the vector z_{t+1} is distributed normal with mean $\mu_{m,t} - \hat{\mu}_t$ and variance $\bar{\Sigma}_{m,t}$ because[14]

$$E(z_{t+1}|m,\mathcal{F}_t) = E(R_{t+1}|m,\mathcal{F}_t) - E(\hat{\mu}_t|m,\mathcal{F}_t) = \mu_{m,t} - \hat{\mu}_t$$

and

$$V(z_{t+1}|m,\mathcal{F}_t) = V(R_{t+1}|m,\mathcal{F}_t) = \bar{\Sigma}_{m,t}.$$

Using standard results the integral in Equation (18) is

$$\int \exp\left[-\tau\phi_t' z_{t+1} + \frac{\theta\tau}{2}\left(\phi_t' z_{t+1}\right)^2\right] f(z_{t+1}|\mathcal{F}_t)\, dz_{t+1}$$

$$= \sum_{m \in M_t} U_{m,t} P_t(m|\mathcal{F}_t), \tag{20}$$

where

$$U_{m,t} = \begin{cases} \frac{1}{\sqrt{q_{m,t}}} \exp\left[\frac{\tau^2 \phi_t' \ \bar{\Sigma}_{m,t}\phi_t - 2\tau\xi_{m,t} + \theta\tau\xi_{m,t}^2}{2q_{m.t}}\right] & \text{when } q_{m,t} > 0 \\ +\infty & \text{when } q_{m,t} \leq 0 \end{cases} \tag{21}$$

[14] The expected value of $\hat{\mu}_t$ given model m is $\hat{\mu}_t$ since $\hat{\mu}_t$ is a known constant at time t. We are not taking the expectation with respect to multiple samples. Asymptotically, if model m is correct, then we would expect $\hat{\mu}_t$ and $\mu_{m,t}$ to be identical. However, we only have one finite sample. If model m is correct, in a given finite sample, $\hat{\mu}_t$ will not usually equal $\mu_{m,t}$.

with

$$q_{m,t} = 1 - \theta\tau\phi_t' \bar{\Sigma}_{m,t}\phi_t, \qquad \xi_{m,t} = \phi_t'(\mu_{m,t} - \hat{\mu}_t).$$

When the integral is $+\infty$, the value of the robust problem (Formula [18]) is $-\infty$. It is not possible to derive an analytical formula for optimal portfolio choices. However, because the objective is concave in ϕ_t and its straightforward to derive an analytical first derivative of the objective, this optimization problem is easy to solve numerically.

In summary, robust BA portfolio choices solve Formula (18) where the integral is given by Equation (20).

6. Alternative Portfolio Rules

We review several well-known portfolio choice algorithms for allocating investments among a risk-free asset and n risky assets. All of the algorithms select an n-dimensional vector of portfolio weights, which gives the fraction of wealth invested in each risky asset. We propose robust and constrained versions of many of the algorithms. Section 7 compares the out-of-sample performance of each algorithm to robust and nonrobust BA.

6.1 Rolling expectations

The rolling expectations method, sometimes referred to as the rolling (windows) approach, approximates future means and variances of returns with the sample means and covariances from a fixed window of recent data. Investors who use rolling expectations believe $E(R_{t+1}|\mathcal{F}_t) = \hat{\mu}_t$ and $V(R_{t+1}|\mathcal{F}_t) = \hat{\Sigma}_t$, where

$$\hat{\mu}_t = \frac{1}{w} \sum_{s=t-w+1}^{t} R_s, \tag{22}$$

$$\hat{\Sigma}_t = \frac{1}{c} \sum_{s=t-w+1}^{t} (R_s - \hat{\mu}_t)(R_s - \hat{\mu}_t)', \tag{23}$$

and w is the window size. Priors are not used, and data are equally weighted within the window. Different variations of this algorithm use different values of the scaling coefficient, c. Some algorithms use a maximum likelihood estimate of the covariance matrix which sets $c = w$. Other algorithms use $c = w - n - 2$ so that $\hat{\Sigma}_t^{-1}$ is an unbiased estimator of the population inverse covariance matrix, under standard assumptions. Another possibility is to set $c = w - 1$ so that $\hat{\Sigma}_t$ is an unbiased estimator of the population covariance matrix, under standard assumptions.[15] In our applications of the rolling expectations method, we set $c = w - 1$. A nonrobust investor with rolling expectations uses the formula $\frac{1}{\theta}\hat{\Sigma}_t^{-1}\hat{\mu}_t$ to compute optimal mean-variance portfolios, where θ is the investor's risk aversion.

15 The standard assumptions used to justify estimates of covariances do not necessarily hold in our setup.

6.2 Historical expectations

The historical-expectations approach approximates future means and variances of returns using all available historical returns. Investors who use historical expectations believe $E(R_{t+1}|\mathcal{F}_t)=\hat{\mu}_t$ and $V(R_{t+1}|\mathcal{F}_t)=\hat{\Sigma}_t$, where

$$\hat{\mu}_t=\frac{1}{t}\sum_{s=1}^{t}R_s, \tag{24}$$

$$\hat{\Sigma}_t=\frac{1}{t-1}\sum_{s=1}^{t}(R_s-\hat{\mu}_t)(R_s-\hat{\mu}_t)'. \tag{25}$$

These investors treat all past data as equally important and do not use priors. Similar to investors with rolling expectations, a nonrobust investor with historical expectations uses the formula $\frac{1}{\theta}\hat{\Sigma}_t^{-1}\hat{\mu}_t$ to compute optimal mean-variance portfolios.[16]

If means and covariances are constant over time, then the historical-expectations approach should outperform the rolling expectations approach. However, if means and covariances change over time, then the rolling expectations approach can perform better if its window size is chosen wisely.

6.3 1/N

The $1/N$ strategy invests an equal amount in each available risky asset and nothing in the risk-free asset. This strategy has been discussed for a long time. Duchin and Levy (2009) trace its origins to the Babylonian Talmud which said: "A man should always place his money, one-third in land, a third into merchandise, and keep a third in hand." The modern formulation of $1/N$ has been analyzed by DeMiguel, Garlappi, and Uppal (2009) and Tu and Zhou (2011). The portfolio choice rule for $1/N$ is prespecified to be

$$\phi_t=\left(\frac{1}{n}\right)\mathbf{1},$$

where **1** is a vector of ones. Agents invest an equal amount in each risky asset regardless of their beliefs about means and variances. By design, this approach never sells assets short.

6.4 Market

The market strategy always invests 100% of wealth in the value-weighted market. This approach does not involve any optimization and by design does not engage in short selling. This is the optimal investment strategy if all investors have identical mean-variance preferences and the equilibrium in the capital asset pricing model (CAPM) describes reality.

---

[16] In this study, all applications of historical expectations compute $\hat{\Sigma}_t$ using the $t-1$ divisor in Formula (25). Similar to the discussion in the rolling windows section, possible variations on this algorithm could divide by t or $t-n-2$.

6.5 The minimum variance strategy

The minimum variance strategy computes the portfolio, consisting of only risky assets, which minimizes the variance of returns. The investor's problem is to minimize

$$V\left(\phi_t' R_{t+1} + R_{ft+1} | \mathcal{F}_t\right)$$

by choice of portfolio weights ϕ_t subject to the constraint $\mathbf{1}'\phi_t = 1$. The optimal solution is the minimum variance portfolio

$$\phi_t = \frac{\hat{\Sigma}_t^{-1}\mathbf{1}}{\mathbf{1}'\hat{\Sigma}_t^{-1}\mathbf{1}},$$

where $\hat{\Sigma}_t$ is an estimate of conditional variance. The examples in later sections use the rolling window estimate of $\hat{\Sigma}_t$, in Equation (23), with $c = w - 1$. We use a window size of 250 days for daily data, 100 weeks for weekly data, and 60 months for monthly data. This method can sell assets short.

6.6 Jorion's Bayes-Stein procedure

Jorion (1986) assumes investors use shrinkage estimators when estimating means and variances. The estimate of the mean is a weighted average of the rolling mean estimator and the mean of the past return on the minimum variance portfolio:

$$\hat{\mu}_t^* = (1 - v_t)\hat{\mu}_t + v_t \hat{\mu}_t^g,$$

where

$$\hat{\mu}_t^g = \left(\frac{\hat{\mu}_t' \hat{\Sigma}_t^{-1}\mathbf{1}}{\mathbf{1}'\hat{\Sigma}_t^{-1}\mathbf{1}}\right)\mathbf{1}, \qquad v_t = \frac{n+2}{(n+2) + w(\hat{\mu}_t - \hat{\mu}_t^g)'\hat{\Sigma}_t^{-1}(\hat{\mu}_t - \hat{\mu}_t^g)}.$$

Here $\hat{\mu}_t$ is a rolling estimator of past mean returns; $\hat{\Sigma}_t^{-1}$ is the rolling unbiased estimator of the inverse covariance matrix with window size w and with $c = w - n - 2$; $\hat{\mu}_t^g$ is a vector whose each element is the investor's beliefs of the mean return on the minimum variance portfolio; and v_t captures the optimal weights to minimize utility loss (under assumptions described by Jorion[1986]). In later sections, we use a window size of 250 days for daily data, 100 weeks for weekly data, and 60 months for monthly data.

Jorion estimates the variance by combining $\hat{\Sigma}_t$ with the variance of the minimum variance portfolio:

$$\hat{\Sigma}_t^* = \left(1 + \frac{1}{w + J_t}\right)\hat{\Sigma}_t + \frac{J_t}{w(w + 1 + J_t)} \frac{\mathbf{1}\mathbf{1}'}{\mathbf{1}'\tilde{\Sigma}_t^{-1}\mathbf{1}},$$

where

$$J_t = \frac{n+2}{\left[(\hat{\mu}_t - \hat{\mu}_t^g 1_n)'\hat{\Sigma}_t^{-1}(\hat{\mu}_t - \hat{\mu}_t^g 1_n)\right]}.$$

The ratio $1/\left(\mathbf{1}'\tilde{\Sigma}_t^{-1}\mathbf{1}\right)$ is the variance of the minimum-variance portfolio return. The Jorion Bayes-Stein optimal portfolio choice rule is $\frac{1}{\theta}\hat{\Sigma}_t^{*-1}\hat{\mu}_t^*$.

6.7 Kan and Zhou's three-fund rule

Kan and Zhou (2007) modify shrinkage estimators to minimize the effect on utility loss of errors in estimating means and variances. As in Jorion's rule, estimates of mean returns are a linear combination of the rolling estimate and the mean of the minimum variance portfolio:

$$\hat{\mu}_t^* = (1-v_t)\hat{\mu}_t + v_t\hat{\mu}_t^g,$$

where

$$v_t = \frac{n}{w\xi_t + n}, \qquad \psi_t = (\hat{\mu}_t - \hat{\mu}_t^g)'\hat{\Sigma}_t^{-1}(\hat{\mu}_t - \hat{\mu}_t^g),$$

$$\xi_t = \frac{1}{w}[(w-n-1)\psi_t - (n-1)] + \frac{1}{wB_t}\left[2(\psi_t)^{\frac{n-1}{2}}(1+\psi_t)^{-\frac{(w-2)}{2}}\right],$$

and where

$$B_t = \int_0^{x_t} y^{a-1}(1-y)^{b-1}dy$$

is the incomplete beta function with parameters $x_t = \psi_t/(1+\psi_t)$, $a = (n-1)/2$, and $b = (w+1)/2$. Beliefs about the covariance of returns are given by the maximum likelihood estimator with $c = w$. The weight v_t differs from Jorion's weight. Kan and Zhou use the same formula for $\hat{\mu}_t^g$ as Jorion. In later sections, we use a window size of 250 days for daily data, 100 weeks for weekly data, and 60 months for monthly data.

The optimal portfolio choice rule is $(h/\theta)\hat{\Sigma}_t^{-1}\hat{\mu}_t^*$, where the constant

$$h = \frac{(w-n-1)(w-n-4)}{w(w-2)}$$

minimizes utility loss, under Kan and Zhou's assumptions.

6.8 Robust portfolio choices for alternative algorithms

We consider robust versions of the rolling, historical, Jorion, and Kan-Zhou approaches. Because we need to know the best available approximation to the distribution of future returns to compute robust portfolio choices, we assume at time t an investor's best approximation is normal: $R_{t+1} \sim \mathcal{N}\left(\mu_t^*, \Sigma_t^*\right)$, where μ_t^* and Σ_t^* vary across the four methods.[17] Because the robust investor worries that his approximation is wrong, he solves the robust mean-variance optimization problem in Section 5.1, which simplifies to:

$$\max_{\phi_t}\left(\phi_t'\mu_t^* + R_{ft+1} + \frac{\log q_t}{2\tau} - \frac{\tau\phi_t'\Sigma_t^*\phi_t}{2q_t}\right)$$

[17] For rolling and historical expectations, $\mu_t^* = \hat{\mu}_t$ and $\Sigma_t^* = \hat{\Sigma}_t$. For Jorion's procedure, $\mu_t^* = \hat{\mu}_t^*$ and $\Sigma_t^* = \hat{\Sigma}_t^*$. For Kan-Zhou's rule, $\mu_t^* = \hat{\mu}_t^*$ and $\Sigma_t^* = (1/h)\hat{\Sigma}_t$.

when the best approximation of returns is normal. Here

$$q_t = 1 - \theta\tau\phi_t' \Sigma_t^* \phi_t,$$

and when $q_t \leq 0$, the objective is $-\infty$.[18] As in Section 5.2, it is straightforward to derive an analytical first derivative, which makes it possible to compute numerical solutions quickly and accurately.

6.9 Constrained portfolio choices

Optimal portfolio choices sometimes require investors to take large negative positions in assets which may not be advisable or allowed under current regulations. It may be more realistic to consider problems with restrictions on short selling. In this study we provide examples with strong restrictions that prohibit any short selling.

We consider the modified mean-variance problem: maximize the objective in Formula (9) by choice of portfolio weights ϕ_t subject to the constraints

$$\phi_t \geq 0, \qquad \mathbf{1}'\phi_t \leq 1, \tag{26}$$

where $\mathbf{1}$ is a n-dimensional vector of ones. This approach forbids short-selling in any asset, including the risk-free rate. We also consider robust portfolio choice problems when short selling is prohibited by adding the constraints in Equation (26) to the investor's problems described in Sections 5.2 and 6.8.

