Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Portfolio Choices with Many Big Models

Evan Anderson, Ai-ru (Meg) Cheng

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–26
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Portfolio Choices with Many Big Models

Evan Anderson,[a] Ai-ru (Meg) Cheng[a]

[a] Department of Economics, Northern Illinois University, DeKalb, Illinois 60115

Contact: ewanderson@niu.edu, https://orcid.org/0000-0002-8541-9120 (EA); acheng1@niu.edu (AC)

Received: January 19, 2019
Revised: July 6, 2020
Accepted: September 24, 2020
Published Online in Articles in Advance: April 30, 2021

https://doi.org/10.1287/mnsc.2020.3876

Abstract. This paper proposes a Bayesian-averaging heterogeneous vector autoregressive portfolio choice strategy with many big models that outperforms existing methods out-of-sample on numerous daily, weekly, and monthly datasets. The strategy assumes that excess returns are approximately determined by a time-varying regression with a large number of explanatory variables that are the sample means of past returns. Investors consider the possibility that every period there is a regime change by keeping track of many models, but doubt that any specification is able to perfectly predict the distribution of future returns, and compute portfolio choices that are robust to model misspecification.

History: Accepted by Tyler Shumway, finance.
Funding: This work was supported by the Center for Robust Decision-Making Tools for Climate and Energy Policy (RDCEP) at the University of Chicago under the National Science Foundation (NSF) Decision Making Under Uncertainty Program [Grant 1463644].
Supplemental Material: The online appendices are available at https://doi.org/10.1287/mnsc.2020.3876.

Keywords: finance • portfolio • investment • economics • econometrics • model uncertainty

We propose a Bayesian-averaging heterogeneous vector autoregressive (BA-HVAR) portfolio choice strategy that assumes that excess returns are approximately determined by a time-varying vector autoregression with a large number of explanatory variables that are the sample means of past returns. The coefficients and covariances in the vector autoregression shift when regimes change and are constant within a regime. Investors consider the possibility that every period there is a regime change by keeping track of many models, where each model makes different assumptions on the date of the most recent change and on hyperparameters. Investors do not know which model is correct and worry that none of the models are correct. Investors apply standard Bayesian statistics to form probabilities over the models and then use robust control methods to compute optimal portfolio choices, which take into account the possibility that all models are misspecified. We consider examples with 1,000 models, where each model has over 12,000 estimated parameters, and we show that BA-HVAR portfolio choices achieve higher certainty equivalents and Shape ratios than many other strategies, out-of-sample on 20 daily, weekly, and monthly data sets.

DeMiguel et al. (2009) and other recent out-of-sample evaluations of portfolio choices show that a $1/N$ strategy, which invests an equal amount in each risky asset, often performs better than other leading strategies. It is well known that most financial economists and many practitioners avoid using sophisticated strategies when making their own portfolio choices. For example, a few years after publishing a paradigm-changing paper (Markowitz 1952) on mean-variance portfolio choices, Harry Markowitz said that he uses a $1/N$ strategy over asset classes when constructing his own portfolios:

> I should have computed the historical co-variances of the asset classes and drawn an efficient frontier. Instead, I visualized my grief if the stock market went way up and I wasn't in it—if it went way down and I was completely in it. My intention was to minimize my future regret. So I split my contributions 50/50 between bonds and equities.[1]

A primary goal of this paper is to provide a sophisticated method with good out-of-sample performance that financial economists and practitioners will want to use.

The poor out-of-sample performance of mean-variance portfolio choices happens mainly because of the difficultly of estimating the covariances and means of asset returns. The distribution of asset returns likely changes over time, and it is challenging to forecast future returns with historical data when regime changes are frequent. Modern refinements, such as Kan and Zhou (2007)'s three-fund rule, greatly improve upon standard mean-variance allocations by mitigating the consequences of estimation error, but these methods sometimes still struggle to reliably achieve good out-of-sample performance.

Recent innovations in portfolio choice theory usually emphasize methods with a limited number of parameters.[2] One reason for this is that, although more complex methods may appear to perform much better in-sample, it is feared that they will overfit in-sample data and lead to portfolio choices that perform worse than parsimonious methods out-of-sample. A key contribution of this paper is to document that this intuition is not always correct and that an approach with an extremely large number of models and parameters can achieve excellent out-of-sample performance. In some of our applications of BA-HVAR, we estimate 8,646,000 parameters that determine covariances and 3,406,000 coefficients that determine means, across 1,000 possible models. We avoid the poor out-of-sample performance of other strategies by placing informative priors on parameters that prevent unreasonable estimates. The large number of parameters gives BA-HVAR ample capability to adapt to regime changes and time-varying relationships among economic variables.

BA-HVAR uses explanatory variables that are inspired by Corsi (2009)'s heterogeneous autoregressive (HAR) model. Corsi's model forecasts volatility with realized volatility over the past day, week, and month. According to Bollerslev et al. (2016), Corsi's model has become a dominant specification for volatility forecasting. BA-HVAR applies the idea behind Corsi's approach to forecast returns (rather than volatility) with past realized returns, and uses returns from the previous day, week, month, year, and 10 years, as explanatory variables. We document that optimal robust portfolios that use the HAR forecasts of returns (as a reference specification) perform much better out-of-sample than alternative approaches.

The foundations for the BA-HVAR portfolio choice strategy were developed by Zellner and Chetty (1965), who began the literature on optimal Bayesian portfolio choices by studying a simple optimal investment problem with returns determined by a vector autoregression, and DeMiguel et al. (2014), who recently have shown that portfolio choices based on a first-order vector autoregressive process for returns perform well. BA-HVAR is also related to Avramov (2002), who argued that Bayesian model averaging is a good way of combining models, and Frost and Savarino (1986), who studied Bayesian portfolio choices with an informative prior that all securities have the same mean and variance and all pairs of securities have the same covariance. For surveys of Bayesian approaches in financial economics, see Bawa et al. (1979) and Avramov and Zhou (2010). Also see Pastor and Veronesi (2009) for a survey of related learning models.[3]

The BA-HVAR portfolio choice strategy builds on the recent portfolio choice strategy proposed by Anderson and Cheng (2016), who show that a Bayesian-averaging portfolio choice strategy can achieve good out-of-sample performance. Anderson and Cheng (2016)'s approach uses a large number of simple models, where each model assumes that the means and covariances of stock returns are constant over time, and models differ on the amount of data used to approximate means and covariances. Our paper is more general than Anderson and Cheng (2016). We allow each model to have a time-varying mean that depends on past returns, include more than one new model each period, use a more natural prior for the probability of regime changes that is grounded in past data, alter the timing of the creation of new models to improve forecasts of asset returns after a regime change, and provide a mechanism for removing poorly performing models. We show that these extensions lead to major performance improvements over Anderson and Cheng (2016), because it gives models more flexibility to adapt to data. We show that the method of Anderson and Cheng (2016) detects an extremely large number of regimes in data, because each of its models cannot adapt to changing means, whereas BA-HVAR finds far fewer regimes in data because its models can better adapt to changing environments.

Despite considering a large number of models, we assume investors are worried that none of the models are correct and compute robust portfolio choices. The robust portfolio choices follow the approach used by Uppal and Wang (2003), Maenhout (2004, 2006), and Anderson and Cheng (2016), who apply the general robust control methodology of Hansen and Sargent (2007b) by computing decisions with good performance on alternative specifications that are near a benchmark (or reference) specification. There are many other related ways for computing robust portfolio choices, such as the methods used by Dow and Werlang (1992), Garlappi et al. (2007), Epstein and Schneider (2010), and Liu (2010).

1. Overview

The BA-HVAR approach assumes a good approximation of the world is that excess returns depend linearly on past returns with time-varying coefficients so that

$$\mathbf{r}_t = B_t'\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t,$$

where $\mathbf{r}_t$ is a vector of excess returns, B_t is a time-varying matrix of coefficients, $\mathbf{x}_{t-1}$ is a vector of explanatory variables, and $\boldsymbol{\epsilon}_t$ is a multivariate normal random shock with a time-varying variance-covariance matrix, Σ_t. The explanatory variables are inspired by HAR and, in some of our examples, include average excess returns over the past day, week, month, year, and 10 years for all stocks (as well as a constant).

At each date, a regime change is possible that shifts the coefficients (B_t) and the covariances (Σ_t). When there is a regime change, new values of the coefficients and covariances are drawn from a normal-inverse Wishart distribution. In the absence of a regime change, the coefficients and covariances have the same values as in the previous period. Unlike in regime-switching models, such as Tu (2010), regimes in the BA-HVAR approach are not recurrent. Once regimes change, asset returns follow a new process that is different from previous processes. Investors do not know when regimes change and are not able to directly observe the values of the coefficients and covariances within a regime.

Investors learn about regime changes, coefficients, and covariances from past asset returns by keeping track of many models, where each model is a Bayesian heterogeneous vector autoregression with constant coefficients and covariances, and models differ on when the last regime change occurred and on their hyperparameters. Every period, many new models are born that assume a regime change occurs next period, and, for computational tractability, poorly performing older models are removed (after a burn-in period). Investors use Bayesian methods to keep track of the probability that each model is right, under the assumption that one of the models is right, and compute optimal robust mean-variance portfolio choices that take into account the possibility that all of the models are wrong.