7. Out-of-Sample Performance

We compare the out-of-sample performance of the BA approach with other strategies on 24 daily, weekly, and monthly datasets from the Center for Research in Security Prices (CRSP) and Kenneth French's data library.[19] We also examine performance on simulated i.i.d. data and simulated data with regime changes.[20] Following DeMiguel, Garlappi, and Uppal (2009), we add the value-weighted market to each portfolio in the actual data sets, but not

[18] The simplification follows because investors believe the conditional distribution of z_{t+1} is normal with mean zero and variance Σ_t^*. Thus,

$$\int \exp\left[-\tau\phi_t' z_{t+1} + \frac{\theta\tau}{2}\left(\phi_t' z_{t+1}\right)^2\right] f(z_{t+1}|\mathcal{F}_t)dz_{t+1} = \frac{1}{\sqrt{q_t}}\exp\left(\frac{\tau^2\phi_t'\Sigma_t^*\phi_t}{2q_t}\right)$$

when $q_t > 0$.

[19] The data were downloaded from Ken French's Web site and Wharton Data Services (WRDS) in early 2012. See CRSP's documentation http://www.crsp.com/documentation/index.html for more details on the CRSP portfolios. See Ken French's Web site http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ for detailed descriptions of his data. In early 2012, French's daily data began in July 1963 (and later for some datasets). We integrate both CRSP and French's daily data to obtain weekly data. For CRSP, we also integrate daily data to obtain monthly data.

[20] Summary statistics for daily data are provided in Table 1. Summary statistics for weekly, monthly, and simulated data are available upon request. Appendix A describes the artificial data.

the simulated data sets. To facilitate the replication of our results, when using CRSP data we use CRSP's value-weighted market return on NYSE, Amex, and NASDAQ stocks and we approximate the daily risk-free rate from CRSP's monthly data on the return of 30-day Treasury bill. When using Ken French's data, we use the value-weighted market return and the daily risk-free rate reported by Ken French.

We measure out-of-sample performance using Sharpe ratios and certainty equivalents. For all results in this study, risk aversion is one ($\theta = 1$). For nonrobust approaches, model uncertainty aversion is zero ($\tau = 0$). In this section, for all robust approaches, model uncertainty aversion is four ($\tau = 4$) and for BA strategies the prior rule is perfect sharing (the fixed sharing prior rule with $\alpha = 1$).

7.1 Sharpe ratios

The Sharpe ratio on the investor's portfolio is the mean excess return on the portfolio divided by its standard deviation:

$$\frac{\bar{E}\left(R_{pt+1}\right)}{\bar{\sigma}\left(R_{pt+1}\right)}, \tag{27}$$

where $R_{pt+1} = \phi_t' R_{t+1}$ is the excess return on the investor's portfolio, ϕ_t is the portfolio weight vector, $\bar{E}$ denotes sample mean, and $\bar{\sigma}$ denotes sample standard deviation. For all reported portfolio choice strategies, ϕ_t only uses information available at time t. For unconstrained nonrobust portfolio optimization problems, when there are no restrictions on choices, the out-of-sample Sharpe ratios do not depend on risk aversion. Otherwise, if short-selling constraints are present then out-of-sample Sharpe ratios can depend on risk aversion. For robust problems, Sharpe ratios can depend on risk-aversion and model uncertainty aversion, both with and without short-selling constraints.

7.1.1 Unconstrained portfolios. Table 2 presents out-of-sample Sharpe ratios for robust and nonrobust strategies when there are no short-selling constraints. Over 24 daily datasets, the robust BA algorithm earns the highest out-of-sample Sharpe ratio on 12 datasets and the BA algorithm earns the highest out-of-sample Sharpe ratio on nine datasets. Other approaches earn the highest Sharpe ratio on the following three datasets:

1. the value-weighted portfolios sorted on short-term reversal,
2. the value-weighted portfolios sorted on size and short-term reversal, and
3. the equally-weighted portfolios sorted on size and short-term reversal,

which involve the short-term reversal portfolios. Reversal portfolios are by design a challenge for the BA method because reversal portfolios are constructed so that well- (poorly) performing portfolios in the past are likely to perform poorly (well) in the future. Implicitly, the BA algorithm assumes

Table 1
Daily summary statistics

Portfolios	Name	Number	Sample size	Start date	End date	Lowest mean	Highest mean	Lowest std	Highest std	Lowest return	Highest return
Panel A: CRSP daily data											
Beta	beta	11	22780	1926-01-02	2011-12-30	0.040	0.113	0.741	1.993	−22.370	41.830
Size	size	11	22780	1926-01-02	2011-12-30	0.039	0.066	1.070	1.300	−18.279	30.859
Standard deviation	std	11	22780	1926-01-02	2011-12-30	0.040	0.228	0.585	1.631	−18.592	30.413
Equally-weighted market	em	2	22780	1926-01-02	2011-12-30	0.040	0.080	1.070	1.078	−17.135	20.302
Panel B: Ken French's value-weighted daily data											
Industry	ind	11	12211	1963-07-01	2011-12-30	0.038	0.053	0.836	1.453	−20.000	19.330
Size	size	11	12211	1963-07-01	2011-12-30	0.039	0.050	0.872	1.078	−19.640	11.920
Book-to-market	beme	11	12211	1963-07-01	2011-12-30	0.038	0.058	0.966	1.138	−18.530	13.850
Long-term reversal	ltr	11	11212	1967-06-19	2011-12-30	0.037	0.054	0.980	1.354	−20.800	17.020
Momentum	mom	11	12211	1963-07-01	2011-12-30	0.007	0.070	0.981	1.617	−20.100	21.400
Short-term reversal	str	11	12211	1963-07-01	2011-12-30	−0.039	0.164	0.998	1.658	−19.220	25.170
Size and book-to-market	size-beme	26	12211	1963-07-01	2011-12-30	0.021	0.065	0.912	1.254	−20.430	13.050
Size and long-term reversal	size-ltr	26	11212	1967-06-19	2011-12-30	0.035	0.068	0.962	1.380	−22.390	17.000
Size and momentum	size-mom	26	12211	1963-07-01	2011-12-30	0.021	0.086	0.894	1.597	−21.000	22.670
Size and short-term reversal	size-str	26	8361	1978-11-13	2011-12-30	−0.103	0.247	0.968	1.709	−20.200	26.230
Panel C: Ken French's equal-weighted daily data											
Industry	ind	11	12211	1963-07-01	2011-12-30	0.041	0.100	0.700	1.356	−17.133	20.430
Size	size	11	12211	1963-07-01	2011-12-30	0.041	0.105	0.748	1.103	−19.460	14.110
Book-to-market	beme	11	12211	1963-07-01	2011-12-30	0.041	0.124	0.774	1.144	−17.133	31.660
Long-term reversal	ltr	11	11212	1967-06-19	2011-12-30	0.041	0.158	0.870	1.207	−17.133	15.120
Momentum	mom	11	12211	1963-07-01	2011-12-30	0.041	0.122	0.809	1.138	−17.133	11.690
Short-term reversal	str	11	12211	1963-07-01	2011-12-30	−0.150	0.507	0.813	1.332	−17.133	21.600
Size and book-to-market	size-beme	26	12211	1963-07-01	2011-12-30	0.041	0.125	0.724	1.281	−19.360	27.700
Size and long-term reversal	size-ltr	26	11212	1967-06-19	2011-12-30	0.037	0.153	0.825	1.362	−21.530	17.240
Size and momentum	size-mom	26	12211	1963-07-01	2011-12-30	0.020	0.126	0.731	1.542	−20.710	17.300
Size and short-term reversal	size-str	26	8361	1978-11-13	2011-12-30	−0.206	0.505	0.769	1.684	−20.910	23.910

Table 1 lists properties of daily asset returns from CRSP and Kenneth French's Web site. The second column gives short labels for the assets that are used in subsequent tables. The value-weighted market return is added to each group of assets. The third and fourth columns provide the number of assets in the portfolio and the total sample size. Columns 5 and 6 show the beginning and end dates of the data. The seventh and eighth columns calculate the lowest and highest mean daily (net) percentage returns, over all assets in the portfolio. The ninth and tenth columns provide the lowest and highest standard deviations (of percentage returns). The last two columns provide the lowest and highest daily percentage return. In this table, the risk-free rate is not subtracted from the nominal (net) percentage returns. Data from the burn-in period of 1,000 days are included.

Table 2
Daily Sharpe ratios for unconstrained portfolio choices

	Robust strategies							Nonrobust strategies									
	BA	R100	R250	R500	H	JOR	KZ	BA	R100	R250	R500	H	JOR	KZ	1/N	MKT	MINV
Panel A: CRSP daily data																	
Beta	**0.181**	0.152⋆	0.162	0.149⋆	0.117⋆	0.170	0.170	0.179	0.154⋆	0.162	0.149⋆	0.117⋆	0.169	0.169	0.053⋆	0.023⋆	0.049⋆
Size	0.206	0.048⋆	0.068⋆	0.053⋆	0.022⋆	0.078⋆	0.081⋆	**0.210**	0.050⋆	0.069⋆	0.053⋆	0.022⋆	0.078⋆	0.081⋆	0.030⋆	0.023⋆	0.027⋆
Std	**0.240**	0.193⋆	0.198⋆	0.176⋆	0.152⋆	0.200⋆	0.198⋆	0.230⋆	0.194⋆	0.197⋆	0.176⋆	0.152⋆	0.199⋆	0.197⋆	0.059⋆	0.023⋆	0.046⋆
Em	**0.160**	0.139	0.123⋆	0.107⋆	0.094⋆	0.118⋆	0.119⋆	0.154⋆	0.139	0.122⋆	0.106⋆	0.094⋆	0.117⋆	0.119⋆	0.043⋆	0.023⋆	0.037⋆
Panel B: Ken French's value-weighted daily data																	
Ind	**0.201**	0.107⋆	0.087⋆	0.054⋆	0.032⋆	0.086⋆	0.082⋆	0.195⋆	0.108⋆	0.088⋆	0.054⋆	0.032⋆	0.086⋆	0.082⋆	0.024⋆	0.019⋆	0.037⋆
Size	0.265	0.132⋆	0.109⋆	0.070⋆	0.026⋆	0.111⋆	0.110⋆	**0.268**	0.133⋆	0.109⋆	0.070⋆	0.026⋆	0.111⋆	0.109⋆	0.023⋆	0.019⋆	0.019⋆
Beme	**0.132**	0.067⋆	0.067⋆	0.041⋆	0.034⋆	0.067⋆	0.064⋆	0.129⋆	0.067⋆	0.067⋆	0.041⋆	0.034⋆	0.067⋆	0.064⋆	0.025⋆	0.019⋆	0.027⋆
Ltr	**0.131**	0.063⋆	0.062⋆	0.043⋆	0.026⋆	0.061⋆	0.057⋆	0.126⋆	0.063⋆	0.062⋆	0.043⋆	0.026⋆	0.060⋆	0.057⋆	0.027⋆	0.021⋆	0.020⋆
Mom	**0.185**	0.090⋆	0.086⋆	0.066⋆	0.066⋆	0.084⋆	0.081⋆	0.178⋆	0.091⋆	0.086⋆	0.066⋆	0.066⋆	0.084⋆	0.081⋆	0.018⋆	0.019⋆	0.006⋆
Str	0.174	0.203⋆	0.227⋆	0.217⋆	0.210⋆	0.224⋆	0.222⋆	0.171	0.204⋆	**0.228⋆**	0.217⋆	0.211⋆	0.224⋆	0.222⋆	0.024⋆	0.019⋆	−0.015⋆
Size-beme	0.255	0.156⋆	0.169⋆	0.133⋆	0.105⋆	0.182⋆	0.182⋆	**0.257**	0.158⋆	0.170⋆	0.134⋆	0.105⋆	0.182⋆	0.182⋆	0.027⋆	0.019⋆	0.082⋆
Size-ltr	0.216	0.121⋆	0.134⋆	0.107⋆	0.062⋆	0.147⋆	0.150⋆	**0.222**	0.123⋆	0.135⋆	0.107⋆	0.062⋆	0.147⋆	0.150⋆	0.034⋆	0.021⋆	0.054⋆
Size-mom	0.303	0.201⋆	0.197⋆	0.170⋆	0.123⋆	0.200⋆	0.196⋆	**0.306**	0.201⋆	0.198⋆	0.170⋆	0.123⋆	0.200⋆	0.195⋆	0.026⋆	0.019⋆	0.047⋆
Size-str	0.540	0.639⋆	0.645⋆	**0.651⋆**	0.601⋆	0.634⋆	0.629⋆	0.586⋆	0.596⋆	0.602⋆	0.605⋆	0.598⋆	0.590⋆	0.587⋆	0.028⋆	0.026⋆	−0.088⋆
Panel C: Ken French's equal-weighted daily data																	
Ind	**0.414**	0.320⋆	0.273⋆	0.208⋆	0.108⋆	0.277⋆	0.276⋆	0.404⋆	0.321⋆	0.274⋆	0.209⋆	0.108⋆	0.277⋆	0.276⋆	0.058⋆	0.019⋆	0.076⋆
Size	0.333	0.241⋆	0.245⋆	0.230⋆	0.173⋆	0.249⋆	0.249⋆	**0.336**	0.243⋆	0.247⋆	0.232⋆	0.172⋆	0.250⋆	0.249⋆	0.032⋆	0.019⋆	0.119⋆
Beme	**0.339**	0.239⋆	0.265⋆	0.250⋆	0.194⋆	0.270⋆	0.269⋆	0.339	0.241⋆	0.266⋆	0.251⋆	0.194⋆	0.269⋆	0.268⋆	0.061⋆	0.019⋆	0.138⋆
Ltr	**0.326**	0.244⋆	0.256⋆	0.250⋆	0.197⋆	0.257⋆	0.255⋆	0.325	0.245⋆	0.257⋆	0.251⋆	0.197⋆	0.257⋆	0.255⋆	0.061⋆	0.021⋆	0.075⋆
Mom	**0.382**	0.233⋆	0.223⋆	0.182⋆	0.124⋆	0.226⋆	0.226⋆	0.379	0.233⋆	0.224⋆	0.183⋆	0.124⋆	0.227⋆	0.226⋆	0.060⋆	0.019⋆	0.033⋆
Str	**0.880**	0.696⋆	0.705⋆	0.707⋆	0.783⋆	0.694⋆	0.693⋆	0.846⋆	0.567⋆	0.588⋆	0.597⋆	0.785⋆	0.579⋆	0.580⋆	0.034⋆	0.019⋆	−0.116⋆
Size-beme	0.350	0.249⋆	0.287⋆	0.274⋆	0.238⋆	0.298⋆	0.297⋆	**0.353**	0.251⋆	0.289⋆	0.276⋆	0.238⋆	0.298⋆	0.296⋆	0.041⋆	0.019⋆	0.159⋆
Size-ltr	0.308	0.225⋆	0.261⋆	0.255⋆	0.217⋆	0.272⋆	0.272⋆	**0.317⋆**	0.229⋆	0.264⋆	0.257⋆	0.217⋆	0.272⋆	0.271⋆	0.042⋆	0.021⋆	0.098⋆
Size-mom	0.386	0.256⋆	0.268⋆	0.250⋆	0.202⋆	0.274⋆	0.271⋆	**0.390**	0.258⋆	0.270⋆	0.251⋆	0.202⋆	0.274⋆	0.271⋆	0.035⋆	0.019⋆	0.083⋆
Size-str	0.882	0.829	0.843	0.853	**0.939⋆**	0.825⋆	0.819⋆	0.810⋆	0.689⋆	0.675⋆	0.692⋆	0.937⋆	0.664⋆	0.663⋆	0.032⋆	0.026⋆	−0.147⋆

Table 2 calculates Sharpe ratios for several strategies on daily datasets when there are no restrictions on portfolio selection. Columns 2 through 8 provide results for robust algorithms, Columns 9 through 15 provide results for nonrobust algorithms, and Columns 16 through 18 provide results for other strategies. BA computes expectations using Bayesian averaging. See Section 4.2 for nonrobust BA portfolio choices and Section 5.2 for robust BA portfolio choices. R100, R250, and R500 are rolling expectation approaches that use predetermined window sizes of 100, 250, and 500 days, respectively. H is the historical-expectations approach. JOR is Jorion's Bayes-Stein procedure. KZ is Kan and Zhou's three-fund rule. The $1/N$ strategy invests an equal amount in each risky asset. MKT is the market strategy. MINV is the minimum variance strategy. A burn-in period of 1,000 days is used, and Sharpe ratios are computed over the remaining sample. The highest Sharpe ratio among all the strategies, for each dataset, is in boldface. For nonrobust strategies, the value of risk aversion does not affect results. For robust strategies, risk aversion is one and model uncertainty aversion is four. For robust BA, we use a fixed sharing prior with $\alpha = 1$. A ⋆ indicates that the strategy significantly underperforms or outperforms robust BA at the 95% confidence level.

that assets with high returns in the past will have high returns in the future and reversals are not foreseen.