We show that optimal robust BA-HVAR portfolios achieve higher certainty equivalents and Sharpe ratios out-of-sample than many other leading strategies (including the 1/N, Jorion, Kan-Zhou, and market strategies) on 20 daily, weekly, and monthly data sets from the Center for Research in Security Prices (CRSP) and from Ken French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/).

2. An Approximation of the World

We begin the description of the technical details of the BA-HVAR method by formally describing the approximation to reality that it uses in later sections to estimate parameters and determine model probabilities.[4] Let $\mathbf{r}_t$ be an $n \times 1$ vector of excess returns on risky stocks between time $t-1$ and time t; let $\mathbf{x}_{t-1}$ be a $k \times 1$ vector of explanatory variables, known at time $t-1$, that are believed to determine the conditional expectation of $\mathbf{r}_t$; let B_t be a $k \times n$ matrix of coefficients that weight $\mathbf{x}_{t-1}$; and let Σ_t be a $n \times n$ variance-covariance matrix. Let $r_{f,t}$ be the nominal risk-free rate between times $t-1$ and t whose value is known at time $t-1$. The excess returns $\mathbf{r}_t$ are constructed by subtracting the nominal risk-free rate from the nominal return on the risky assets.

The approximating framework assumes that excess returns are conditionally normal.

Assumption 1. *The vector $\mathbf{r}_t$ is conditionally normal with mean $B_t'\mathbf{x}_{t-1}$ and variance-covariance matrix Σ_t:*

$$\mathbf{r}_t | \mathbf{x}_{t-1}, B_t, \Sigma_t \sim \mathcal{N}\left(B_t'\mathbf{x}_{t-1}, \Sigma_t\right),$$

where at time $t-1$, the value of $\mathbf{x}_{t-1}$ is known but the values of B_t and Σ_t are not observed.

We let $\mathbf{x}_{t-1}$ consist of a constant and the sample means of past excess returns for each of the n risky stocks measured using $(\ell-1)$ different windows. The first element of $\mathbf{x}_{t-1}$ is the constant, and the other elements are the sample means of excess returns between periods $t-d_i$ and $t-1$ for $i = 2, 3, \ldots \ell$, where $d_2, d_3, \ldots d_\ell$ is an increasing sequence of numbers. More formally, we have the following.

Assumption 2. *The $k \times 1$ vector of explanatory variables is*

$$\mathbf{x}_{t-1} = \left[1 \quad \overline{E}_{s=t-d_2}^{t-1}(\mathbf{r}_s)' \quad \overline{E}_{s=t-d_3}^{t-1}(\mathbf{r}_s)' \quad \ldots \quad \overline{E}_{s=t-d_\ell}^{t-1}(\mathbf{r}_s)'\right]',$$

where the $n \times 1$ vector

$$\overline{E}_{s=t-a}^{t-1}(\mathbf{r}_s) = \frac{1}{a}\sum_{s=t-a}^{t-1}\mathbf{r}_s$$

is the sample mean of $\mathbf{r}_s$ between times $t-a$ and $t-1$.

The number of elements in the vector $\mathbf{x}_{t-1}$ is $k = 1 + n(\ell-1)$, which can be large. For example, in some of our applications, $n = 26$ and $\ell = 6$, so that $k = 131$. In this case, there are $kn = 3{,}406$ coefficients in the matrix B_t since the lags of all assets are used as explanatory variables for each asset.

The approximating framework assumes that a regime change is possible each period. The coefficients and covariances are constant within a regime but jump and are drawn from a normal-inverse Wishart distribution when regimes change.

Assumption 3. *With probability π_t, there is a regime change between periods $t-1$ and t, and the time t coefficients and covariances are drawn from the normal inverse-Wishart distribution:*

$$\Sigma_t \sim \mathcal{IW}(\Lambda, \nu), \tag{1a}$$

$$vec(B_t)|\Sigma_t \sim \mathcal{N}\left(vec(\overline{B}), \Sigma_t \otimes \Phi\right). \tag{1b}$$

With probability $(1-\pi_t)$, there is not a regime change and the coefficients and covariances do not change from the previous period:

$$B_t = B_{t-1} \quad and \quad \Sigma_t = \Sigma_{t-1}.$$

The notation $\mathcal{IW}(\Lambda, \nu)$ denotes the inverse-Wishart distribution with the $n \times n$ positive-definite scale matrix Λ and (scalar) degrees of freedom ν. Investors do not directly observe regime changes and also do not directly observe the draws of the coefficients and covariances. The $k \times n$ matrix $\overline{B}$ is the mean of the B_t draw. The $k \times k$ matrix Φ along with the Σ_t draw determine the variance-covariance matrix of the draw of $\text{vec}(B_t)$. The mean of the Σ_t draw is

$$\frac{\Lambda}{\nu - n - 1} \tag{2}$$

and is defined when $\nu > n + 1$. The size of the windows in $\mathbf{x_{t-1}}$ is fixed across regimes. All regimes use the same number of explanatory variables and the same values for $d_2, d_3, \ldots d_\ell$.

We set the value of ν so that the covariances can take on a wide range of values when regimes change. We assume that Φ is a constant matrix, but we do not assume its value is known.

Assumption 4. *We set $\nu = n + 2$ and Φ to be a constant matrix.*

We set $\nu = n + 2$ because this is the smallest integer value of ν for which the mean of the Σ_t draw is defined. In later sections, we propose a method for estimating the values of $\overline{B}$, Λ, and Φ.

3. A Typical Model

This section describes the inner workings of one of the many possible vector heterogeneous autoregression models of returns used by BA-HVAR, and the next section discusses the collection of all available models.

Investors have many different models of returns, where each model makes a different assumption about the date of the last regime change and/or a different assumption about hyperparameters. Investors do not know which, if any, of the models is correct and are worried that none of the models are correct. In this section, we describe a typical model that is a Bayesian vector autoregression that uses explanatory variables inspired by Corsi (2009)'s HAR model. For more background and details about Bayesian vector autoregressions, see Del Negro and Schorfheide (2011), Sims and Zha (1998), and Litterman (1986).

Let m indicate a particular model of returns that assumes that there is a regime change at time $c = c(m)$ and that the hyperparameters are $i = i(m)$. We say that model m is *born* at time $(c - 1)$, and that before time $(c - 1)$, model m did not exist. Model m is born before time c information is observed, is first used at time $c - 1$, and assumes that excess returns are conditionally normal with the same coefficients and covariances at all dates on and after date c. Model m believes that between time $(c - 1)$ and time c there is a regime change and that, starting at date c, the world is different than it was in the past.

Assumption 5. *In model m, $\mathbf{r_t}$ is conditionally normal with mean $B_c{}'\mathbf{x_{t-1}}$ and covariance matrix Σ_c:*

$$\mathbf{r_t} | \mathbf{x_{t-1}}, B_c, \Sigma_c \sim \mathcal{N}(B_c{}'\mathbf{x_{t-1}}, \Sigma_c)$$

for $t \geq c$.

In other words, model m takes the position that there are no further regime changes after date c so that the covariances $\Sigma_t = \Sigma_c$ and the coefficients $B_t = B_c$ for all $t \geq c$. In model m, the values of the coefficients and covariances are time-invariant, but their actual values are not directly revealed to investors. Section 3.1 specifies the initial beliefs of model m about the coefficients and covariances, and Section 3.2 shows how those beliefs are updated over time.

3.1. Initial Beliefs

We specify the initial beliefs of model m, which are the beliefs of model m about the time $c = c(m)$ values of the covariance of returns and the coefficients, formed at time $(c - 1)$, before time c information is observed. Model m assumes that, between times $c - 1$ and c, there is a regime change and, as described in Section 2, the coefficients and covariances evolve. Following Assumptions 3 and 4, model m believes that the unobserved constants Σ_c and $vec(B_c)$ are drawn from a normal-inverse Wishart distribution:

$$\Sigma_c \sim \mathcal{IW}(\Lambda, \nu), \tag{3a}$$

$$\text{vec}(\text{B}_c) \,|\, \Sigma_c \sim \mathcal{N}\big(\text{vec}(\overline{B}), \Sigma_c \otimes \Phi\big), \tag{3b}$$

where $\nu = n + 2$ and Φ is a constant matrix. The values of Λ, $\overline{B}$, and Φ are not known. We now describe the estimates used by model m.

Model m estimates Λ and $\overline{B}$ by assuming that, after a regime change, there is no *initial* information about how stocks will be different from each other, so that the mean of each stock's excess return is drawn from the same distribution, and the variance of each stock's return is also drawn from the same distribution. Model m also assumes that the mean draw of covariances between any pair of stocks is zero and that the mean dependence on all lag returns is zero. However, the actual draws of means, covariances, and lag dependencies will typically vary across stocks. Since it is difficult to determine a single estimate of Φ, we assume that model m's estimate of Φ is determined by its hyperparameters. We now describe each of the estimates in more detail.

Model m estimates $\overline{B}$ by assuming that the mean of the *draw* of the coefficient on the constant for each stock is the common past sample mean of all stocks and that the means of the draws of the other coefficients are zero.

Assumption 6. *Model m estimates $\overline{B}$ with*

$$\overline{B}_{m,c-1} = \begin{bmatrix} \bar{\mu}_{c-1}\mathbf{1}' \\ O \end{bmatrix}, \tag{4}$$

where O is a $(k-1)\times n$ matrix of zeros and $\bar{\mu}_{c-1}\mathbf{1}'$ is an n-dimensional row vector with all elements equal to the common past mean of all stocks:

$$\bar{\mu}_{c-1} = \frac{1}{n(c-a)}\sum_{j=1}^{n}\sum_{s=a}^{c-1}\mathbf{r_s}(j).$$

The notation $\mathbf{r_s}(j)$ represents the excess return on stock j between dates $s-1$ and s, and the constant a is the initial date that data are available.

Model m estimates Λ by assuming that the mean of the *draw* of the variance for each stock is the common past sample variance of all stocks and that the mean of the draw of the covariances, between any pair of stocks, are zero.

Assumption 7. *Model m estimates Λ with*

$$\Lambda_{m,c-1} = \bar{\sigma}^2_{c-1} I_n, \tag{5}$$

where $\bar{\sigma}^2_{c-1}I_n$ is a diagonal matrix with all diagonal elements equal to the common past variance of all stocks,

$$\bar{\sigma}^2_{c-1} = \frac{1}{n(c-a)-1}\sum_{j=1}^{n}\sum_{s=a}^{c-1}\left[\mathbf{r_s}(j) - \bar{\mu}_{c-1}\right]^2,$$

and the notation I_n represents an $n\times n$ identity matrix.

These are not the only possible estimates of $\overline{B}$ and Λ, but, since regime changes are likely infrequent, it is difficult to form better estimates than $\overline{B}_{m,c-1}$ and $\Lambda_{m,c-1}$.

Model m estimates Φ by assuming that it is a diagonal matrix that is determined by two scalar hyperparameters. The hyperparameters vary across models born at time $(c-1)$. Since it is likely that the variance of the constant coefficients in B_c is different from the variance of the other terms, we let the $(1,1)$ element of Φ be determined by one hyperparameter and the rest of the diagonal elements of Φ be determined by the other hyperparameter.

Assumption 8. *Model m estimates Φ with the diagonal matrix*

$$\Phi_{m,c-1} = \begin{bmatrix} \alpha_i & \mathbf{0}' \\ \mathbf{0} & \beta_i I_{k-1} \end{bmatrix}, \tag{6}$$

where $i = i(m)$ is the setting of model m's hyperparameters, $\beta_i I_{k-1}$ is a diagonal matrix with all diagonal elements equal to β_i, and $\mathbf{0}$ is a column vector of zeros. We let $\alpha_i \in \{\frac{1}{100}, \frac{1}{10}, 1\}$ and $\beta_i \in \{1, 10, 100\}$.

We allow the possible values for β_i to be much larger than the values of α_i because we have very little information about the coefficients that multiply previous returns, whereas it is likely that we have much more information about the first coefficient (which is the constant term).[5]

Model m's beliefs about the coefficients and covariances, at time $c-1$, are obtained by plugging the estimates provided in this section, into the normal-inverse Wishart distribution specified in Equations (3a) and (3b).

Model *m*'s Initial Beliefs. *At time $c-1$, model m believes that B_c and Σ_c have a normal-inverse Wishart distribution:*

$$P(\Sigma_c|\mathcal{F}_{c-1}) = \mathcal{IW}(\Lambda_{m,c-1}, \nu_{m,c-1}), \tag{7a}$$

$$P[vec(B_c)|\mathcal{F}_{c-1}, \Sigma_c] = \mathcal{N}\left(vec(\overline{B}_{m,c-1}), \Sigma_c \otimes \Phi_{m,c-1}\right), \tag{7b}$$

where $\nu_{m,c-1} = \nu = n+2$. Here $\Lambda_{m,c-1}$, $\overline{B}_{m,c-1}$, and $\Phi_{m,c-1}$ are given by Equations (4), (5), and (6).

The notation $\mathcal{F}_{c-1}$ denotes the information available at time $c-1$, which consists of historical returns up until time $(c-1)$ and thus includes $\mathbf{x_{c-1}}$. The notation $P(\Sigma_c|\mathcal{F}_{c-1})$ denotes the probability density function (pdf) of Σ_c given information at time $c-1$, and the notation $P[\text{vec}(B_c)|\mathcal{F}_{c-1}, \Sigma_c]$ denotes the pdf of $vec(B_c)$ given information at time $c-1$ and Σ_c. Since $\nu_{m,c-1} = n+2$, the value of $\Lambda_{m,c-1}$ is the belief of model m about the mean of Σ_c [see Equation (2)].

Our specification of initial beliefs builds upon Frost and Savarino (1986) and Anderson and Cheng (2016). In an application with a single model, Frost and Savarino set the prior means and variances of asset returns to be the same for all assets, and they set all pairwise prior covariances to be the same. We follow Anderson and Cheng (2016) by assuming that all pairwise prior covariances are zero, and we extend their approach by assuming that the prior values of the coefficients on past returns are zero.

This section described the initial beliefs about the draws of the coefficients and covariances. Over time, as the next section describes, investors are able to form more precise estimates of the actual draws by observing realized returns.

3.2. Updating Parameters

At each date, investors update their beliefs about the coefficients and covariances after excess returns are observed. We assume that this is the only new information available to investors and that investors use Bayesian methods to update beliefs. The beliefs about the coefficients and covariances in model m are normal-inverse Wishart at every date, since the normal-inverse Wishart distribution is a conjugate distribution for the likelihood function of excess returns, and, as discussed in Section 3.1, the initial beliefs are normal-inverse Wishart.

Parameters are updated as follows.

Parameter Updating. *Before observing time t information, investors' prior beliefs about the constants B_c and Σ_c in model m are normal-inverse Wishart:*

$$P(\Sigma_c|\mathcal{F}_{t-1}) = \mathcal{IW}(\Lambda_{m,t-1}, \nu_{m,t-1}), \tag{8a}$$

$$P[vec(B_c)|\mathcal{F}_{t-1}, \Sigma_c] = \mathcal{N}\left(vec(\overline{B}_{m,t-1}), \Sigma_c \otimes \Phi_{m,t-1}\right), \tag{8b}$$

where $c = c(m)$, $\Lambda_{m,t-1}$ and $\Phi_{m,t-1}$ are positive-definite matrices, $\overline{B}_{m,t-1}$ is a matrix, and $\nu_{m,t-1}$ is a scalar. After observing $\mathbf{r_t}$ investors' posterior beliefs about the constants B_c and Σ_c in model m are also normal-inverse Wishart:

$$P(\Sigma_c|\mathcal{F}_t) = \mathcal{IW}(\Lambda_{m,t}, \nu_{m,t}), \tag{9a}$$

$$P[vec(B_c)|\mathcal{F}_t, \Sigma_c] = \mathcal{N}\left(vec(\overline{B}_{m,t}), \Sigma_c \otimes \Phi_{m,t}\right), \tag{9b}$$

where

$$\Lambda_{m,t} = \Lambda_{m,t-1} + \left(\frac{1}{1 + \mathbf{x}'_{t-1}\Phi_{m,t-1}\mathbf{x}_{t-1}}\right)\boldsymbol{e}_{m,t}\boldsymbol{e}'_{m,t}, \tag{10a}$$

$$\nu_{m,t} = \nu_{m,t-1} + 1, \tag{10b}$$

$$\Phi_{m,t} = \left(\Phi_{m,t-1}\mathbf{x}_{t-1}\mathbf{x}'_{t-1} + I_k\right)^{-1}\Phi_{m,t-1}, \tag{10c}$$

$$\overline{B}_{m,t} = \Phi_{m,t}\mathbf{x}_{t-1}\mathbf{r}'_t + \left(\Phi_{m,t-1}\mathbf{x}_{t-1}\mathbf{x}'_{t-1} + I_k\right)^{-1}\overline{B}_{m,t-1}, \tag{10d}$$

and where we define

$$\boldsymbol{e}_{m,t} = \mathbf{r_t} - \overline{B}'_{m,t-1}\mathbf{x_{t-1}}$$

to be the prediction error.

These updating formulas follow from Bayes's rule and were derived by Tiao and Zellner (1964). The posterior distribution of the coefficients and covariances at time t becomes the prior distribution for time $t+1$. So, at each date in the future, the prior and posterior will be normal-inverse Wishart, since the initial prior distribution is normal-inverse Wishart.