For each strategy, Table 2 indicates if the difference in Sharpe ratios from the robust BA strategy is significant at the 95% level. To compute significance, we calculate t-statistics using GMM with the moment conditions described in Appendix B.1, where the robust BA method is the benchmark and 22 lags are used to compute the Newey-West spectral density. Robust BA is significantly better than 1/N, the market strategy, and the minimum-variance procedure on all 24 datasets. The overall performance of robust and nonrobust BA can not be distinguished: on 13 datasets, they are statistically identical, on nine datasets, robust BA significantly outperforms BA, and on two datasets BA significantly outperforms robust BA. Robust BA significantly outperforms each of the other alternative algorithms on at least 20 datasets.

Table 3 presents percentiles of Sharpe ratios across datasets, for each portfolio strategy. For daily data, the minimum Sharpe ratio obtained by the robust BA approach across all datasets is 0.131. This is higher than the minimum of any other algorithm. At the 25% percentile, the robust BA achieves a Sharpe ratio of 0.197 which is also the highest across all algorithms. BA has the highest median and highest 75% percentile across all algorithms, slightly beating robust BA by 0.003 and 0.001, respectively. Robust historical expectations achieves the highest maximum Sharpe ratio. In summary, the robust and nonrobust BA approaches win four of the five daily percentiles. For weekly data, robust BA wins four of the five percentiles, losing only the median to the robust Jorion algorithm by 0.008. For monthly data, robust BA wins three of the five percentiles, losing the minimum to the market strategy by 0.004 and the maximum to the robust strategy with a rolling window of 60 months. For simulated i.i.d. data at every percentile, the 1/N strategy wins. The minimum variance strategy is second on four percentiles. For simulated data with regime changes, either robust or nonrobust BA wins all five percentiles.

Why do both robust and nonrobust BA perform well (both absolutely and relative to other algorithms) when there are regime changes and poorly on i.i.d. data? The main reason that robust BA and BA perform better in absolute terms is that the conditional volatility of the underlying factor is lower in the regime-change specification, which helps make forecasts of future means more accurate.[21] The main reason that robust and nonrobust BA perform better relative to other procedures is that they can better predict time-varying expected returns (caused by the changing mean of the factor) than other algorithms.

7.1.2 Constrained portfolios. Table 4 shows that the robust BA algorithm achieves much higher out-of-sample Sharpe ratios than other strategies do on 19 of 24 daily datasets, when short selling is prohibited. Four of the five

[21] As described in Appendix A, we set the *unconditional* variance of the factor to be identical in the i.i.d. and regime-change cases.

Table 3
Sharpe ratio percentiles

	Robust strategies							Nonrobust strategies									
	BA	R_1	R_2	R_3	H	JOR	KZ	BA	R_1	R_2	R_3	H	JOR	KZ	1/N	MKT	MINV
Panel A: Daily data																	
Min	**0.131**	0.048	0.062	0.041	0.022	0.061	0.057	0.126	0.050	0.062	0.041	0.022	0.060	0.057	0.018	0.019	−0.147
25%	**0.197**	0.129	0.119	0.098	0.065	0.116	0.117	0.191	0.131	0.119	0.097	0.065	0.116	0.116	0.027	0.019	0.020
Median	0.284	0.202	0.211	0.179	0.123	0.212	0.210	**0.287**	0.203	0.211	0.179	0.123	0.212	0.210	0.033	0.019	0.041
75%	0.358	0.245	0.266	0.250	0.204	0.273	0.271	**0.359**	0.246	0.267	0.251	0.204	0.273	0.271	0.045	0.021	0.077
Max	0.882	0.829	0.843	0.853	**0.939**	0.825	0.819	0.846	0.689	0.675	0.692	0.937	0.664	0.663	0.061	0.026	0.159
Panel B: Weekly data																	
Min	**0.102**	0.071	0.055	0.041	0.030	0.052	0.045	0.098	0.071	0.055	0.041	0.030	0.052	0.045	0.039	0.042	−0.031
25%	**0.197**	0.152	0.143	0.133	0.120	0.153	0.153	0.192	0.151	0.142	0.133	0.120	0.152	0.152	0.054	0.042	0.046
Median	0.300	0.303	0.304	0.284	0.228	**0.308**	0.307	0.290	0.304	0.307	0.288	0.227	0.308	0.306	0.069	0.042	0.081
75%	**0.391**	0.359	0.374	0.380	0.353	0.379	0.378	0.386	0.364	0.376	0.384	0.351	0.375	0.372	0.102	0.046	0.120
Max	**1.269**	0.996	1.030	1.094	1.113	0.991	0.970	0.985	0.681	0.713	0.790	1.114	0.696	0.695	0.125	0.058	0.271
Panel C: Monthly data																	
Min	0.123	0.054	0.090	0.078	0.057	0.099	0.098	0.118	0.051	0.090	0.079	0.057	0.099	0.099	0.119	**0.127**	0.067
25%	**0.166**	0.087	0.130	0.128	0.135	0.145	0.148	0.155	0.078	0.131	0.128	0.134	0.144	0.147	0.145	0.130	0.129
Median	**0.252**	0.125	0.205	0.223	0.210	0.205	0.200	0.241	0.112	0.202	0.221	0.209	0.203	0.195	0.157	0.131	0.151
75%	**0.350**	0.200	0.286	0.339	0.331	0.289	0.285	0.322	0.180	0.287	0.334	0.331	0.288	0.283	0.172	0.143	0.162
Max	0.482	0.468	**0.525**	0.522	0.451	0.517	0.513	0.452	0.457	0.520	0.522	0.446	0.509	0.505	0.236	0.151	0.252
Panel D: Simulated i.i.d. data																	
Min	−0.020	−0.019	−0.019	−0.016	−0.011	−0.015	−0.014	−0.020	−0.019	−0.019	−0.016	−0.011	−0.015	−0.014	**0.008**	—	−0.002
25%	−0.001	0.001	−0.002	0.001	0.005	−0.001	−0.002	−0.001	0.001	−0.002	0.001	0.005	−0.001	−0.002	**0.027**	—	0.013
Median	0.007	0.007	0.005	0.005	0.013	0.005	0.006	0.007	0.007	0.005	0.005	0.013	0.005	0.006	**0.031**	—	0.019
75%	0.013	0.010	0.011	0.011	0.020	0.012	0.012	0.013	0.011	0.011	0.011	0.019	0.012	0.012	**0.036**	—	0.025
Max	0.031	0.028	0.029	0.026	0.053	0.030	0.028	0.032	0.029	0.029	0.026	0.053	0.030	0.028	**0.054**	—	0.041
Panel E: Simulated regime changes																	
Min	**0.086**	0.054	0.023	−0.009	−0.007	0.021	0.019	0.084	0.054	0.023	−0.009	−0.007	0.021	0.019	−0.047	—	−0.039
25%	**0.106**	0.074	0.050	0.036	0.013	0.053	0.054	0.105	0.076	0.050	0.036	0.013	0.053	0.054	0.015	—	0.004
Median	0.117	0.090	0.065	0.049	0.024	0.071	0.070	**0.118**	0.091	0.065	0.049	0.025	0.071	0.069	0.040	—	0.027
75%	0.125	0.099	0.076	0.056	0.044	0.083	0.084	**0.126**	0.100	0.076	0.056	0.044	0.083	0.084	0.059	—	0.039
Max	**0.152**	0.129	0.107	0.088	0.081	0.113	0.114	0.152	0.130	0.108	0.088	0.081	0.113	0.114	0.091	—	0.061

Table 3 summarizes the distribution of Sharpe ratios on daily, weekly, monthly, and simulated data sets (when there are no restrictions on portfolio selection). For daily data, for each algorithm, we show the percentiles of the Shape ratios listed in Table 2. For weekly and monthly data, we also report percentiles across the same 24 datasets. Panels D and E report certainty equivalent percentiles for 41 draws of 12,000 periods from the i.i.d. and regime-change data-generating processes described in Appendix A. The burn-in period is 1,000 days for daily and simulated data, 250 weeks for weekly data, and 120 months for monthly data. The highest Sharpe ratio among all strategies, for each percentile is in boldface. For nonrobust strategies, the value of risk aversion does not affect results. For robust strategies, risk aversion is one and model uncertainty aversion is four. For robust BA, we use a fixed sharing prior with $\alpha = 1$. In the simulated data sets there is not a well-defined market portfolio. R_1, R_2, and R_3 refer to rolling approaches with window sizes of 100, 250, and 500 days for daily data; 50, 100, and 150 weeks for weekly data; and 30, 60, and 90 months for monthly data.

Table 4
Daily Sharpe ratios for constrained portfolio choices

	Robust strategies							Nonrobust strategies						
	BA	R100	R250	R500	H	JOR	KZ	BA	R100	R250	R500	H	JOR	KZ
Panel A: CRSP daily data														
Beta	**0.105**	0.077⋆	0.085⋆	0.085⋆	0.096	0.093	0.095	0.099⋆	0.071⋆	0.070⋆	0.066⋆	0.053⋆	0.074⋆	0.077⋆
Size	**0.100**	0.070⋆	0.062⋆	0.049⋆	0.039⋆	0.069⋆	0.069⋆	0.094⋆	0.067⋆	0.051⋆	0.045⋆	0.043⋆	0.056⋆	0.060⋆
Std	**0.174**	0.154⋆	0.145⋆	0.137⋆	0.130⋆	0.145⋆	0.143⋆	0.166⋆	0.152⋆	0.144⋆	0.133⋆	0.130⋆	0.143⋆	0.142⋆
Em	**0.137**	0.078⋆	0.069⋆	0.061⋆	0.060⋆	0.073⋆	0.072⋆	0.132⋆	0.077⋆	0.063⋆	0.060⋆	0.060⋆	0.070⋆	0.067⋆
Panel B: Ken French's value-weighted daily data														
Ind	**0.100**	0.037⋆	0.040⋆	0.026⋆	0.024⋆	0.039⋆	0.041⋆	0.094⋆	0.037⋆	0.041⋆	0.021⋆	0.020⋆	0.039⋆	0.037⋆
Size	**0.100**	0.052⋆	0.046⋆	0.032⋆	0.022⋆	0.046⋆	0.045⋆	0.092⋆	0.047⋆	0.043⋆	0.030⋆	0.023⋆	0.041⋆	0.042⋆
Beme	**0.068**	0.040⋆	0.041⋆	0.027⋆	0.029⋆	0.041⋆	0.039⋆	0.064⋆	0.037⋆	0.038⋆	0.027⋆	0.029⋆	0.041⋆	0.041⋆
Ltr	**0.078**	0.039⋆	0.043⋆	0.023⋆	0.022⋆	0.044⋆	0.041⋆	0.075	0.038⋆	0.038⋆	0.022⋆	0.024⋆	0.041⋆	0.040⋆
Mom	**0.083**	0.033⋆	0.034⋆	0.017⋆	0.033⋆	0.031⋆	0.028⋆	0.081	0.032⋆	0.036⋆	0.017⋆	0.034⋆	0.027⋆	0.031⋆
Str	0.092	**0.102**	0.102	0.097	0.082	0.100	0.099	0.088	0.093	0.096	0.092	0.082	0.097	0.097
Size-beme	**0.093**	0.051⋆	0.046⋆	0.036⋆	0.041⋆	0.059⋆	0.063⋆	0.084⋆	0.043⋆	0.038⋆	0.032⋆	0.040⋆	0.049⋆	0.053⋆
Size-ltr	**0.076**	0.039⋆	0.048⋆	0.038⋆	0.041⋆	0.054	0.058	0.069⋆	0.037⋆	0.039⋆	0.034⋆	0.036⋆	0.046⋆	0.049⋆
Size-mom	**0.098**	0.056⋆	0.060⋆	0.049⋆	0.054⋆	0.066⋆	0.071⋆	0.088⋆	0.050⋆	0.056⋆	0.044⋆	0.055⋆	0.058⋆	0.062⋆
Size-str	0.189	0.213	0.204	0.197	0.166⋆	0.216	**0.226⋆**	0.166⋆	0.197	0.189	0.184	0.164⋆	0.190	0.197
Panel C: Ken French's equal-weighted daily data														
Ind	**0.212**	0.128⋆	0.106⋆	0.077⋆	0.063⋆	0.112⋆	0.113⋆	0.206⋆	0.122⋆	0.098⋆	0.065⋆	0.052⋆	0.103⋆	0.104⋆
Size	**0.158**	0.135	0.127⋆	0.105⋆	0.107⋆	0.132	0.134	0.146⋆	0.123⋆	0.114⋆	0.089⋆	0.107⋆	0.116⋆	0.121⋆
Beme	**0.172**	0.148	0.143⋆	0.136⋆	0.127⋆	0.152	0.153	0.168⋆	0.134⋆	0.131⋆	0.125⋆	0.127⋆	0.136⋆	0.138⋆
Ltr	**0.212**	0.177⋆	0.156⋆	0.152⋆	0.136⋆	0.162⋆	0.163⋆	0.204⋆	0.172⋆	0.156⋆	0.146⋆	0.136⋆	0.157⋆	0.162⋆
Mom	**0.217**	0.141⋆	0.134⋆	0.109⋆	0.086⋆	0.135⋆	0.135⋆	0.214	0.135⋆	0.129⋆	0.106⋆	0.084⋆	0.127⋆	0.122⋆
Str	**0.464**	0.400⋆	0.380⋆	0.369⋆	0.356⋆	0.393⋆	0.395⋆	0.449⋆	0.369⋆	0.359⋆	0.356⋆	0.356⋆	0.365⋆	0.367⋆
Size-beme	0.143	0.125	0.130	0.128	0.141	0.145	**0.154**	0.131⋆	0.110⋆	0.112⋆	0.100⋆	0.140	0.119	0.130
Size-ltr	0.143	0.144	0.146	0.150	0.143	0.153	**0.158**	0.132⋆	0.126	0.135	0.133	0.143	0.142	0.143
Size-mom	**0.154**	0.132	0.140	0.123⋆	0.100⋆	0.146	0.149	0.141⋆	0.122⋆	0.129⋆	0.114⋆	0.099⋆	0.134	0.137
Size-str	0.441	0.460	0.437	0.424	0.420	0.461	**0.466**	0.410⋆	0.422	0.416	0.411⋆	0.420	0.422	0.423