4. All Models

There are many possible models in the BA-HVAR approach. Each model is a Bayesian heterogeneous vector autoregression, as described in Section 3, but makes a different assumption about the date of the most recent regime change and/or a different assumption about hyperparameters. At each date, we assume that there are h models born, where h is the number of possible settings of hyperparameters.[6] At time $t-1$, for predicting time t returns, there are $(h \cdot t)$ possible models since time t returns could be part of a regime that started at any date between 1 and t and could have any one of h settings of hyperparameters.

Model m assumes that there is a new regime at time $c = c(m)$ and no subsequent regime changes, so that the most recent regime change occurred between times $(c-1)$ and c. Model m uses the hyperparameter setting $i = i(m)$, where $i \in \{1, 2, 3, \ldots, h\}$. We assume that there always is a new regime at time 1, but there may or may not be other regime changes. We let $\mathcal{M}_t$ be the set of available models, at time $t-1$, for predicting time t returns.

We keep track of the time-varying probability that each model is correct, under the assumption that one of the models is correct. We define $P_s(m|\mathcal{F}_{t-1})$ to be the probability of model m using time $t-1$ information, when the set of available models is $\mathcal{M}_s$. Section 4.1 describes how probabilities are updated when new information arrives. At each date, there are additional models that can lead to an explosion of models, and so in Section 4.2, we discuss how to remove poorly performing models, in order to limit the number of models. Section 4.3 discusses how probabilities are assigned to new models.

4.1. Updating Model Probabilities

At each date, model probabilities are updated after excess returns are observed. At time t, after observing excess returns $\mathbf{r_t}$ the probability of each model is updated from $P_t(m|\mathcal{F}_{t-1})$ to $P_t(m|\mathcal{F}_t)$.

Probability Updating. *After observing $\mathbf{r_t}$, the probability of each model $m \in \mathcal{M}_t$ is updated to*

$$P_t(m|\mathcal{F}_t) = \frac{P(\mathbf{r_t}|m, \mathcal{F}_{t-1})P_t(m|\mathcal{F}_{t-1})}{\sum_{m\in\mathcal{M}_t} P(\mathbf{r_t}|m, \mathcal{F}_{t-1})P_t(m|\mathcal{F}_{t-1})}, \tag{11a}$$

where

$$P(\mathbf{r_t}|m, \mathcal{F}_{t-1}) = \frac{\Gamma_n(\nu_{m,t}/2)|\Lambda_{m,t-1}|^{\frac{\nu_{m,t-1}}{2}}}{\Gamma_n(\nu_{m,t-1}/2)|\Lambda_{m,t}|^{\frac{\nu_{m,t}}{2}}\left[\pi(1 + \mathbf{x}'_{t-1}\Phi_{m,t-1}\mathbf{x}_{t-1})\right]^{\frac{n}{2}}} \tag{11b}$$

is the likelihood of observing $\mathbf{r_t}$ in model m.

The likelihood function in Equation (11b) follows from Bayes's rule and is described in Zellner (1971). The notation Γ_n denotes the n-dimensional multivariate gamma function.

4.2. Removing Poorly Performing Models

Every date, after a burn-in period, we remove poorly performing models in order to reduce the computational burden of keeping track of an extremely large collection of models. If we did keep track of all conceivable models, then there would be an explosion of models, since on every date many new models are born. For example, after 24,000 daily observations, there would be 216,000 models, when $h = 9$. It is computationally overwhelming to keep track of so many models (using an ordinary workstation). Thus, for tractability, we keep track of only $M = 1{,}000$ models, by removing poorly performing models. This subsection describes how we remove models, and the next subsection discusses how we add new models.

At each date, after a burn-in period, we permanently remove the h models with the lowest probability, which are at least $M/(2h)$ periods old. Young models [models that have existed less than $M/(2h)$ periods] are not fully developed, and we keep them around until they are mature, regardless of their performance. We let $\mathcal{P}_t$ be the subset of models, which includes only the poorly performing models that will be removed.[7]

We renormalize the model probabilities, $m \in \mathcal{M}_t$:

$$\hat{P}_t(m|\mathcal{F}_t) = \begin{cases} 0 & \text{if } m \in \mathcal{P}_t \\ \left(\frac{1}{s_t}\right) P_t(m|\mathcal{F}_t) & \text{if } m \notin \mathcal{P}_t, \end{cases}$$

where

$$s_t = \sum_{m \in \mathcal{M}_t \text{ such that } m \notin \mathcal{P}_t} P_t(m|\mathcal{F}_t),$$

so that the sum of the probabilities of surviving models is one. Here $\hat{P}_t(m|\mathcal{F}_t)$ is the probability of model m *after* poorly performing models have been removed. The removed models are not included in $\mathcal{M}_{t+1}$, which is the set of available models next period.

4.3. Regime Changes and the Probability of New Models

Every period, many new models are born, and we set the total probability of all the new models equal to the probability of a regime change and discount the probabilities of older models so that the probabilities over all models sum to one. The probability of a particular new model depends on the probability of a regime change and the past performance of models with the same hyperparameters. We assign more probability to new models with hyperparameters that have worked well in the past, which allows investors to learn the best hyperparameter setting.

At time t, h models are born and included in the set of available models, $\mathcal{M}_{t+1}$, that are used to forecast time $t+1$ returns. Let

$$\omega_{i,t} = \sum_{m \in \mathcal{M}_t \text{ such that } i(m)=i} \hat{P}_t(m|\mathcal{F}_t)$$

be the sum of the current probabilities of all models of type i born in previous periods, and (as in earlier sections) let π_{t+1} be the probability of a regime change between times t and $t+1$. If we are completely sure that the best setting of hyperparameters is constant over time, then it would be reasonable to set the probability of a new model, of type i, to $\pi_{t+1}\,\omega_{i,t}$, since exactly one model of type i is born every period. However, we are not sure that the best setting is constant over time. In order to guard against the possibility that the best setting is time-varying, we set the probability of a new model, of type i, to $\pi_{t+1}\,\bar{\omega}_{i,t}$, where

$$\bar{\omega}_{i,t} = \frac{\bar{\omega}}{h} + (1-\bar{\omega})\omega_{i,t}$$

and where $\bar{\omega} = 0.10$. This guarantees that each new model has at least probability $(\pi_{t+1}\bar{\omega})/h$, regardless of the past performance of models of the same type. If the best setting is time-varying, then this allows investors to quickly adapt to the current best setting. The probabilities of models $m \in \mathcal{M}_{t+1}$, after the new models are born, become

$$P_{t+1}(m|\mathcal{F}_t) = \begin{cases} \pi_{t+1}\,\bar{\omega}_{i,t} & \text{if } c(m) = t+1 \text{ and } i(m) = i \\ (1-\pi_{t+1})\hat{P}_t(m|\mathcal{F}_t) & \text{if } c(m) < t+1 \end{cases}$$

when $t \geq 1$. At date one, we assume that there is a regime change and that all h initial models have the same probability so that $P_1(m|\mathcal{F}_0) = 1/h$ for $m \in \mathcal{M}_1$.

In the rest of this section, we describe an empirical prior for approximating the prior probability of a regime change. The empirical prior uses the probability-weighted inverse-age of existing models to approximate the probability of a regime change:

$$\pi_{t+1} = \sum_{m \in \mathcal{M}_t} \left(\frac{1}{t - c(m) + 2}\right) \hat{P}_t(m|\mathcal{F}_t), \qquad t \geq 1, \tag{12}$$

where $t - c(m) + 2$ is the age of model m at time $t+1$. We call this the *empirical prior* because it is determined by existing model probabilities and does not have any free parameters.

To motivate the empirical prior, consider the following example. Suppose that some model, model m^*, has probability one at time t and all other models have probability zero. We suggest that a reasonable setting for the probability of a regime change between times t and $t+1$ is

$$\frac{1}{t - c(m^*) + 2}. \tag{13}$$

This sets the probability of a regime change equal to the average probability of all models, born on or after time $c(m^*)$, that have ever existed, including models that have been removed and no longer exist. If $c(m^*) = 1$, so that one of the oldest models, born at time zero, has probability one, then this prior gives probability $1/(t+1)$ to the new models. If $c(m^*) = t$, so that one of the models born at time $t-1$ has probability one, then this prior gives probability $1/2$ to the new models. More generally, Equation (12) is a probability-weighted average of Equation (13) and π_{t+1} will always be between $1/(t+1)$ and $1/2$.

The applications of BA-HVAR, in later sections, use the empirical prior. The empirical prior is much different from the sharing prior used in Anderson and Cheng (2016). The sharing prior transfers a large amount of probability to younger models, because, at the beginning of each period, every older model shares its previous probability equally with *all* younger models. In practice, the empirical prior transfers much less probability and only newly born models receive the transfers.

5. Predicted Returns

We discuss BA-HVAR's predictive distribution, as well as a related reference distribution that we use in subsequent sections to compute portfolio choices.

5.1. Predicted Distribution

The predicted return distribution for a *single* Bayesian vector autoregressive model, model m, is well known and is multivariate t:[8]

$$P(\mathbf{r}_{t+1}|m,\mathcal{F}_t) = \mathcal{T}\left[\boldsymbol{\mu}_{t+1|m,t}, \left(\frac{\nu_{m,t}-n-1}{\nu_{m,t}-n+1}\right) \times V_{t+1|m,t}, \nu_{m,t}-n+1\right] \tag{14a}$$

with means and covariances,

$$\boldsymbol{\mu}_{t+1|m,t} = \overline{B}_{m,t}{}'\mathbf{x}_t, \quad V_{t+1|m,t} = (1+\mathbf{x}_t{}'\Phi_{m,t}\mathbf{x}_t)\,\Sigma_{t+1|m,t} \tag{14b}$$

where

$$\Sigma_{t+1|m,t} = \left(\frac{1}{\nu_{m,t}-n-1}\right)\Lambda_{m,t} \tag{14c}$$

is model m's mean estimate of Σ_c. The correction term $(\mathbf{x}_t'\Phi_{m,t}\mathbf{x}_t)\Sigma_{t+1|m,t}$ to the covariance of returns results from uncertainty in estimating the value of B_c.

The predicted distribution of asset returns for BA-HVAR is a mixture of multivariate t distributions:

$$P(\mathbf{r}_{t+1}|\mathcal{F}_t) = \sum_{m\in\mathcal{M}_{t+1}} P(\mathbf{r}_{t+1}|m,\mathcal{F}_t)P_{t+1}(m|\mathcal{F}_t),$$

where $P_{t+1}(m|\mathcal{F}_t)$ is the probability of model m at time t.

5.2. Reference Distribution

In the rest of this paper, we assume that BA-HVAR investors' best available description of the world (or reference distribution) ignores parameter uncertainty in coefficients and covariances. In this reference distribution, the conditional distribution of returns is normal (rather than multivariate t) in each model and a mixture of multivariate normals (rather than a mixture of multivariate t's) across all models. More formally, the reference distribution is as follows.

Reference Distribution. *The conditional distribution of $\mathbf{r}_{t+1}$ in model m is*

$$\bar{P}(\mathbf{r}_{t+1}|m,\mathcal{F}_t) = \mathcal{N}\left(\boldsymbol{\mu}_{t+1|m,t}, \Sigma_{t+1|m,t}\right) \tag{15a}$$

and the conditional distribution of $\mathbf{r}_{t+1}$ across all models is

$$\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t) = \sum_{m\in\mathcal{M}_{t+1}} \mathcal{N}\left(\boldsymbol{\mu}_{t+1|m,t}, \Sigma_{t+1|m,t}\right) \times P_{t+1}(m|\mathcal{F}_t) \tag{15b}$$

where $\boldsymbol{\mu}_{t+1|m,t}$ and $\Sigma_{t+1|m,t}$ are defined in Equations (14b) and (14c). The conditional means and covariances of returns are

$$\boldsymbol{\mu}_{t+1|t} = \sum_{m\in\mathcal{M}_{t+1}} \boldsymbol{\mu}_{t+1|m,t}P_{t+1}(m|\mathcal{F}_t), \tag{15c}$$

$$\Sigma_{t+1|t} = \sum_{m\in\mathcal{M}_{t+1}} \left(\Sigma_{t+1|m,t} + \boldsymbol{\mu}_{t+1|m,t}\boldsymbol{\mu}'_{t+1|m,t}\right) \times P_{t+1}(m|\mathcal{F}_t) - \boldsymbol{\mu}_{t+1|t}\boldsymbol{\mu}'_{t+1|t}. \tag{15d}$$

The reference distribution ignores estimation errors in coefficients and covariances (within a model) but does fully account for the uncertainty in coefficients and covariances *across* models. As described in previous sections, we allow estimates to evolve over time as new information arrives. Parameter uncertainty is ignored only for forming predictions that are used to make portfolio choices.

We ignore parameter uncertainty for several reasons. First, although for analytical convenience we assumed a normal-inverse Wishart distribution when estimating parameters, investors may not fully trust all of its implications, especially its implications for tail behavior. Second, for robust BA-HVAR investors, there are no interesting solutions to the robust mean-variance problem (described in later sections) when they believe returns follow a multivariate t or a mixture of multivariate t distributions because, as discussed in Anderson and Cheng (2016), the optimal solution to the robust mean-variance problem, in this case, is to *never* invest in any risky asset. Third, the robust mean-variance problem already takes into account uncertainty in the entire distribution of returns, so that it is not necessary to also include parameter uncertainty. Fourth, multiperiod-ahead return forecasts are very time-consuming to compute in the presence of parameter uncertainty, but easier to compute without parameter uncertainty.[9] Fifth, although it is reasonable to include parameter uncertainty in coefficients (but not covariances) when forecasting one-period-ahead returns, the resulting portfolios are almost identical to the portfolios assuming no parameter uncertainty.[10]

6. Portfolio Choices

We discuss nonrobust and robust mean-variance rules of allocating investments among a risk-free asset and many risky assets, which BA-HVAR investors use to select a vector of portfolio weights that gives the fraction of wealth invested in each risky asset. As in earlier sections, the (nominal) risk-free rate between times t and $t+1$ is denoted by $r_{f,t+1}$ and its value is known at time t. The vector of excess returns between times t and $t+1$ is denoted by $\mathbf{r}_{t+1}$ and its value is realized at time $t+1$ but unknown at time t. During each period, t, investors choose portfolio weights, $\boldsymbol{\phi}_t$ on the risky assets and realize excess returns $\boldsymbol{\phi}_t'\mathbf{r}_{t+1}$ at time $t+1$.

6.1. Nonrobust Mean-Variance Portfolio Choices

We begin by reviewing standard optimal mean-variance portfolio choices. As proposed by Markowitz (1952), an investor with mean-variance preferences prefers portfolios with a high mean and a low variance, and seeks to maximize his mean return minus a constant times his variance,

$$E(\boldsymbol{\phi}_t'\mathbf{r}_{t+1} + r_{f,t+1}|\mathcal{F}_t) - \frac{\gamma}{2}V(\boldsymbol{\phi}_t'\mathbf{r}_{t+1} + r_{f,t+1}|\mathcal{F}_t) \tag{16}$$

by choice of portfolio weights $\boldsymbol{\phi}_t$ where E denotes expectation, V denotes variance, and γ measures risk aversion. The solution to the nonrobust mean-variance problem is as follows.

Mean-Variance Optimal Portfolio Choices. *Optimal (nonrobust) mean variance portfolio choices at time t are*

$$\boldsymbol{\phi}_t = \left(\frac{1}{\gamma}\right)\Sigma_{t+1|t}{}^{-1}\,\boldsymbol{\mu}_{t+1|t} \tag{17}$$

where $\boldsymbol{\mu}_{t+1|t}$ and $\Sigma_{t+1|t}$ are the means and covariances of asset returns in the reference distribution. The optimal choices follow from the first-order condition for maximizing Equation (16) by choice of $\boldsymbol{\phi}_t$.

Investors with BA-HVAR expectations, who wish to compute nonrobust portfolio choices, solve the mean-variance problems in this section with the conditional means and covariances listed in Equations (15c) and (15d). Unlike robust investors, who are described in the next section, nonrobust investors trust that means and covariances of asset returns are really $\boldsymbol{\mu}_{t+1|t}$ and $\Sigma_{t+1|t}$.

6.2. Robust Portfolio Choices for a Fixed Value of Model Uncertainty

We now discuss optimal robust mean-variance portfolio choices that are designed for situations when the reference distribution, described in Section 5.2, may not accurately describe reality and none of the models described in Section 4 are correct. Although robust BA-HVAR investors believe that the reference distribution is the best available model, we assume that they doubt any specification, regardless of its complexity and flexibility, is able to perfectly predict the distribution of future returns. We first describe how robust investors compute portfolio choices for a single (given) value of model uncertainty aversion and then how they dynamically find the optimal value of model uncertainty aversion.

Following Hansen and Sargent (1995), Hansen et al. (1999), Maenhout (2004, 2006), Hansen and Sargent (2007a), and Anderson and Cheng (2016), robust investors compute optimal robust portfolio choices, for a fixed value of model uncertainty, by solving a max-min problem.

Robust Portfolio Choice Problem. *For the fixed value of model uncertainty θ_t, robust optimal mean-variance portfolio choices, $\boldsymbol{\phi}_t$, at time t solve*

$$\begin{aligned}\max_{\boldsymbol{\phi}_t}\min_{\varrho_t} &\int\left[\boldsymbol{\phi}_t'\mathbf{r}_{t+1} + r_{f,t+1} - \frac{\gamma}{2}(\boldsymbol{\phi}_t'z_{t+1})^2\right] \\ &\times \varrho_t(\mathbf{r}_{t+1})\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)\,d\mathbf{r}_{t+1} \\ &+ \frac{1}{\theta_t}\int \varrho_t(\mathbf{r}_{t+1})\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)\log\varrho_t(\mathbf{r}_{t+1})\,d\mathbf{r}_{t+1}\end{aligned} \tag{18}$$

subject to the constraint

$$\int \varrho_t(\mathbf{r}_{t+1})\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)\,d\mathbf{r}_{t+1} = 1, \tag{19}$$

where

$$z_{t+1} = \mathbf{r}_{t+1} - \int \mathbf{r}_{t+1}\overline{\boldsymbol{P}}(\mathbf{r}_{t+1}|\mathcal{F}_t)d\mathbf{r}_{t+1}$$

is the deviation of returns from the investor's best approximation of mean returns and $\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)$ is the reference distribution defined in Equation (15b).