Table 4 calculates Sharpe ratios for several different strategies on daily datasets when short selling is prohibited. This table parallels Table 2. A burn-in period of 1,000 days is used, and Sharpe ratios are computed over the remaining sample. The highest Sharpe ratio in each row is in boldface. For all algorithms, risk aversion is one. For robust algorithms, model uncertainty aversion is four. For robust BA, we use a fixed sharing prior with $\alpha = 1$. A ⋆ indicates that the strategy significantly underperforms or outperforms robust BA at the 95% confidence level.

datasets for which robust BA approach underperforms involve either the short-term or long-term reversal portfolios. Unlike the unconstrained case, there is a statistical difference between robust and nonrobust BA. Robust BA statistically outperforms BA on 20 datasets and is statistically indistinguishable from BA on the other four datasets. Robust BA is significantly better than the robust Jorion and Kan-Zhou's approaches on 14 datasets, the robust strategy with a rolling window of 100 on 16 datasets, and all other algorithms on at least 18 datasets.

7.2 Certainty equivalents

The certainty equivalent is the constant risk-free rate that gives agents the same utility out-of-sample as their optimal choices and is defined as

$$\bar{E}\left(R_{pt+1}\right)+\bar{E}\left(R_{ft+1}\right)-\frac{\theta}{2}\bar{V}\left(R_{pt+1}\right), \tag{28}$$

where $\bar{V}$ is the sample variance. We assume that the goal of agents is to maximize the mean-variance criteria out of sample, whether or not they have robust preferences. Thus, investors would like to maximize Formula (28) regardless of their model uncertainty aversion.[22]

7.2.1 Unconstrained portfolios. Table 5 shows that robust BA earns decisively higher out-of-sample certainty equivalents on 20 of the 24 daily datasets, when short selling is allowed. The four datasets for which robust BA underperforms involve short-term reversal portfolios. The robust BA approach is statistically significantly better than the 1/N strategy, the market strategy, the minimum-variance strategy, the BA approach, and the nonrobust technique with a rolling window of 250 on all 24 datasets. Robust BA significantly outperforms each of the other alternative algorithms on at least 21 datasets.[23]

Table 6 presents the percentiles of certainty equivalents across datasets, for each algorithm. For daily and weekly data, robust BA achieves the highest certainty equivalent on four of the five percentiles by a large margin, and the nonrobust historical-expectations method achieves the best maximum certainty equivalent. For monthly data, the robust BA achieves the highest certainty equivalent on three of the five percentiles. The robust approaches with rolling windows of 90 and 60 months achieve (respectively) the highest certainty equivalents on the 75% and 100% percentiles. For simulated i.i.d. data, 1/N wins four percentiles, and the nonrobust historical-expectations approach wins the 100% percentile. For simulated data with regime changes, robust BA

---

[22] Appendix C discusses the reasons for using Formula (28) to compute certainty equivalents when investors have robust preferences.

[23] To compute significance, we calculate t-statistics using GMM with the moment conditions described in Appendix B.2, where the robust BA method is the benchmark and 22 lags are used to compute the Newey-West spectral density.

Table 5
Daily certainty equivalents for unconstrained portfolio choices

	Robust strategies							Nonrobust strategies									
	BA	R100	R250	R500	H	JOR	KZ	BA	R100	R250	R500	H	JOR	KZ	1/N	MKT	MINV
Panel A: CRSP daily data																	
Beta	**1.44**	1.10*	0.96*	0.72*	0.29*	0.83*	0.78*	−14.11*	−10.25*	−1.52*	−0.06*	0.68*	0.54*	0.85	0.07*	0.03*	0.04*
Size	**1.88**	−1.87	0.24*	0.15*	0.03*	0.29*	0.28*	−21.17*	−80.62	−5.47*	−2.03*	−0.01*	−1.66*	−0.93*	0.04*	0.03*	0.03*
Std	**2.29**	1.66*	1.32*	0.95*	0.45*	1.12*	1.04*	−16.53*	−11.27*	−1.13*	0.21*	1.16*	0.90*	1.19*	0.07*	0.03*	0.03*
Em	**0.97**	0.71*	0.48*	0.33*	0.17*	0.41*	0.43*	−3.18*	−1.23*	0.04*	0.31*	0.46*	0.25*	0.23*	0.05*	0.03*	0.05*
Panel B: Ken French's value-weighted daily data																	
Ind	**2.01**	0.60*	0.38*	0.16*	0.06*	0.30*	0.26*	−46.26*	−12.00*	−2.86*	−1.43*	−0.14*	−0.59*	−0.28*	0.04*	0.04*	0.04*
Size	**2.87**	0.88*	0.55*	0.25*	0.05*	0.48*	0.44*	−23.24*	−13.19*	−3.24*	−1.63*	−0.12*	−0.83*	−0.41*	0.04*	0.04*	0.03*
Beme	**0.86**	0.19*	0.25*	0.11*	0.07*	0.20*	0.17*	−39.80*	−10.45*	−2.69*	−1.34*	−0.12*	−0.52*	−0.22*	0.04*	0.04*	0.04*
Ltr	**0.85**	0.16*	0.21*	0.11*	0.05*	0.18*	0.15*	−37.76*	−10.27*	−2.68*	−1.14*	−0.13*	−0.58*	−0.29*	0.04*	0.04*	0.03*
Mom	**1.71**	0.41*	0.37*	0.23*	0.15*	0.29*	0.25*	−39.13*	−11.41*	−2.74*	−1.56*	0.08*	−0.47*	−0.17*	0.03*	0.04*	0.02*
Str	1.52	1.71	1.57	1.31	1.18	1.23	1.14*	−37.48*	−7.50*	0.39*	1.20	1.46	2.03	**2.17**	0.04*	0.04*	0.00*
Size-beme	**3.22**	1.13*	1.29*	0.77*	0.33*	1.11*	0.93*	−109.96*	−52.38*	−8.21*	−3.07*	0.30*	−0.64*	0.84*	0.04*	0.04*	0.06*
Size-ltr	**2.36**	0.47*	0.88*	0.54*	0.16*	0.79*	0.66*	−97.58*	−52.19*	−8.28*	−3.25*	−0.19*	−0.94*	0.48*	0.05*	0.04*	0.05*
Size-mom	**4.33**	2.01*	1.69*	1.17*	0.46*	1.35*	1.10*	−120.84*	−62.56*	−9.15*	−3.51*	0.28*	−0.83*	0.77*	0.04*	0.04*	0.04*
Size-str	9.33	**19.07***	16.15*	14.64*	7.44*	14.67*	13.71*	−118.26*	−885.76*	−283.79*	−152.71*	16.11*	−195.05*	−142.35*	0.04*	0.04*	−0.05*
Panel C: Ken French's equal-weighted daily data																	
Ind	**6.10**	4.20*	2.64*	1.49*	0.34*	2.32*	2.21*	−40.11*	−22.81*	−4.27*	−1.35*	0.35*	0.20*	0.90*	0.07*	0.04*	0.06*
Size	**4.04**	2.48*	1.98*	1.56*	0.71*	1.77*	1.68*	−23.18*	−16.12*	−1.78*	0.50*	1.32*	0.85*	1.37*	0.05*	0.04*	0.09*
Beme	**4.34**	2.44*	2.11*	1.65*	0.86*	1.86*	1.76*	−33.47*	−14.29*	0.25*	1.92*	1.73*	2.34*	2.69*	0.07*	0.04*	0.10*
Ltr	**4.01**	2.45*	2.01*	1.65*	0.84*	1.71*	1.61*	−27.63*	−12.22*	−0.12*	1.92*	1.87*	2.09*	2.38*	0.07*	0.04*	0.07*
Mom	**5.16**	2.31*	1.70*	1.09*	0.39*	1.45*	1.36*	−30.27*	−13.23*	−1.68*	−0.36*	0.67*	1.15*	1.57*	0.07*	0.04*	0.04*
Str	15.56	21.62*	18.72*	16.82*	13.05*	17.78*	17.36*	−12.98*	−1186.44*	−432.61*	−233.81*	**25.26***	−375.00*	−334.68*	0.05*	0.04*	−0.07*
Size-beme	**5.45**	3.10*	3.00*	2.32*	1.12*	2.51*	2.19*	−126.02*	−64.87*	−6.79*	−0.26*	2.83*	1.32*	3.05*	0.06*	0.04*	0.09*
Size-ltr	**4.33**	2.54*	2.56*	2.06*	0.98*	2.12*	1.84*	−101.82*	−59.57*	−7.03*	−0.61*	2.32*	1.00*	2.58*	0.06*	0.04*	0.07*
Size-mom	**6.27**	3.27*	2.77*	2.09*	1.03*	2.26*	1.95*	−122.64*	−72.75*	−8.73*	−1.76*	1.58*	0.12*	1.99*	0.05*	0.04*	0.06*
Size-str	18.52	33.91*	30.55*	27.66*	14.61*	28.28*	26.80*	−269.31*	−5158.95*	−1965.47*	−984.33*	**43.63***	−1492.08*	−1157.89*	0.05*	0.04*	−0.07*

Table 5 calculates certainty equivalents for several strategies on daily datasets when there are no restrictions on portfolio selection. Certainty equivalents are the daily percentage return on a risk-free asset that would give investors the same utility as their optimal strategy. See the caption to Table 2 for a detailed description of the investment strategies. A burn-in period of 1,000 days is used and the certainty equivalents are computed over the remaining sample. The highest certainty equivalent in each row is in boldface. For all algorithms (which require a risk-aversion parameter), risk aversion is one. For robust algorithms, model uncertainty aversion is four. For robust BA, we use a fixed sharing prior with $\alpha = 1$. A $\star$ indicates that the strategy significantly underperforms or outperforms robust BA at the 95% confidence level.

Table 6
Certainty equivalent percentiles

	Robust strategies							Nonrobust strategies									
	BA	R_1	R_2	R_3	H	JOR	KZ	BA	R_1	R_2	R_3	H	JOR	KZ	1/N	MKT	MINV
Panel A: Daily data																	
Min	**0.85**	−1.87	0.21	0.11	0.03	0.18	0.15	−269.31	−5158.95	−1965.47	−984.33	−0.19	−1492.08	−1157.89	0.03	0.03	−0.07
25%	**1.84**	0.68	0.53	0.31	0.16	0.46	0.44	−103.85	−63.14	−8.23	−2.29	0.06	−0.83	−0.28	0.04	0.04	0.03
Median	**3.62**	1.86	1.63	1.13	0.42	1.29	1.12	−38.45	−15.21	−3.05	−1.35	0.68	−0.17	0.80	0.05	0.04	0.04
75%	**5.23**	2.68	2.58	1.75	0.99	2.16	1.87	−23.23	−11.38	−1.64	0.01	1.76	0.93	1.68	0.06	0.04	0.06
Max	18.52	33.91	30.55	27.66	14.61	28.28	26.80	−3.18	−1.23	0.39	1.92	**43.63**	2.34	3.05	0.07	0.04	0.10
Panel B: Weekly data																	
Min	**0.49**	−3.81	0.15	0.13	0.12	0.24	0.20	−746.41	−75083.23	−13564.09	−4282.21	−1.15	−6753.64	−3450.64	0.17	0.17	0.02
25%	**1.86**	0.36	1.04	0.95	0.59	0.90	0.91	−126.08	−612.70	−71.89	−29.73	−0.22	−10.70	−1.42	0.20	0.17	0.15
Median	**3.80**	3.42	3.57	3.10	1.42	2.91	2.61	−48.57	−71.52	−17.02	−8.79	2.35	−4.14	−0.52	0.23	0.17	0.21
75%	**5.86**	5.75	5.60	5.24	3.13	4.41	3.83	−28.75	−38.28	−10.87	−5.32	5.29	−2.01	1.05	0.30	0.18	0.27
Max	32.13	40.74	51.45	48.75	25.71	44.74	39.15	−3.25	−3.38	−0.01	1.84	**52.75**	5.66	7.38	0.38	0.18	0.47
Panel C: Monthly data																	
Min	**1.02**	−1476.52	−1.20	0.58	0.43	0.79	0.72	−76.67	−946550.80	−469.38	−128.79	−0.35	−55.33	−4.42	0.79	0.79	0.51
25%	**1.44**	−837.35	0.77	1.03	0.80	1.35	1.15	−25.56	−418822.07	−229.33	−73.24	0.68	−23.47	−1.56	0.95	0.81	0.72
Median	**2.90**	−1.35	2.30	2.29	1.27	1.90	1.59	−12.08	−235.93	−36.72	−11.83	2.08	−4.94	−0.50	1.02	0.81	0.83
75%	4.32	0.83	4.08	**5.38**	3.25	4.01	2.84	−5.52	−168.91	−23.18	−10.11	2.99	−3.82	1.17	1.08	0.87	0.93
Max	6.39	11.17	**13.02**	11.27	4.37	8.88	6.84	3.30	−3.50	1.78	2.26	10.25	2.69	6.77	1.54	0.90	1.30
Panel D: Simulated i.i.d. data																	
Min	−1.29	−0.37	−0.15	−0.07	0.00	−0.04	−0.02	−44.52	−7.43	−2.53	−1.18	−0.15	−0.70	−0.43	**0.01**	—	0.00
25%	−1.01	−0.26	−0.09	−0.03	0.01	−0.02	−0.01	−41.70	−7.23	−2.33	−1.06	−0.09	−0.64	−0.38	**0.03**	—	0.02
Median	−0.89	−0.21	−0.06	−0.02	0.02	−0.01	0.00	−40.35	−7.03	−2.26	−1.02	−0.06	−0.61	−0.35	**0.03**	—	0.02
75%	−0.83	−0.19	−0.04	0.00	0.02	0.01	0.01	−38.94	−6.91	−2.16	−0.97	−0.05	−0.56	−0.31	**0.04**	—	0.02
Max	−0.52	−0.07	0.03	0.04	0.08	0.05	0.04	−35.50	−6.71	−2.01	−0.83	**0.11**	−0.46	−0.24	0.06	—	0.04
Panel E: Simulated regime changes																	
Min	**0.12**	0.11	0.00	−0.08	−0.01	0.03	0.02	−44.85	−7.85	−2.74	−1.67	−0.18	−0.85	−0.55	−0.05	—	−0.02
25%	**0.46**	0.28	0.13	0.07	0.02	0.13	0.12	−38.42	−7.06	−2.29	−1.09	−0.09	−0.54	−0.26	0.02	—	0.01
Median	**0.65**	0.42	0.21	0.12	0.03	0.21	0.18	−36.28	−6.81	−2.15	−1.00	−0.05	−0.44	−0.20	0.04	—	0.02
75%	**0.73**	0.49	0.27	0.15	0.06	0.25	0.23	−34.62	−6.62	−1.91	−0.88	0.06	−0.24	−0.01	0.06	—	0.03
Max	**1.15**	0.78	0.48	0.29	0.16	0.41	0.37	−30.79	−5.92	−1.41	−0.51	0.31	0.19	0.40	0.10	—	0.05