The robust problem uses the scalar model uncertainty parameter, θ_t, and the endogenous density perturbation function ϱ_t. Although investors believe that the reference distribution $\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)$ is a reasonable approximation of $\mathbf{r}_{t+1}$'s distribution, they worry that the actual density might be given by alternative distributions of the form $\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)\varrho_t(\mathbf{r}_{t+1})$. Investors choose the function ϱ_t to minimize their objective while penalizing deviations from their reference specification with the relative entropy of the reference distribution to alternative distributions. If $\varrho_t(\mathbf{r}_{t+1})$ equals the constant one, for all $\mathbf{r}_{t+1}$, then the reference and alternative densities are identical and the penalty is zero. For any other setting of ϱ_t, the penalty is positive and increases, as the reference and alternative densities become farther apart.[11]

The minimizing choice of ϱ_t depends on portfolio choices and is

$$\varrho_t^*(\mathbf{r}_{t+1}) = \frac{\exp\left[-\theta_t \boldsymbol{\phi}_t' \mathbf{r}_{t+1} + \frac{\gamma\theta_t}{2}(\boldsymbol{\phi}_t' z_{t+1})^2\right]}{E\left(\exp\left[-\theta_t \boldsymbol{\phi}_t' \mathbf{r}_{t+1} + \frac{\gamma\theta_t}{2}(\boldsymbol{\phi}_t' z_{t+1})^2\right] \middle| \mathcal{F}_t\right)}, \quad (20)$$

so that investors focus their concerns on the alternative distribution $\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)\varrho_t^*(\mathbf{r}_{t+1})$. Plugging $\varrho_t^*(\mathbf{r}_{t+1})$ into the robust objective yields a risk-sensitive mean-variance portfolio choice optimization problem:

$$\max_{\boldsymbol{\phi}_t} \left(-\frac{1}{\theta_t}\right) \log \int \exp\Big[-\theta_t \boldsymbol{\phi}_t' \mathbf{r}_{t+1} + r_{f,t+1} + \frac{\gamma\theta_t}{2}(\boldsymbol{\phi}_t{}' \mathbf{z}_{t+1})^2\Big] \overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)\, d\mathbf{r}_{t+1}. \quad (21)$$

In the risk-sensitive problem, expectations are taken with respect to the reference distribution, $\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)$. As discussed by Hansen and Sargent (2007b), risk-sensitive optimization problems are observationally equivalent to robust optimization problems.

The model uncertainty parameter $\theta_t \geq 0$ measures investors' confidence in their reference model, where larger values indicate less confidence and smaller values indicate more confidence. As $\theta_t \downarrow 0$, investors have complete confidence in their approximating framework and portfolio choices converge to optimal nonrobust portfolio choices.[12] We index the model uncertainty parameter by time because, as discussed in Section 6.3, investors' confidence in their reference model can vary over time.

Although there are no known formulas for optimal robust mean-variance portfolio choices when the reference distribution is normal or a mixture of normals, both cases can be accurately solved using a nonlinear numerical optimization routine since analytical first derivatives of the objectives are available. For other reference distributions, there may be no interesting solutions to robust portfolio optimization problems. For example, if the reference distribution is multivariate t or a mixture of multivariate t's, then the objective is always negative infinity unless $\boldsymbol{\phi}_t = 0$.

Robust BA-HVAR investors assume that the reference model is a mixture of normals and compute robust portfolio choices using Equation (21), for the value of model uncertainty determined in the next section.[13]

6.3. Optimal Model Uncertainty Aversion

The model uncertainty aversion parameter, θ_t, is not a structural preference parameter but rather depends upon investors' confidence in their model. In different environments, the same investor likely will have different levels of model uncertainty aversion. As described by Anderson et al. (2003), one way to calibrate model uncertainty aversion is with detection probabilities. This approach chooses the largest value of model uncertainty aversion for which the alternative model, which investors focus on, in the robust optimal portfolio choice problem cannot be empirically distinguished from the reference specification. Following Anderson and Cheng (2016), we take a different approach and let investors choose the model uncertainty aversion parameter with the best past performance.

For tractability, we assume that there are a finite number of possible values for the model uncertainty parameter, θ_t, and we let Θ be the set of all possible values. The set Θ is time-invariant, but the choice of θ_t can be time-varying. In the applications presented in later sections, we let there be 10 possible values of model uncertainty:

$$\Theta = \left\{0, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16\right\}.$$

One possible choice of model uncertainty is zero, so that investors can choose not to be robust and use standard mean-variance portfolio choices. For comparison purposes, we also consider examples where the only possible value of model uncertainty is zero, so that investors always use nonrobust portfolio choices.

Let $\bar{\theta}$ be a predetermined time-invariant desired level of robustness that is used to calibrate the time-varying model uncertainty parameter θ_t. Let $\boldsymbol{\phi}_{s,\theta}$ be the optimal portfolio choices at time s, if the model uncertainty parameter is θ.[14] At each date, investors choose the model uncertainty aversion parameter to be the constant value that maximizes the historical robust certainty equivalent for the predetermined parameter $\bar{\theta}$.

Optimal Model Uncertainty Aversion. *Let $b > 1$ be a fixed date. At each date $t >> b$, robust investors choose the model uncertainty aversion parameter that maximizes the previous in-sample robust certainty equivalent:*

$$\theta_t = \underset{\theta \in \Theta}{\operatorname{argmax}} \left\{ -\left(\frac{1}{\bar{\theta}}\right) \log \bar{E}_{s=b}^t \exp\left[-\bar{\theta}\, r_{p,s,\theta} + \left(\frac{\gamma\bar{\theta}}{2}\right) z_{p,s,\theta}{}^2\right] \right\}, \quad (22)$$

where

$$r_{p,s,\theta} = \boldsymbol{\phi}_{s-1,\theta}{}' \mathbf{r}_s$$

and where

$$z_{p,s,\theta} = r_{p,s,\theta} - \bar{E}_{s=b}^t[r_{p,s,\theta}]$$

is the deviation of returns from their sample mean when model uncertainty aversion is fixed at θ at all past dates. If multiple values of θ achieve the maximum, then we randomly choose θ_t to be one of the values that achieves the maximum.[15]

The historical robust certainty equivalent [in Equation (22)] is the sample analogue of the robust (or risk-sensitive) objective in Equation (21) when the desired level of robustness is $\bar{\theta}$. The risk-free rate does not affect the calibration and is not used in Equation (22). At date t, investors set their model uncertainty parameter to be θ_t and choose portfolio weights $\boldsymbol{\phi}_t \equiv \boldsymbol{\phi}_{t,\theta_t}$ that realize a return of

$$r_{p,t+1} = r_{p,t+1,\theta_t} = \boldsymbol{\phi}_{t,\theta_t}{}' \boldsymbol{r}_{t+1}$$

at time $t+1$. The calibration should be based on many periods, so that t should be much greater than b. Date b should be large enough for the portfolio choice strategy to initialize.

The predetermined parameter $\bar{\theta}$ reflects the investors' desired level of robustness. One possible choice of $\bar{\theta}$ is zero, in which case the investors choose the model uncertainty aversion parameter to be the constant value that maximizes the historical nonrobust certainty equivalent. This is the case considered by Anderson and Cheng (2016). We set the desired level of robustness to a small number ($\bar{\theta} = \frac{1}{16}$). For most applications, results are identical if $\bar{\theta}$ is zero or a small positive number. However, we find in simulated results that setting $\bar{\theta}$ to a small positive number can avoid occasional bad outcomes, which occur when investors become too aggressive because they were lucky in the past.[16]

7. Out-of-Sample Performance

We compare the performance of nonrobust and robust BA-HVAR portfolio methods to other strategies on a variety of daily, weekly, and monthly data sets from the Center for Research of Security Prices (CRSP) and Kenneth French's website. The other strategies include the $1/N$ and market methods, as well as robust and nonrobust versions of the first-order Bayesian vector autoregression, Bayesian averaging, rolling, historical, Jorion, and Kan-Zhou methods. The $1/N$ method always invests an equal amount in each risky asset and nothing in the risk-free asset. The market method always invests all wealth in the value-weighted market. The first-order Bayesian vector autoregression method uses a constant and the first lag of returns as explanatory variables and is a Bayesian variation on the first-order vector autoregression proposed by DeMiguel et al. (2014). The Bayesian averaging method is Anderson and Cheng (2016)'s Bayesian-averaging strategy. The rolling method estimates the means and covariances of excess returns using a fixed window. The historical method estimates the means and covariances of excess returns using all available historical data. The Jorion method uses Jorion (1986)'s shrinkage estimates of the means and covariances of asset returns. The Kan-Zhou method is Kan and Zhou (2007)'s three-fund rule that modifies shrinkage estimators to minimize the utility loss from estimation errors.[17]

The data sets include assets that are portfolios of stocks sorted on beta, size, standard deviation, industry, book-to-market, momentum, short-term reversal, and long-term reversal.[18] Excess returns are formed by subtracting the risk-free rate from returns. For CRSP portfolios, we approximate the risk-free rate from CRSP's monthly data on the 30-day Treasury bill return, and, for Ken French's portfolios, we use the risk-free rate reported by Ken French. Following DeMiguel et al. (2009), the value-weighted market return is added to all of the data sets. For CRSP portfolios, we use the value-weighted market return on NYSE, AMEX, and NASDAQ stocks provided by CRSP. For Ken French's portfolios, we use the value-weighted market return provided by Ken French. A summary of the daily datasets is provided in Table 1, and summaries for other frequencies are available upon request.