Table 6 summarizes the distribution of certainty equivalents on daily, weekly, monthly, and simulated data sets (when there are no restrictions on portfolio selection). For daily data, for each algorithm, we show the percentiles of the certainty equivalents listed in Table 5. For weekly and monthly data, we also report percentiles across the same 24 datasets. Panels D and E report certainty equivalent percentiles for 41 draws of 12,000 periods from the i.i.d. and regime-change data generating processes described in Appendix A. The number of burn-in periods, the risk-aversion parameter, and the model uncertainty aversion parameter follow Table 3. The highest certainty equivalent among all strategies, for each percentile is in boldface. In the simulated data sets, there is not a well-defined market portfolio. R_1, R_2, and R_3 refer to rolling approaches with window sizes of 100, 250, and 500 days for daily data; 50, 100, and 150 weeks for weekly data; and 30, 60, and 90 months for monthly data.

outperforms the other algorithms on all five percentiles, usually by a large margin.

7.2.2 Constrained portfolios. Table 7 presents out-of-sample certainty equivalents for robust and nonrobust strategies with short-selling constraints. Over the 24 daily datasets considered, robust BA and BA perform nearly identically and are statistically indistinguishable on 22 datasets. Robust BA is significantly better than each of the alternative algorithms is on at least 18 datasets.

8. Uncertainty and Prior Selection

Often in Bayesian settings, results depend crucially on the choice of priors and other arbitrarily fixed parameters and mechanisms. If future information is inadvertently introduced into priors, parameters, or mechanisms, then out-of-sample results may be contaminated. To avoid using future information, we tie prior means and covariances to past sample data. We set the confidence in our prior means and covariances, κ and δ, to be small so that they carry very little weight. Small changes in the specification of prior means and covariances will have little effect on our results. Two of our other arbitrary choices potentially could have a large effect on our results: the choice of a prior rule for updating the probabilities of all models when a new model is born and the choice of the model uncertainty aversion level.

In this section, we show that robust BA, when paired with several different prior rules and different values of model uncertainty aversion, also achieves excellent out-of-sample certainty equivalents.[24] We begin by discussing methods for selecting uncertainty aversion and prior rules each period using only currently available information. For comparison purposes, we also discuss alternative fixed values of uncertainty aversion and alternative fixed prior rules.

8.1 Robust Bayesian averaging with optimal uncertainty and priors

In earlier sections, we proposed a robust BA portfolio choice strategy with a fixed model uncertainty aversion parameter, and a fixed rule for determining the prior probabilities of models. In reality, the optimal setting of model uncertainty aversion may vary over time and there are many possible settings of prior probabilities. At each date, we now let investors choose the model uncertainty aversion parameter and the prior rule with the best past performance using currently available information.

We index portfolio choices and portfolio returns by model uncertainty aversion and the prior rule. The prior rule is a function that maps model

[24] The choice of the risk-aversion parameter also could potentially influence our results. In results available upon request, we demonstrate that we achieve similar results for several different values of risk aversion.

Table 7
Daily certainty equivalents for constrained portfolio choices

	Robust strategies							Nonrobust strategies						
	BA	R100	R250	R500	H	JOR	KZ	BA	R100	R250	R500	H	JOR	KZ
Panel A: CRSP daily data														
Beta	0.128	0.088⋆	0.090⋆	0.085⋆	0.082⋆	0.081⋆	0.077⋆	**0.131**	0.095⋆	0.101⋆	0.092⋆	0.089⋆	0.097⋆	0.093⋆
Size	**0.099**	0.069⋆	0.060⋆	0.050⋆	0.031⋆	0.058⋆	0.055⋆	0.099	0.071⋆	0.059⋆	0.053⋆	0.053⋆	0.060⋆	0.059⋆
Std	0.207	0.198	0.199	0.192	**0.213**	0.181⋆	0.169⋆	0.207	0.208	0.209	0.204	0.213	0.202	0.196
Em	**0.115**	0.067⋆	0.063⋆	0.058⋆	0.067⋆	0.063⋆	0.063⋆	0.114	0.070⋆	0.064⋆	0.065⋆	0.073⋆	0.065⋆	0.064⋆
Panel B: Ken French's value-weighted daily data														
Ind	**0.124**	0.052⋆	0.056⋆	0.042⋆	0.038⋆	0.049⋆	0.047⋆	0.121	0.055⋆	0.061⋆	0.039⋆	0.037⋆	0.057⋆	0.052⋆
Size	**0.091**	0.053⋆	0.052⋆	0.041⋆	0.036⋆	0.048⋆	0.045⋆	0.090	0.053⋆	0.053⋆	0.043⋆	0.039⋆	0.049⋆	0.049⋆
Beme	**0.078**	0.049⋆	0.050⋆	0.039⋆	0.044⋆	0.047⋆	0.044⋆	0.077	0.050⋆	0.050⋆	0.042⋆	0.047⋆	0.052⋆	0.050⋆
Ltr	**0.092**	0.051⋆	0.056⋆	0.037⋆	0.035⋆	0.052⋆	0.048⋆	0.092	0.053⋆	0.055⋆	0.038⋆	0.042⋆	0.056⋆	0.053⋆
Mom	0.107	0.046⋆	0.047⋆	0.032⋆	0.050⋆	0.041⋆	0.037⋆	**0.111**	0.050⋆	0.053⋆	0.034⋆	0.057⋆	0.042⋆	0.044⋆
Str	0.115	0.109	0.119	0.122	**0.146⋆**	0.105	0.101	0.118	0.111	0.128	0.136⋆	0.146⋆	0.121	0.117
Size-beme	**0.100**	0.060⋆	0.056⋆	0.047⋆	0.055⋆	0.061⋆	0.060⋆	0.097	0.057⋆	0.054⋆	0.048⋆	0.055⋆	0.062⋆	0.061⋆
Size-ltr	**0.091**	0.052⋆	0.061⋆	0.050⋆	0.056⋆	0.059⋆	0.055⋆	0.089	0.054⋆	0.057⋆	0.053⋆	0.055⋆	0.061⋆	0.061⋆
Size-mom	**0.118**	0.065⋆	0.068⋆	0.060⋆	0.073⋆	0.068⋆	0.066⋆	0.115	0.067⋆	0.071⋆	0.061⋆	0.075⋆	0.071⋆	0.071⋆
Size-str	0.210	0.210	0.221	0.229	**0.247⋆**	0.211	0.208	0.205	0.216	0.236⋆	0.241⋆	0.246⋆	0.223	0.222
Panel C: Ken French's equal-weighted daily data														
Ind	**0.224**	0.136⋆	0.122⋆	0.092⋆	0.085⋆	0.119⋆	0.116⋆	0.224	0.139⋆	0.123⋆	0.087⋆	0.083⋆	0.125⋆	0.124⋆
Size	**0.131**	0.099⋆	0.099⋆	0.088⋆	0.100⋆	0.098⋆	0.098⋆	0.130	0.098⋆	0.099⋆	0.087⋆	0.100⋆	0.098⋆	0.099⋆
Beme	0.143	0.110⋆	0.112⋆	0.107⋆	0.121⋆	0.112⋆	0.111⋆	**0.145**	0.109⋆	0.114⋆	0.108⋆	0.121⋆	0.113⋆	0.111⋆
Ltr	0.187	0.150⋆	0.148⋆	0.146⋆	0.156⋆	0.145⋆	0.142⋆	**0.188**	0.153⋆	0.155⋆	0.149⋆	0.156⋆	0.151⋆	0.152⋆
Mom	0.207	0.130⋆	0.125⋆	0.108⋆	0.115⋆	0.117⋆	0.114⋆	**0.212⋆**	0.134⋆	0.136⋆	0.116⋆	0.114⋆	0.130⋆	0.124⋆
Str	0.480	0.471	0.487	0.493⋆	**0.498⋆**	0.479	0.478	0.484⋆	0.486	0.496⋆	0.498⋆	0.498⋆	0.491	0.490
Size-beme	**0.142**	0.107⋆	0.110⋆	0.105⋆	0.122⋆	0.114⋆	0.115⋆	0.138	0.107⋆	0.109⋆	0.098⋆	0.121⋆	0.110⋆	0.112⋆
Size-ltr	**0.151**	0.129⋆	0.135	0.137	0.151	0.133	0.129⋆	0.149	0.126⋆	0.139	0.140	0.151	0.139	0.133
Size-mom	**0.170**	0.123⋆	0.126⋆	0.113⋆	0.118⋆	0.122⋆	0.117⋆	0.167	0.125⋆	0.130⋆	0.118⋆	0.118⋆	0.130⋆	0.127⋆
Size-str	0.479	0.515⋆	0.534⋆	0.532⋆	**0.541⋆**	0.524⋆	0.521⋆	0.476	0.524⋆	0.541⋆	0.536⋆	0.541⋆	0.537⋆	0.536⋆

Table 7 calculates certainty equivalents for several different strategies on daily datasets when short selling is prohibited. This table parallels Table 5. A burn-in period of 1,000 days is used, and certainty equivalents are computed over the remaining sample. The highest certainty equivalent in each row is in boldface. For all algorithms, risk aversion is one. For robust algorithms, model uncertainty aversion is four. For robust BA, we use a fixed sharing prior with $\alpha = 1$. A ⋆ indicates that the strategy significantly underperforms or outperforms robust BA at the 95% confidence level.

probabilities before model t is born, $\{P_{t-1}(m|\mathcal{F}_{t-1})\}_{m=1}^{t-1}$, to the probabilities after model t is born, $\{P_t(m|\mathcal{F}_{t-1})\}_{m=1}^{t}$. Let Ω be the set of possible values of model uncertainty aversion and Π be the set of possible prior rules. Let $\phi_s(\tau,\pi)$ be the optimal portfolio weight, at time s, and $R_{ps+1}(\tau,\pi)=\phi_s(\tau,\pi)' R_{s+1}$ be the subsequent excess portfolio return when model uncertainty aversion is $\tau \in \Omega$ and the prior rule is $\pi \in \Pi$ at all current and past dates.[25]

At each date $t>1$, investors choose the model uncertainty aversion parameter and the prior rule to be the constant value and constant rule that maximize the in-sample certainty equivalent:

$$\tau^*,\pi^* = \mathrm{argmax}_{\tau\in\Omega,\pi\in\Pi}\left(\bar{E}_t\left[R_{ps}(\tau,\pi)\right]-\frac{\theta}{2}\bar{V}_t\left[R_{ps}(\tau,\pi)\right]\right),$$

where $\bar{E}_t\left[R_{ps}(\tau,\pi)\right]$ and $\bar{V}_t\left[R_{ps}(\tau,\pi)\right]$ denote the sample mean and variance of portfolio returns up until time t, when model uncertainty aversion is fixed at τ and the prior rule is fixed at π at every past period.[26] The choices are used to form portfolio weights $\phi_t(\tau^*,\pi^*)$ that realize a return of $R_{pt+1}(\tau^*,\pi^*)= \phi_t(\tau^*,\pi^*)' R_{t+1}$ at time $t+1$.

We refer to this method as robust BA with optimal uncertainty and priors when investors choose both τ and π, and $\phi_s(\tau,\pi)$ solves the robust BA portfolio choice problem in Section 5.2 for the fixed model uncertainty aversion parameter τ and the fixed prior rule π. We also consider two special cases: robust BA with optimal uncertainty, which takes π as given and lets investors choose τ, and robust BA with optimal priors, which takes τ as given and lets investors choose π.

8.2 Optimal uncertainty for alternative robust algorithms

The recursive approach to estimating model uncertainty aversion can be applied to the other robust algorithms in Section 6.8. When the method is applied to the rolling, historical, Jorion, and Kan-Zhou approaches, $\phi_s(\tau,\pi)$ solves the robust portfolio choice problem in Section 6.8 for the fixed model uncertainty aversion parameter τ. These algorithms do not use priors so their optimal versions only compute τ^*. The prior rule, π, does not affect portfolio choices.

8.3 Robust Bayesian averaging with worst-case priors

This section proposes a portfolio choice problem where investors choose the portfolio that performs the best under the worst-possible prior. As in Section 8.1, we assume there are many possible prior rules and we let Π denote the set of

[25] In general, as the next paragraph describes in more detail actual portfolio choices will depend on time-varying model uncertainty aversion and time-varying prior rules. However, when selecting the current period's model uncertainty aversion and prior rule, investors imagine model uncertainty aversion and the prior rule as being fixed in the past.