We focus on an example of BA-HVAR that uses many of the commonly studied lags in financial economics as explanatory variables. With monthly data, the explanatory variables ($\mathbf{x}_{t-1}$), are a constant and the previous month's excess returns, as well as average excess returns over the past year and 10 years. With weekly data, the explanatory variables are a constant and the previous week's excess returns, as well as average excess returns over the past month, year, and 10 years. With daily data, the explanatory variables are a constant and the previous day's excess returns, as well as average excess returns over the last week, month, year, and 10 years. These explanatory variables were inspired by Corsi (2009)'s HAR volatility model.

In the following subsections, we show that BA-HVAR achieves much higher out-of-sample certainty equivalents and Sharpe ratios of portfolio returns than other methods. In subsequent sections, we show that BA-HVAR can better predict portfolio returns than other methods. Readers interested in immediately seeing the most important results are encouraged to look at Tables 2 through 4 for out-of-sample certainty equivalents, and Tables 5 and 6 for predictions.

7.1. Certainty Equivalents

In this subsection, we compare the certainty equivalents of BA-HVAR to the certainty equivalents obtained by other methods. We consider the standard measure of certainty equivalents that is discussed in many other papers, as well as an alternative measure that we dub the *perceived certainty equivalent*. The perceived certainty equivalent uses investors' conditional expectations for some quantities, rather than unconditional means, and is a more appropriate measure of the certainty equivalent when the means and variances of assets are time-varying. We show

Table 1. Data Sets

Data set	Name	Number	Sample size	Start date	End date
Panel A: CRSP daily data					
Beta	Beta	11	24,038	01-02-1926	12-30-2016
Size	Size	11	24,038	01-02-1926	12-30-2016
Standard deviation	Std	11	24,038	01-02-1926	12-30-2016
Equal-weighted market	Em	2	24,038	01-02-1926	12-30-2016
Panel B: Ken French's value-weighted daily data					
Industry	Ind	11	23,889	07-01-1926	12-30-2016
Size	Size	11	23,889	07-01-1926	12-30-2016
Book-to-market	Beme	11	23,889	07-01-1926	12-30-2016
Long-term reversal	Ltr	11	22,788	03-20-1930	12-30-2016
Momentum	Mom	11	23,788	11-03-1926	12-30-2016
Short-term reversal	Str	11	23,889	07-01-1926	12-30-2016
Size and book-to-market	Size-beme	26	23,889	07-01-1926	12-30-2016
Size and momentum	Size-mom	26	20,453	01-17-1938	12-30-2016
Panel C: Ken French's equal-weighted daily data					
Industry	Ind	11	23,889	07-01-1926	12-30-2016
Size	Size	11	23,889	07-01-1926	12-30-2016
Book-to-market	Beme	11	23,889	07-01-1926	12-30-2016
Long-term reversal	Ltr	11	22,788	03-20-1930	12-30-2016
Momentum	Mom	11	23,788	11-03-1926	12-30-2016
Short-term reversal	Str	11	23,889	07-01-1926	12-30-2016
Size and book-to-market	Size-beme	26	23,889	07-01-1926	12-30-2016
Size and momentum	Size-mom	26	20,453	01-17-1938	12-30-2016

Notes. This table displays information about the daily data sets used in this paper. The name column provides an abbreviated name for the datasets used in subsequent tables. The number column lists the number of assets in the data set. All data sets include the value-weighted market. The sample size column provides the number of observations between the start and end dates of the available data. For each data set, we use the longest block of data that does not have any missing values.

that robust BA-HVAR portfolio choices achieve much higher standard and perceived certainty equivalents than other methods.

7.1.1. Standard and Perceived Certainty Equivalents. We begin by discussing appropriate out-of-sample certainty equivalent measures when the means and variances of asset returns are time-varying and the actual distribution of asset returns is not known. We let $\widetilde{E}y$ denote the expected value of y under the true (unknown) distribution of returns. At time t, we assume that investors would like to maximize

$$\widetilde{E}\left[r_{p,t+1} + r_{f,t+1} - \frac{\gamma}{2}\left(r_{p,t+1} - \tilde{r}_{p,t+1|t}\right)^2 | \mathcal{F}_t\right], \qquad (23)$$

where $r_{p,t+1} = \boldsymbol{\phi}_t'\mathbf{r}_{t+1}$ is the excess return on the investor's portfolio and $\tilde{r}_{p,t+1|t} = \widetilde{E}[r_{p,t+1}|\mathcal{F}_t]$ is the *conditional* expected return on the investor's portfolio. However, investors are unable to find the optimal portfolio choices for this problem, because they do not know the true distribution of returns. Instead, investors solve the nonrobust problem or the robust problem, discussed in previous sections. Nonrobust investors replace the true distribution in Equation (23) with the reference distribution and compute optimal nonrobust portfolio choices using Equation (17). Robust investors hope to achieve good out-of-sample performance by solving the robust optimization problem [in Equation (21)], using the optimal value of model uncertainty aversion (defined in Section 6.3).[19]

When evaluating the performance of portfolio choice methods, we also would like to use the true distribution of returns, to the extent possible. If we observed actual returns and observed the true conditional expectations, then a good measure of performance would be the sample certainty equivalent

$$\bar{E}\left[r_{f,t+1} + r_{p,t+1} - \frac{\gamma}{2}\left(r_{p,t+1} - \tilde{r}_{p,t+1|t}\right)^2\right], \qquad (24)$$

where $\bar{E}y_{t+1}$ denotes the sample mean of y_{t+1}. Unfortunately, we do not observe conditional expectations. Most other studies use an alternative measure of certainty equivalents that replaces conditional means with sample means. The standard certainty equivalent, often referred to as simply the *certainty equivalent* in other papers, is

$$\bar{E}\left[r_{f,t+1} + r_{p,t+1} - \frac{\gamma}{2}\left(r_{p,t+1} - \bar{r}_p\right)^2\right], \qquad (25)$$

where $\bar{r}_p = \bar{E}r_{p,t+1}$ is the sample mean of the portfolio return. The standard measure in Equation (25) differs from the sample certainty equivalent in Equation (24), because it uses the sample mean of portfolio returns, $\bar{r}_p$, rather than the conditional mean $\tilde{r}_{p,t+1|t}$. The standard measure is convenient in practice, because it does not rely on unobservable expectations, but it does not follow from the preferences of investors, unless the conditional means and variances of portfolio returns are constant over time. Equation (24) follows from investor's preferences and is the appropriate certainty equivalent when means and variances are time-varying.

In order to evaluate performance using the sample certainty equivalent, in Equation (24) we need to approximate the conditional expectation $\tilde{r}_{p,t+1|t}$. A natural approximation of the true conditional expectation is the conditional expectation used by an investor's portfolio choice method. We define the perceived certainty equivalent to be the sample certainty equivalent, where investors' expectations are used in place of the true conditional expectations. The perceived certainty equivalent is

$$\bar{E}\left[r_{f,t+1} + r_{p,t+1} - \frac{\gamma}{2}\left(r_{p,t+1} - \hat{r}_{p,t+1|t}\right)^2\right], \tag{26}$$

where $\hat{r}_{p,t+1|t}$ is the conditional expectation used by investors. We call this the *perceived certainty equivalent*, because it is based on an investor's perceptions of conditional means that may be different from the actual conditional means.

This perceived certainty equivalent has a very desirable feature. Although the perceived certainty equivalent involves the possibly inaccurate conditional expectations reported by a portfolio choice method, a method cannot, on average, increase its certainty equivalent by reporting incorrect expectations. Wrong conditional expectations lead to lower certainty equivalents because

$$\widetilde{E}\left(\bar{E}\left[\left(r_{p,t+1} - \hat{r}_{p,t+1|t}\right)^2 - \left(r_{p,t+1} - \tilde{r}_{p,t+1|t}\right)^2\right]\middle|\mathcal{F}_0\right) \geq 0$$

where the expected value is taken with respect to the true distribution.[20] Though, there are two possible drawbacks of perceived certainty equivalents. If we compare different methods using perceived certainty equivalents, then we are evaluating the mean-variance properties of their portfolios, as well as the accuracy of their conditional mean predictions. Also, perceived certainty equivalents cannot be used to evaluate portfolio choice methods that do not make predictions of conditional means such as the $1/N$ method and the market method.