[26] It is possible, though unlikely in practice, that multiple values achieve the maximum. If multiple values achieve the maximum, then we randomly select (τ^*,π^*) from the set of all (τ,π) that achieve the maximum.

all such rules. For the prior rule $\pi \in \Pi$, let $\{P_{\pi t}(m|\mathcal{F}_{t-1})\}_{m=1}^{t}$ be the model probabilities after model t is born but before R_t is observed. After observing R_t, the model probabilities are

$$P_{\pi t}(m|\mathcal{F}_t) = \frac{L(R_t|m, \mathcal{F}_{t-1})\, P_{\pi t}(m|\mathcal{F}_{t-1})}{\sum_{m \in M_t} L(R_t|m, \mathcal{F}_{t-1})\, P_{\pi t}(m|\mathcal{F}_{t-1})}$$

by Bayes rule.

Investors are unsure which prior rule is appropriate and want to choose portfolios that perform well on all possible priors. The optimal portfolio choices for a robust BA investor, who uses the worst-case prior, solve the max-min problem:

$$\max_{\phi_t} \min_{\pi \in \Pi} \left(\phi_t' \hat{\mu}_{\pi t} + R_{f t+1} - \frac{1}{\tau} \log \sum_{m \in M_t} U_{m,t}\, P_{\pi t}(m|\mathcal{F}_t) \right),$$

where

$$\hat{\mu}_{\pi t} = \sum_{m \in M_t} \mu_{m,t}\, P_{\pi t}(m|\mathcal{F}_t)$$

and $U_{m,t}$, which depends on ϕ_t, was defined in Equation (21). Let π^* be the minimizing choice of the prior rule.[27] For all m, the investor chooses the probabilities $P_t(m|\mathcal{F}_t) = P_{\pi^* t}(m|\mathcal{F}_t)$. Investors achieve good performance on all priors by choosing the best portfolio under the worst prior. The worst prior is endogenous and depends upon portfolio choices.

This formulation of worst-case priors incorporates many features from the recursive multiple-priors method (Epstein and Schneider [2003]); the smooth ambiguity literature (Klibanoff, Marinacci, and Mukerji [2005]); and, especially, the T1-T2 approach (Hansen and Sargent [2007a]). The T1-T2 approach introduces two operators (T1 and T2) and two corresponding robustness parameters. Our robust optimization problem for a fixed prior, described in Section 5.2, is a version of the T1 operator, and our τ is its corresponding robustness parameter. The minimization over prior rules in this section is an extreme variation on the T2 operator, where investors select prior rules as the corresponding robustness parameter tends toward infinity.

8.4 Evaluating uncertainty selections

Columns 2 through 8 of Table 8 present certainty equivalent percentiles for robust optimal uncertainty versions of many algorithms when the set of possible model uncertainty aversion parameters is $\Omega = \{0, 1, 2, 4, 8\}$. For robust BA with optimal uncertainty, we use a fixed sharing prior rule with $\alpha = 1$ (perfect sharing). Robust BA with optimal uncertainty performs much better than

[27] If multiple values of π achieve the minimum, then we randomly select π^* to be one of them.

Table 8
Certainty equivalent percentiles for robust strategies with optimal and alternative fixed model uncertainty aversion parameters

	Robust strategies with optimal uncertainty							Robust strategies with $\tau=2$							Robust strategies with $\tau=8$						
	BA	R_1	R_2	R_3	H	JOR	KZ	BA	R_1	R_2	R_3	H	JOR	KZ	BA	R_1	R_2	R_3	H	JOR	KZ
Panel A: Daily data																					
Min	**0.82**	−0.34	0.20	0.09	0.03	0.19	0.16	0.18	−6.31	0.06	0.07	0.04	0.20	0.18	0.63	−0.29	0.17	0.09	0.02	0.12	0.10
25%	**1.90**	0.82	0.70	0.46	0.21	0.65	0.68	1.69	0.23	0.65	0.41	0.21	0.60	0.60	1.36	0.65	0.37	0.20	0.11	0.30	0.28
Median	**3.92**	2.00	2.20	1.59	0.72	2.15	1.99	3.92	1.77	2.08	1.48	0.60	1.81	1.63	2.67	1.38	1.04	0.73	0.26	0.80	0.69
75%	**6.26**	2.96	3.46	2.98	1.89	3.62	3.40	6.11	2.88	3.34	2.49	1.48	3.04	2.72	3.65	2.20	1.69	1.09	0.60	1.34	1.15
Max	32.22	34.07	34.60	36.59	**41.39**	34.81	35.96	24.92	31.36	36.33	35.34	22.06	34.50	33.46	12.76	27.62	22.29	19.35	8.89	20.33	19.00
Panel B: Weekly data																					
Min	0.46	−0.17	0.17	0.11	0.12	0.18	0.15	−0.60	−16.51	−0.59	−0.47	0.12	0.15	0.17	**0.61**	−0.16	0.25	0.19	0.11	0.22	0.18
25%	**1.94**	0.85	1.06	0.96	0.72	1.20	1.23	1.51	−7.54	0.24	0.57	0.76	1.20	1.21	1.35	0.88	0.71	0.63	0.40	0.59	0.60
Median	4.25	4.59	4.55	3.83	2.31	**4.63**	4.38	4.17	−0.77	4.54	3.94	2.00	4.04	3.71	2.68	3.47	2.54	2.06	0.90	1.86	1.65
75%	7.24	6.32	6.65	6.94	5.55	7.02	6.64	**7.25**	5.31	6.62	6.46	4.59	5.98	5.47	4.46	4.75	3.83	3.56	1.93	2.86	2.46
Max	59.53	44.75	56.00	58.34	51.97	48.09	52.49	45.02	33.31	52.64	**60.14**	37.28	51.08	48.15	21.66	47.78	40.64	35.15	16.36	33.53	28.18
Panel C: Monthly data																					
Min	0.85	−632.41	0.63	0.43	0.39	0.60	0.59	**1.05**	−3303.05	−7.89	−1.68	0.47	0.73	0.79	0.81	−632.41	0.72	0.60	0.39	0.67	0.59
25%	**1.55**	−347.13	0.93	1.03	0.84	1.09	1.21	1.43	−1910.87	−0.72	0.85	1.07	1.19	1.40	1.12	−347.13	0.97	1.06	0.60	0.96	0.86
Median	2.95	0.29	2.20	2.42	1.53	2.04	2.00	**3.30**	−6.67	0.95	2.46	1.69	2.29	2.05	2.12	0.29	1.89	1.71	0.90	1.37	1.13
75%	5.11	1.97	4.19	5.58	4.91	3.94	3.87	5.62	−4.01	2.75	**5.68**	4.53	4.43	3.81	2.99	1.85	3.88	3.69	2.15	2.90	1.92
Max	9.79	10.91	13.39	**13.61**	7.43	13.30	12.28	8.67	9.70	13.20	13.22	6.27	11.65	9.73	4.33	9.01	10.28	8.06	2.83	5.95	4.40
Panel D: Simulated i.i.d. data																					
Min	−0.50	−0.14	−0.06	−0.04	−0.01	−0.03	−0.02	−3.10	−0.88	−0.33	−0.17	−0.02	−0.10	−0.06	−0.50	−0.14	−0.06	−0.03	**0.00**	−0.02	−0.01
25%	−0.34	−0.07	−0.03	−0.01	**0.01**	−0.00	0.00	−2.60	−0.71	−0.25	−0.11	0.00	−0.06	−0.04	−0.34	−0.07	−0.03	−0.00	0.01	0.00	0.00
Median	−0.29	−0.05	−0.01	0.00	**0.01**	0.00	0.00	−2.42	−0.64	−0.21	−0.08	0.01	−0.04	−0.02	−0.29	−0.05	−0.01	0.00	0.01	0.01	0.01
75%	−0.24	−0.04	0.00	0.01	0.02	0.01	0.01	−2.33	−0.61	−0.17	−0.06	**0.03**	−0.02	0.00	−0.24	−0.04	0.01	0.01	0.02	0.01	0.02
Max	−0.07	0.04	0.05	0.03	**0.12**	0.04	0.04	−1.79	−0.44	−0.07	0.00	0.11	0.04	0.04	−0.07	0.04	0.05	0.04	0.05	0.04	0.03
Panel E: Simulated regime changes																					
Min	0.36	0.12	0.03	−0.03	−0.01	0.02	0.02	−1.01	−0.19	−0.13	−0.21	−0.02	−0.01	0.00	**0.38**	0.15	0.03	−0.03	0.00	0.03	0.02
25%	**0.55**	0.27	0.11	0.06	0.01	0.13	0.14	−0.38	0.04	0.07	0.04	0.01	0.15	0.15	0.55	0.26	0.11	0.06	0.01	0.09	0.08
Median	0.64	0.41	0.20	0.10	0.03	0.23	0.22	−0.13	0.26	0.18	0.11	0.04	0.26	0.23	**0.66**	0.34	0.16	0.09	0.02	0.14	0.11
75%	0.74	0.48	0.26	0.14	0.10	0.32	0.33	0.05	0.38	0.28	0.16	0.08	0.33	0.32	**0.74**	0.38	0.19	0.11	0.04	0.16	0.15
Max	**1.13**	0.74	0.56	0.36	0.28	0.63	0.62	0.58	0.81	0.58	0.38	0.23	0.56	0.53	1.00	0.56	0.32	0.19	0.10	0.26	0.23

Table 8 summarizes the distribution of certainty equivalents (when there are no restrictions on portfolio selection) for robust versions of the BA, rolling, historical, Jorion, and Kan-Zhou approaches when uncertainty aversion is estimated optimally (Columns 2 through 8), when uncertainty aversion is fixed at 2 (Columns 9 through 15), and when uncertainty aversion is fixed at 8 (Columns 16 through 22). The possible values of model uncertainty aversion are $\Omega=\{0,1,2,4,8\}$ for the optimal uncertainty versions in Columns 2 through 8. The highest certainty equivalent in each row is in boldface. We set risk aversion to one for all algorithms. The robust BA strategies, with and without optimal uncertainty, use a fixed sharing prior ($\alpha=1$). R_1, R_2, and R_3 refer to rolling approaches with window sizes of 100, 250, and 500 days for daily data; 50, 100, and 150 weeks for weekly data; and 30, 60, and 90 months for monthly data. The number of burn-in periods follows Table 3.

optimal uncertainty versions of other algorithms. It also usually outperforms robust BA with a fixed uncertainty aversion parameter when data are plentiful (daily data), but it slightly underperforms when data are limited (monthly data).

Columns 9 through 22 of Table 8 display certainty-equivalent percentiles when model uncertainty aversion is fixed at two or eight. This complements results in Table 6 for model uncertainty aversion parameters of zero (the nonrobust case) and four. Overall, on actual data, model uncertainty aversion parameters of two and four achieve excellent performance, with two slightly outperforming four. Although a model uncertainty aversion parameter of eight underperforms, it still usually obtains higher out-of-sample certainty equivalents than other robust (non-BA) algorithms. On simulated data with regime changes, model-uncertainty aversion parameters of four and eight achieve excellent performance, with eight slightly outperforming four.

8.5 Evaluating prior selections

Table 9 presents certainty-equivalent percentiles for robust BA with optimal uncertainty and priors (BAUP), robust BA with optimal priors (BAP), and robust BA with worst-case priors (BAW), when the value of α for the sharing prior is selected from the menu {1, 0.75, 0.5, 0.25, 0.001} and the value of β for the power prior is selected from the menu $\{0.5, 0.6, 0.75, 0.9, 1\}$. Results are also provided for several fixed-sharing and power priors. Generally, the BAUP algorithm with sharing priors achieves the highest certainty equivalents. When BAUP underperforms, its certainty equivalents are nearly identical to the best performer.

Table 9 shows that robust BA with a fixed sharing prior or a fixed power prior earns higher certainty equivalents than other robust (non-BA) algorithms listed in Table 6 for a wide range of values for α or β on actual daily, weekly, and monthly data. The perfect sharing prior ($\alpha = 1$) obtains the best performance, but other sharing priors with $\alpha \geq 0.25$ also usually achieve higher certainty equivalents than robust versions of the rolling, historical, Jorion, and Kan-Zhou approaches do. Among fixed power priors, $\beta = 0.6$ obtains the best results, but other values between 0.5 and 0.9 also have good performance. One reason large values of α and relatively small values of β work well is that they both downweigh older models. The $1/t$ prior (which is equivalent to the power prior with $\beta = 1$) and the equal-weighted prior do not perform as well but are competitive with other (non-BA) strategies.[28]

To illustrate the perfect sharing prior (the sharing prior with $\alpha = 1$), Figure 1 plots prior, $P_t(m|\mathcal{F}_{t-1})$, and posterior, $P_t(m|\mathcal{F}_t)$, model probabilities for December 17, 18, and 19, 2008 using the BA algorithm, where the excess returns

[28] Results for simulated data are different. With simulated i.i.d. data, it is better to use as long a sample as possible to estimate means and variances. Hence, sharing priors with small values of α and the power prior with $\beta = 1$ (which is identical to the $1/t$ prior) produce the best results. For the simulated data with regime changes, sharing priors with $\alpha \geq 0.25$ work well, but power priors, for all values of β, do not work well.

Table 9
Certainty equivalent percentiles for robust BA strategies with different priors:

	Sharing priors								Power priors								
	BAUP	BAP	BAW	1.00	0.75	0.50	0.25	0.001	BAUP	BAP	BAW	0.50	0.60	0.75	0.90	1.00	EW
Panel A: Daily data																	
Min	0.82	**0.85**	0.34	0.85	0.70	0.56	0.43	0.20	0.64	0.37	0.55	0.06	0.14	0.45	0.41	0.03	0.05
25%	**1.99**	1.81	0.99	1.84	1.61	1.36	1.11	0.69	1.74	1.66	1.56	1.35	1.42	1.59	1.59	0.36	0.26
Median	**3.92**	3.63	2.38	3.62	2.87	2.52	2.26	1.72	3.49	3.53	3.08	3.35	3.50	3.55	3.31	1.36	1.07
75%	**6.40**	5.20	4.43	5.23	4.76	4.45	4.15	3.02	5.68	5.34	4.95	5.29	5.37	5.29	4.87	2.07	1.68
Max	39.97	22.26	19.27	18.52	21.44	21.97	22.22	22.56	**40.43**	23.65	21.49	21.51	22.44	23.33	23.81	21.90	19.40
Panel B: Weekly data																	
Min	0.42	0.44	0.28	**0.49**	0.44	0.42	0.35	0.17	0.26	−0.22	0.08	−0.18	−0.14	−0.07	0.07	0.02	0.13
25%	**2.10**	1.84	0.95	1.86	1.29	1.17	1.13	0.82	1.79	1.93	1.48	1.98	1.81	1.54	1.28	0.50	0.71
Median	**5.69**	3.88	3.74	3.80	3.95	3.81	3.67	3.25	4.29	4.00	3.67	3.88	3.92	4.01	3.61	2.55	2.48
75%	**8.27**	5.93	5.81	5.86	6.18	6.13	6.04	5.24	6.80	5.89	5.71	6.28	6.16	6.09	5.88	3.67	4.21
Max	59.09	35.51	32.62	32.13	34.36	34.83	35.10	35.85	**62.35**	35.85	34.30	34.34	35.11	35.50	36.13	32.67	28.17
Panel C: Monthly data																	
Min	0.51	0.78	0.48	1.02	0.79	0.63	0.52	0.40	0.59	0.96	0.68	1.06	**1.17**	0.83	0.75	0.45	0.51
25%	**1.51**	1.38	1.06	1.44	1.37	1.28	1.25	1.23	1.15	1.38	1.41	1.50	1.51	1.48	1.34	1.21	1.13
Median	2.67	2.56	1.75	**2.90**	2.67	2.36	2.18	1.58	2.17	2.19	2.18	2.58	2.73	2.50	2.18	1.67	1.69
75%	**6.73**	4.28	3.57	4.32	4.00	4.02	4.07	4.08	4.95	3.94	4.12	4.21	4.31	4.32	4.10	4.05	3.97
Max	**11.62**	6.59	6.40	6.39	6.52	6.51	6.50	6.59	11.22	6.47	6.12	5.70	5.71	6.09	6.47	6.48	6.24
Panel D: Simulated i.i.d. data																	
Min	−0.03	−0.05	−1.30	−1.29	−0.79	−0.49	−0.24	−0.01	−0.03	**0.00**	−2.72	−3.30	−3.35	−3.36	−3.13	0.00	−0.02
25%	−0.01	0.00	−0.99	−1.01	−0.69	−0.38	−0.16	**0.01**	0.01	0.01	−2.31	−2.78	−2.83	−2.85	−2.55	0.01	0.01
Median	0.00	0.01	−0.90	−0.89	−0.58	−0.31	−0.14	**0.02**	0.01	0.02	−2.17	−2.62	−2.70	−2.71	−2.40	0.02	0.01
75%	0.01	0.02	−0.82	−0.83	−0.51	−0.28	−0.11	**0.03**	0.02	0.02	−2.06	−2.53	−2.58	−2.58	−2.28	0.02	0.02
Max	0.10	0.05	−0.33	−0.52	−0.39	−0.18	0.01	0.08	**0.12**	0.08	−1.72	−2.25	−2.25	−2.21	−1.77	0.08	0.06
Panel E: Simulated regime changes																	
Min	**0.39**	0.26	0.13	0.12	0.25	0.31	0.21	0.02	−0.02	−0.16	−0.99	−1.18	−1.24	−1.18	−1.04	−0.00	0.04
25%	0.55	0.53	0.46	0.46	**0.56**	0.55	0.42	0.06	0.28	−0.04	−0.45	−0.71	−0.72	−0.80	−0.69	0.01	0.08
Median	0.65	0.65	0.63	0.65	**0.69**	0.68	0.56	0.10	0.46	0.01	−0.29	−0.47	−0.53	−0.57	−0.45	0.03	0.10
75%	0.75	0.76	0.72	0.73	**0.78**	0.74	0.61	0.13	0.60	0.04	−0.06	−0.25	−0.28	−0.36	−0.24	0.06	0.14
Max	1.08	1.13	1.14	1.15	**1.19**	1.03	0.85	0.25	0.90	0.35	0.65	0.41	0.40	0.44	0.52	0.16	0.21

Table 9 summarizes the distribution of certainty equivalents (when there are no restrictions on portfolio selection) for various sharing priors (Columns 2 through 9) and various power priors (Columns 10 through 17). BAUP is robust BA with optimal uncertainty and priors. BAP is robust BA with optimal priors. BAW is robust BA with worst-case priors. The BAUP, BAP, and BAW strategies in Columns 2 through 4 select the value of α for the sharing prior from the menu $\{1, 0.75, 0.5, 0.25, 0.001\}$. The strategies in Columns 5 through 9 use fixed sharing priors with α equal to 1, 0.75, 0.5, 0.25, and 0.001 (respectively). The BAUP, BAP, and BAW strategies in Columns 10 through 12 select the value of β for the power prior from the menu $\{0.5, 0.6, 0.75, 0.9, 1\}$. The strategies in Columns 13 through 17 use fixed power priors with β equal to 0.5, 0.6, 0.75, 0.9, and 1 (respectively). For BAUP, the possible values of model uncertainty aversion are $\Omega = \{0, 1, 2, 4, 8\}$. EW is the equal weighted prior. The highest certainty equivalent in each row is in boldface. For all algorithms, risk aversion is one. All strategies, except BAUP, fix model uncertainty aversion at four. The number of burn-in periods follows Table 3.

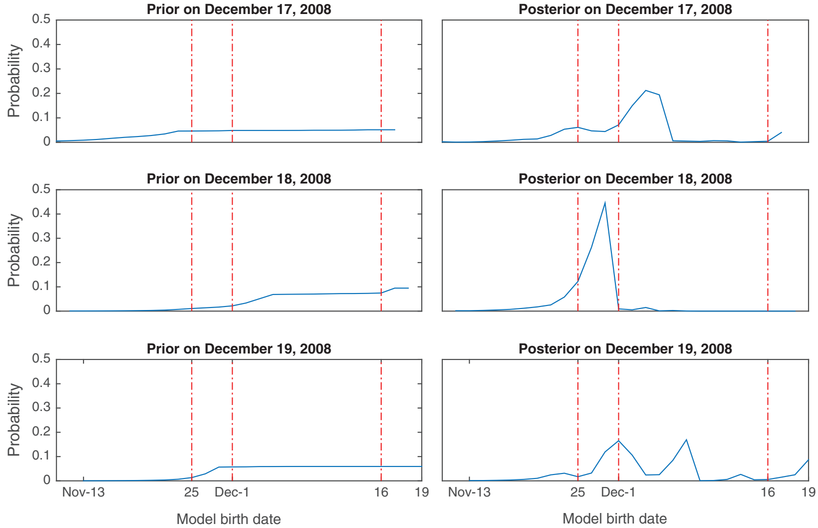

Figure 1
An example of model probabilities
Figure 1 plots prior, $P_t(m|\mathcal{F}_{t-1})$, and posterior, $P_t(m|\mathcal{F}_t)$, model probabilities for 3 consecutive days in December 2008 for the market and the 10 value-weighted industry portfolios using the BA algorithm with the sharing prior, $\alpha = 1$. Model probabilities are identical under the robust BA algorithm. For each plot, the y-axis provides the probabilities for the 25 youngest models born on the days listed on the x-axis. The three dashed lines correspond to three major financial events. On November 25, 2008, the Federal Reserve announced the first quantitative easing (QE) policy. On December 1, the Federal Reserve released more details about QE. On December 16, the Federal Reserve lowered the federal funds rate.

are on the market and the 10 value-weighted industry portfolios.[29] Recall that November 25 and December 1 and 16 are important dates in financial history. On November 25, 2008, the Federal Reserve announced the first quantitative easing (QE) policy. On December 1, the Federal Reserve released more details about QE, and on December 16, the Federal Reserve lowered the federal funds rate.[30] The priors and posteriors show that the total probability of all models born before November 25 is much less than the total probability of models born after. The posteriors substantially vary from day to day: on December 18, models born after December 1 receive negligible probabilities, whereas, on December 17 and 19, they receive substantial probability. The priors for models born after December 1 are similar and smooth on December 17, 18, and 19.

The solution to the BAW problem, where the agent selects from a menu of possible sharing priors (different values of α), entails that the agent chooses large values of α a majority of the time, on most datasets. This happens because newer models are less precise and thus usually more worrisome for a robust

29 The probabilities are identical under the robust BA algorithm.

30 See Gagnon and others (2010) for a discussion of these events.

agent than older models are.[31] The solutions to the BAP and BAUP problems lead to investors choosing α to be near one, most of the time. This happens because large values of α consistently work well throughout our sample. For example, investors choosing the best prior in 1960, looking only at data between 1926 and 1959, typically choose the same prior rule, as investors who have access to all data between 1926 and 2011. Even though there is little evidence that investors in 1926 used a perfect sharing prior, they would have learned to use a large value of α over time.

In detailed results available upon request, when investors use the BAP algorithm, they choose the perfect sharing prior 50% or more of the time on 19 of 24 daily datasets. On 11 of these datasets, they choose the perfect sharing prior 100% of the time. Three of the five datasets for which perfect sharing is chosen less than 50% of time, involve the short-term reversal portfolios. Results are virtually identical for the BAUP algorithm and similar for the BAW algorithm. Because, investors usually choose large values of α when given the option, we view robust BA with a fixed perfect sharing rule ($\alpha = 1$) as producing a computationally simpler approximate solution to the BAP and BAUP problems, which imposes the usually optimal prior rule.

9. Reasonable Uncertainty

In previous sections, because investors suspect that their best available approximating specification of excess returns could be wrong, they compute portfolio choices by solving a robust optimization problem. The solution to the robust optimization problem determines an alternative specification that investors, with a given value of model uncertainty aversion, should be concerned about. As model uncertainty aversion increases, the alternative specification will be farther from the approximating specification and represent a relatively worse outcome. This section investigates if a model uncertainty aversion parameter of four, in the robust BA algorithm, yields a plausible alternative specification that is not too far from the approximating specification. We measure plausibility by computing detection-error probabilities, which are the probabilities that an econometrician would mistakenly classify observations as coming from the approximating specification when they are actually generated by the alternative specification or vice versa.

Using the notation in Section 5, the approximating and alternative specifications take different stands on the probability density function (pdf) of z_{t+1}. The approximating specification assumes that the pdf of z_{t+1} is $f(z_{t+1}|\mathcal{F}_t)$ and the alternative specification takes the pdf of z_{t+1} as $f(z_{t+1}|\mathcal{F}_t)\varrho_t^*(z_{t+1})$, where the formula for ϱ^* is presented in Equation (17). The specification of f and the optimal perturbation ϱ^* vary across algorithms. For example, robust

[31] Robust investors with worst-case priors do not always find newer models more worrisome. Even though newer models are less precise, the worst-possible model can be an older model with dire predictions.

investors with Bayesian averaging beliefs take f to be a mixture of normals and investors with rolling expectations take f to be normal.

At each date, we compute detection-error probabilities by simulating a large number of time series from the approximating density and an equal number of time series from the alternative density. We classify each realization as either coming from f or $f\varrho_t^*$, using likelihood methods. We estimate the detection-error probability with the simulated rate of misclassification. If the approximating and alternative specifications are nearly indistinguishable, we expect the detection error to be near 50% since (asymptotically as the number of simulations tends towards infinity) our selection method will never be wrong more than 50% of the time, no matter how similar the specifications. If the approximating and alternative distributions are far apart, we expect the detection-error probability to be close to zero because the approximating and alternative specifications can easily be distinguished.

If the detection-error probabilities are large, then a robust agent's worries about the approximating specification are justified. If detection errors are small, then the robust agent may be worrying too much about nearly impossible outcomes. The threshold level of the detection-error probability, that determines large and small, is a matter of personal preference. On our view, a reasonable threshold is about 10% or 20%, so that robust agents should worry about specifications with detection-error probabilities greater than 10% or 20%. We find that the median (over time) detection-error probability for the robust BA algorithm is greater than 30% for all of our daily, weekly and monthly actual datasets when model uncertainty aversion is four and the prior rule is perfect sharing. This suggests that robust agents are justified in worrying about the alternative specification implied by the robust BA algorithm.[32]

10. Conclusion

This study proposes robust and nonrobust BA algorithms with a large number of models that preserve the simplicity of Markowitz's approach and achieve superior out-of-sample performance on a majority of 24 datasets. The approaches are fast, are adaptable to a large number of assets, and work in nonstationary environments. Our methods are similar to standard rolling window methods except that we do not prespecify the window size. Instead, we estimate the probability that each possible window size describes past data. The probabilities are re-computed each period based on available statistical evidence.

On the vast majority of datasets, the robust BA algorithm achieves decisively higher Sharpe ratios and certainty equivalents than rolling window schemes, the historical-expectations method, the 1/N approach, and other leading strategies

[32] We compute detection-error probabilities by adapting methods proposed by Hansen and Sargent (2007b) and others. The computational details are available upon request.

do. The robust BA algorithm also outperforms the nonrobust BA algorithm by achieving statistically higher out-of-sample Sharpe ratios (when short selling is prohibited) and statistically higher certainty equivalents (when short selling is allowed). The robust and nonrobust BA algorithms are statistically indistinguishable on out-of-sample Sharpe ratios (when short selling is allowed) and certainty equivalents (when short selling is prohibited).

To avoid arbitrary selections of a model uncertainty aversion parameter and a prior rule, we extend robust BA to allow agents to recursively choose the optimal value of model uncertainty aversion and the optimal prior rule at each date using only currently available information. On most datasets, the recursive extensions improve the performance of robust BA. We also show that robust BA achieves excellent performance for a wide range of fixed model uncertainty aversion parameters and a wide range of fixed prior rules.

We leave it to future work to improve the performance of the robust BA algorithm on reversal portfolios. Other limitations of our work include that we ignore background risk, taxes, transaction costs, and the dynamic nature of decision making. An interesting extension is to combine the robust BA algorithm with beliefs in asset pricing models as in Pastor (2000).

Appendix A. Simulated Data

We simulate daily data under two different processes. The first data generating process assumes i.i.d. returns and follows the specification of MacKinlay and Pastor (2000) and Garlappi, Uppal, and Wang (2007). We let the daily risk-free rate be i.i.d. normal

$$R_{ft} \sim \mathcal{N}\left(\frac{0.02}{252}, \frac{0.02^2}{252}\right)$$

so that the yearly risk-free rate approximately has a mean of 2% and a standard deviation of 2%. There is a single underlying factor that drives excess returns on 10 available risky assets. The daily distribution of the factor is

$$g_t \sim \mathcal{N}\left(\bar{\mu}_g, \bar{\sigma}_g^2\right),$$

where $\bar{\mu}_g = 0.08/252$ and $\bar{\sigma}_g = 0.16/\sqrt{252}$ are constant over time. The yearly approximate mean of the factor is 8% with a standard deviation of 16%. Excess returns on the 10 assets are distributed

$$R_t \sim \mathcal{N}(Bg_t, \Sigma_s),$$

where B is a 10-dimensional vector and Σ_s is a 10×10 diagonal matrix. The values of B are constant over time and evenly spaced between 0.5 and 1.5. The diagonal values of Σ_s are also constant over time and randomly selected from a uniform distribution over the interval $[\frac{0.1}{252}, \frac{0.3}{252}]$, so that yearly idiosyncratic variances are distributed uniformly over the interval $[0.1, 0.3]$.[33]

The second data-generating process allows for regime changes in the mean of the single factor. We assume

$$g_t \sim \mathcal{N}\left(\mu_{gt}, \sigma_g^2\right)$$

[33] The parameters used in the i.i.d. simulations are the same as in Garlappi, Uppal, and Wang (2007), except we adapt them to the daily frequency by assuming there are 252 trading days in a year.

and every day μ_{gt} shifts with 50% probability. When there is a regime change:

$$\mu_{gt+1} \sim \mathcal{N}\left[\bar{\mu}_g + \rho\left(\mu_{gt} - \bar{\mu}_g\right), \sigma_{\mu g}^2\right],$$

and when there is no regime change: $\mu_{g+1} = \mu_{gt}$. We let $\rho = 0.95$, $\sigma_{\mu g} = 0.001$, and set

$$\sigma_g = \sqrt{\bar{\sigma}_g^2 - \frac{\sigma_{\mu g}^2}{1-\rho^2}}$$

so that the unconditional variance of the factor g_t is the same as in the i.i.d. case above.[34] We limit the scope of regime changes to the mean of the single factor and assume the rest of the specification and parameters are identical to the i.i.d. case.[35]

For each data-generating process we create 41 datasets. Each dataset has 12,000 observations on daily returns.

Appendix B. Standard Errors of Sharpe Ratios and Certainty Equivalents

To evaluate the performance of portfolio choice rules we need the standard errors of the difference of out-of-sample Sharpe ratios and the difference of out-of-sample certainty equivalents. We use generalized method of moments (GMM) (Hansen [1982]) to derive asymptotic standard errors that are valid under many specifications of disturbances. We choose GMM because we want to make as few assumptions as possible on out-of-sample quantities. Our standard errors do assume that asset returns are stationary so that their unconditional means and variances exist.[36] To compute standard errors, we use standard GMM formulas with analytical derivatives of the relevant moment conditions and the Newey and West (1987) estimate of the spectral density at frequency zero. The following subsections provide the moment conditions for the difference of out-of-sample Sharpe ratios and the difference of out-of-sample certainty equivalents.[37]

B.1 Sharpe ratios

Let R^*_{pt+1} be a benchmark excess portfolio return and R_{pt+1} be the portfolio return using an alternative method. We construct moment conditions so that the Sharpe ratio of the alternative portfolio return is one parameter and the difference of the Sharpe ratios is another parameter. The population moment conditions are

$$E\left[R_{pt+1} - s\sigma\right] = 0, \tag{B1a}$$

$$E\left[\left(R_{pt+1} - s\sigma\right)^2 - \sigma^2\right] = 0, \tag{B1b}$$

$$E\left[R^*_{pt+1} - (s-d)\sigma^*\right] = 0, \tag{B1c}$$

$$E\left[\left(R^*_{pt+1} - (s-d)\sigma^*\right)^2 - \sigma^{*2}\right] = 0, \tag{B1d}$$

where σ is the standard deviation of the alternative portfolio, σ^* is the standard deviation of the benchmark portfolio, s is the Sharpe ratio of the alternative portfolio, and d is the difference of the Sharpe ratios of the two portfolios.

---

[34] We assume the random variables determining regime changes and shifts in μ_{gt} are independent of each other and independent of the random variables determining factor values and asset returns.

[35] We choose the new parameters in the regime-change specification so that simulated autocorrelations of the factor roughly match the actual autocorrelations of daily, monthly, and yearly value-weighted market excess returns.

[36] As discussed earlier, the BA portfolio choice method is designed to work well even when returns are not stationary.

[37] In this appendix we recycle notation. Some of the notation conflicts with definitions in earlier sections.

Let $\bar{E}$ denote the sample mean and $\bar{\sigma}$ denote the sample standard deviation. GMM estimates

$$\hat{\sigma}=\bar{\sigma}\left(R_{pt+1}\right),\ \ \hat{s}=\frac{\bar{E}R_{pt+1}}{\bar{\sigma}\left(R_{pt+1}\right)},\ \ \hat{\sigma}^{*}=\bar{\sigma}\left(R^{*}_{pt+1}\right),\ \ \hat{d}=\frac{\bar{E}R_{pt+1}}{\bar{\sigma}\left(R_{pt+1}\right)}-\frac{\bar{E}R^{*}_{pt+1}}{\bar{\sigma}\left(R^{*}_{pt+1}\right)}$$

of the parameters (σ, s, σ^*, and d) from the sample moments corresponding to the population moments (B1a) through (B1d) are exactly identified because there are four parameters and four moments. The formulas verify that $\hat{s}$ is the sample Sharpe ratio of the portfolio of interest and $\hat{d}$ is the sample difference of the Sharpe ratio of the portfolio of interest and the Sharpe ratio of the benchmark portfolio.

B.2 Certainty equivalents

We imitate the approach in the previous section and construct moment conditions so that the certainty equivalent of the alternative portfolio return is one parameter and the difference of the certainty equivalents is another parameter. The population moment conditions are

$$E\left[R_{pt+1}-\mu\right]=0, \tag{B2a}$$

$$E\left[\mu+R_{ft+1}-\frac{\theta}{2}\left(R_{pt+1}-\mu\right)^{2}-q\right]=0, \tag{B2b}$$

$$E\left[R^{*}_{pt+1}-\mu^{*}\right]=0, \tag{B2c}$$

$$E\left[q-\mu^{*}-R_{ft+1}+\frac{\theta}{2}\left(R^{*}_{pt+1}-\mu^{*}\right)^{2}-\epsilon\right]=0, \tag{B2d}$$

where μ is the mean of the alternative portfolio, μ^* is the mean of the benchmark portfolio, q is the certainty equivalent of the alternative portfolio, and ϵ is the difference of the certainty equivalents of the two portfolios.

Let $\bar{V}$ denote the sample variance. GMM estimates

$$\hat{\mu}=\bar{E}\left(R_{pt+1}\right), \qquad \hat{q}=\bar{E}\left(R_{pt+1}\right)+R_{ft+1}-\frac{\theta}{2}\bar{V}\left(R_{pt+1}\right),$$

$$\hat{\mu}^{*}=\bar{E}\left(R^{*}_{pt+1}\right), \qquad \hat{\epsilon}=\bar{E}\left(R_{pt+1}\right)-\frac{\theta}{2}\bar{V}\left(R_{pt+1}\right)-\bar{E}\left(R^{*}_{pt+1}\right)+\frac{\theta}{2}\bar{V}\left(R^{*}_{pt+1}\right)$$

of the parameters (μ, q, μ^*, and ϵ) from the sample moments corresponding to the population moments (B2a) through (B2d) are exactly identified because there are four parameters and four moments. The formulas verify that the GMM estimate of $\hat{q}$ is the sample certainty equivalent of the alternative portfolio and $\hat{\epsilon}$ is the sample difference of the certainty equivalent of the portfolio of interest and the certainty equivalent of the benchmark portfolio.

Appendix C. Certainty Equivalents for Robust Preferences

Which preferences to use when computing certainty equivalents is partly a philosophical question. On our view, the goal of agents is to maximize their nonrobust preferences. Having a large degree of model uncertainty aversion is interpreted as a device to obtain good out-of-sample returns and maximize Equation (28). Consider three distributions:

1. $f(z_{t+1})$ is the investor's best (*ex-ante*) approximation to the distribution of z_{t+1},
2. $\varrho^{*}(z_{t+1})f(z_{t+1})$ is an alternative (or a constrained worst-case) distribution of z_{t+1} that robust agents worry about, and

3. $q(z_{t+1})$ is the actual (unknown to the investors) distribution of z_{t+1}.

The objective of both robust and nonrobust investors is to maximize the mean-variance criteria where the distribution of z_{t+1} is $q(z_{t+1})$. If the distribution q was known, then

$$f(z_{t+1})=\varrho^*(z_{t+1})f(z_{t+1})=q(z_{t+1})$$

for all z_{t+1} and robust and nonrobust investors would make the same portfolio choices.[38] However, in practice, because neither investor knows the actual distribution q, they use approximations when making portfolio choices. Nonrobust investors use the best approximation from previous data, f. Robust investors use the alternative distribution, $\varrho^* f$. *Ex post*, after observing aspects of the distribution q, all investors form better estimates of the distribution q, and use the estimates and standard mean-variance preferences to evaluate their past portfolio choices. Consequently, we (as econometricians) also use *ex-post* estimates of q and standard mean-variance preferences, to compute out-of-sample certainty equivalents.

There are other plausible interpretations of certainty equivalents that could lead to different techniques for computing certainty equivalents. For example, robust preferences are observationally equivalent to risk-sensitive preferences. If we assume investors have risk-sensitive preferences (see Formula 18, as well as Hansen and Sargent [2007b]) then it is natural to compute certainty equivalents using these preferences.[39]

We follow a similar approach when computing out-of-sample certainty equivalents for robust and nonrobust versions of Jorion's approach and Kan-Zhou's rule, as well as the 1/N portfolio, the market portfolio and the minimum variance portfolio. For each algorithm, we use Formula (28) to compute out-of-sample certainty equivalents.

References

Anderson, E., L. Hansen, and T. Sargent. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1:68–123.

Avramov, D. 2002. Stock Return Predictability and Model Uncertainty. *Journal of Financial Economics* 64: 423–58.

Avramov, D., and G. Zhou. 2010. Bayesian Portfolio Analysis. *Annual Review of Financial Economics* 2: 25–47.

Bishop, C. 2006. *Pattern Recognition and Machine Learning*. New York: Springer.

Brandt, M. 2010. Portfolio Choice Problems. In *Handbook of Financial Econometrics*, vol. 1 269–336. Ed. by Y. Aït-Sahalia and L. Hansen. Elsevier.

Campbell, J., and L. Viceira. 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford: Oxford University Press.

Cremers, K. 2002. Stock Return Predictability: A Bayesian Model Selection Perspective. *Review of Financial Studies* 15:1223–49.

---

[38] Robust agents solve the problem in Formula (15). Because q is the only possible distribution, q must also be the worst-possible distribution and $\varrho^*(z_{t+1})=1$. In this case $\tau=0$ in the robust problem so that the penalty associated with any distribution other than q is infinite. In general, robust agents set τ depending upon their uncertainty. In different environments, the same robust agent can use different values of τ.

[39] One reason for treating risk-sensitive preferences differently from robust preferences is that if we begin by assuming investors have risk-sensitive preferences, then τ is naturally interpreted as a preference parameter and usually assumed to be fixed across environments. With robust preferences, the value of τ depends on the amount of uncertainty in the environment and varies over situations.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies* 22:1915–53.

Dow, J., and S. Werlang. 1992. Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio. *Econometrica* 60:197–204.

Duchin, R., and H. Levy. 2009. Markowitz Versus the Talmudic Portfolio Diversification Strategies. *Journal of Portfolio Management* 35:71–4.

Epstein, L., and M. Schneider. 2003. Recursive Multiple-Priors. *Journal of Economic Theory* 113:1–31.

———. 2010. Ambiguity and Asset Markets. *Annual Review of Financial Economics* 2:315–46.

Gagnon, J., M. Raskin, J. Remache, and B. Sack. 2010. Large-scale Asset Purchases by the Federal Reserve: Did They Work? Federal Reserve Bank of New York Staff Report No. 441.

Garlappi, L., R. Uppal, and T. Wang. 2007. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *Review of Financial Studies* 20:41–81.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2005. There is a Risk-Return Tradeoff After All. *Journal of Financial Economics* 76:509–48.

Hansen, L. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50:1029–54.

Hansen, L., and T. Sargent. 1995. Discounted Linear Exponential Quadratic Gaussian Control. *IEEE Transactions on Automatic Control* 40:968–71.

———. 2007a. Recursive Robust Estimation and Control Without Commitment. *Journal of Economic Theory* 136:1–27.

———. 2007b. *Robustness*. Princeton, NJ: Princeton University Press.

Hansen, L., T. Sargent, and T. Tallarini, Jr. 1999. Robust Permanent Income and Pricing. *Review of Economic Studies* 66:873–907.

Hansen, L., T. Sargent, G. Turmuhambetova, and N. Williams. 2006. Robust Control and Model Misspecification. *Journal of Economic Theory* 128:45–90.

Jorion, P. 1986. Bayes-Stein Estimation for Portfolio Analysis. *Journal of Financial and Quantitative Analysis* 21:279–92.

Kan, R., and G. Zhou. 2007. Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis* 42:621–56.

Kandel, S., and R. Stambaugh. 1996. On the Predictability of Stock Returns: An Asset-Allocation Perspective. *Journal of Finance* 51:385–424.

Klibanoff, P., M. Marinacci, and S. Mukerji. 2005. A Smooth Model of Decision Making under Ambiguity. *Econometrica* 73:1849–92.

MacKinlay, A., and L. Pastor. 2000. Asset Pricing Models: Implications for Expected Returns and Portfolio Selection. *Review of Financial Studies* 13:883–916.

Maenhout, P. 2004. Robust Portfolio Rules and Asset Pricing. *Review of Financial Studies* 17:951–83.

Markowitz, H. 1952. Portfolio Selection. *Journal of Finance* 7:77–91.

Merton, R. 1973. An Intertemporal Capital Asset Pricing Model. *Econometrica* 41:867–87.

Meucci, A. 2005. *Risk and Asset Allocation.* Berlin: Springer Verlag.

Murphy, K. 2012. *Machine Learning: A Probabilistic Perspective.* Cambridge, MA: MIT Press.

Newey, W., and K. West, A Simple. Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–8.

Pastor, L. 2000. Portfolio Selection and Asset Pricing Models. *Journal of Finance* 55:179–223.

Pastor, L., and P. Veronesi. 2009. Learning in Financial Markets. *Annual Review of Financial Economics* 1:361–81.

Samuelson, P. 1970. The Fundamental Approximation Theorem of Portfolio Analysis in terms of Means, Variances and Higher Moments. *Review of Economic Studies* 37:537–42.

Tu, J., and G. Zhou. 2004. Data-generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions? *Journal of Financial Economics* 72:385–421.

———. 2010. Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty. *Journal of Financial and Quantitative Analysis* 45:959–86.

———. 2011. Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies. *Journal of Financial Economics* 99:204–15.

Uppal, R., and T. Wang. 2003. Model Misspecification and Underdiversification. *Journal of Finance* 58:2465–86.

Welch, I., and A. Goyal. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21:1455–508.

Xia, Y. 2001. Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation. *Journal of Finance* 56:205–46.