Although perceived certainty equivalents are better measures of certainty equivalents than standard certainty equivalents, we report results for both measures, since standard certainty equivalents are prevalent in the literature.

7.1.2. Empirical Results. Tables 2 and 3 compare the perceived certainty equivalents of BA-HVAR to several alternative methods advocated in the recent literature. Table 2 presents certainty equivalent quantiles over 20 data sets at the monthly, weekly, and daily frequencies. Robust BA-HVAR achieves much higher certainty equivalents on every quartile at the monthly frequency. At the weekly frequency, robust BA-HVAR earns the highest certainty equivalent on four of the five quartiles, losing the highest quartile to the nonrobust Bayesian averaging (BA) strategy. At the daily frequency, robust BA-HVAR also earns the highest certainty equivalent on four of the five quartiles, in this case losing the highest quartile to the nonrobust BA-HVAR strategy. For every frequency, robust BA-HVAR's median certainty equivalent is statistically significantly higher (at the 95% confidence level) than the median certainty equivalents for other methods.[21] We also note that the median certainty equivalents of robust versions of the first-order Bayesian vector autoregression, Bayesian averaging, Jorion, Kan-Zhou, and rolling methods are at least as high as the median certainty equivalents of their nonrobust counterparts for every frequency.

Table 3 displays perceived certainty equivalents for each daily data set. Robust BA-HVAR earns the highest certainty equivalent on 18 of 20 daily data sets. Nonrobust BA-HVAR beats robust BA-HVAR on the equal-weighed market data set and the equal-weighted short-term reversals data set. In additional results available upon request, we find that robust BA-HVAR earns the highest certainty equivalent on 19 of 20 weekly datasets and 16 of 20 monthly data sets. In most cases when robust BA-HVAR wins, the differences are statistically significant, and in most cases in when it loses, the differences are not statistically significant.

Table 4 displays the standard (rather than the perceived) certainty equivalent quantiles at the monthly, weekly, and daily frequencies. We see that robust BA-HVAR earns the highest standard certainty equivalent on every quantile, at every frequency. For every frequency, robust BA-HVAR's median standard certainty equivalent is statistically significantly higher than the median standard certainty equivalents for every other method. The median standard certainty equivalents of the market and $1/N$ strategies are higher than the median standard certainty equivalents of four of the seven nonrobust methods at the monthly frequency, five of the seven nonrobust methods at the weekly frequency, and three of the seven nonrobust methods at the daily frequency. However, the median standard certainty equivalents of the market and $1/N$ strategies are lower than the median standard certainty equivalents of all the robust strategies at every frequency. We also see that, for all methods, the standard certainty

Table 2. Perceived Certainty Equivalent Quartiles

	Robust strategies							Nonrobust strategies						
Quartile	BA-HVAR	B-VAR1	BA	JOR	KZ	ROLL	HIST	BA-HVAR	B-VAR1	BA	JOR	KZ	ROLL	HIST
Panel A: Monthly data														
Min	**1.21**	0.08	0.15	0.19	0.30	–0.00	0.36	–9.98*	–56.43*	–113.90*	–68.02*	–5.68*	–569.67*	–0.67*
25%	**3.10**	1.32*	1.72	0.81*	0.82*	0.82*	0.95*	–2.76*	–17.53*	–39.27*	–10.69*	–3.08*	–49.21*	0.78*
Median	**9.86**	3.44*	4.16*	2.85*	2.28*	2.88*	3.08*	–0.12*	0.09*	–21.83*	–7.10*	–1.54*	–35.23*	2.43*
75%	**12.19**	6.14*	6.81*	4.42*	4.91*	4.68*	4.83*	9.56	5.40*	–2.52*	–3.83*	0.79*	–32.32*	4.38*
Max	**22.75**	13.09*	12.02*	15.74*	15.29*	14.96*	11.99*	18.79	12.55*	6.65*	2.49*	11.19*	–0.43*	11.99*
Panel B: Weekly data														
Min	**0.84**	–2.04*	0.00*	0.17	0.17	0.05	0.16	–73.61*	–48.64*	–143.99*	–0.92*	–0.48*	–8.86*	0.03*
25%	**7.81**	1.37*	0.99*	0.41*	0.42*	0.35*	0.57*	–19.86*	–14.25*	–43.90*	–0.50*	–0.08*	–3.41*	0.60*
Median	**14.67**	5.22*	3.30*	3.54*	3.52*	3.31*	2.34*	–1.03*	–3.05*	–21.31*	1.34*	2.11*	–2.74*	2.31*
75%	**18.27**	6.39*	7.29*	8.02*	8.18*	7.68*	3.17*	2.24*	2.52*	–6.97*	6.32*	6.77*	1.95*	3.19*
Max	70.18	33.22*	96.05	57.99	58.14*	59.48	17.14*	51.12	46.30*	**158.25***	10.08	18.45	5.01	17.14*
Panel C: Daily data														
Min	**2.61**	–0.20*	0.22*	0.06*	0.07*	0.03*	0.03*	–60.70*	–171.60*	–125.31*	–0.21*	–0.15*	–0.62*	0.00*
25%	**7.71**	0.59*	0.81*	0.14*	0.14*	0.13*	0.15*	–26.64*	–79.18*	–39.91*	–0.04*	0.01*	–0.40*	0.15*
Median	**15.46**	1.34*	3.32*	1.21*	1.21*	1.17*	0.60*	–3.76*	–36.80*	–20.92*	0.74*	0.84*	0.41*	0.58*
75%	**17.07**	2.84*	4.62*	2.94*	3.00*	2.86*	0.97*	–1.27*	–23.01*	–9.48*	2.58*	2.69*	1.96*	0.97*
Max	63.90	5.47*	26.00*	19.86*	19.69*	20.04*	13.84*	**98.73**	–4.03*	81.97	22.96*	23.42*	21.69*	13.84*

Notes. This table displays perceived certainty equivalent quartiles over 20 monthly, weekly, and daily data sets for the robust and nonrobust BA-HVAR (Bayesian-averaging heterogeneous vector autoregression), B-VAR1 (first-order Bayesian vector autoregression), BA (Bayesian averaging), JOR (Jorion), KZ (Kan-Zhou), ROLL (rolling), and HIST (historical) methods. We set risk aversion to one, and the robust methods all use the optimal value of model uncertainty when the possible values are $\Theta = \{0, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16\}$. The prior for BA-HVAR is the empirical prior and the explanatory variables include a constant and average excess returns over the past day (only for daily data), week (only for daily and weekly data), month, year, and 10 years. The window sizes for the JOR, KZ, and ROLL methods are 60 months for monthly data, 260 weeks for weekly data, and 1,260 days for daily data. Certainty equivalents are expressed in percent and are reported for all periods between 1977 and 2016. The highest certainty equivalent in each row is in boldface. An asterisk indicates that the quantile is significantly different from the quantile for robust BA-HVAR at the 95% confidence level.

Table 3. Daily Perceived Certainty Equivalents

	Robust strategies							Nonrobust strategies						
Quartile	BA-HVAR	B-VAR1	BA	JOR	KZ	ROLL	HIST	BA-HVAR	B-VAR1	BA	JOR	KZ	ROLL	HIST
Panel A: CRSP daily data														
Beta	**9.93**	1.17*	2.48*	0.90*	0.89*	0.85*	0.59*	–3.94*	–27.87*	–11.14*	0.70*	0.73*	0.38*	0.55*
Size	**16.13**	2.89*	3.80*	0.16*	0.16*	0.16*	0.13*	–33.62*	–22.79*	–13.34*	–0.12*	–0.06*	–0.50*	0.13*
Std	**14.01**	2.96*	3.78*	1.35*	1.36*	1.34*	1.02*	–21.17*	–32.66*	–10.23*	1.33*	1.35*	1.07*	1.03*
Em	6.72	1.26*	0.87*	0.59*	0.60*	0.61*	0.49*	**7.74***	–24.20*	–6.59*	0.44*	0.44*	0.43*	0.51*
Panel B: Ken French's value-weighted daily data														
Ind	**8.70**	0.29*	0.30*	0.07*	0.07*	0.03*	0.03*	1.10*	–20.60*	–54.22*	–0.21*	–0.15*	–0.62*	0.00*
Size	**12.58**	1.31*	1.82*	0.12*	0.12*	0.10*	0.06*	–25.72*	–43.77*	–12.00*	–0.17*	–0.12*	–0.56*	0.06*
Beme	**2.61**	–0.20*	0.22*	0.06*	0.07*	0.04*	0.07*	–3.41*	–4.59*	–28.97*	–0.10*	–0.04*	–0.44*	0.07*
Ltr	**2.89**	0.03*	0.43*	0.12*	0.12*	0.09*	0.07*	–1.23*	–4.03*	–33.01*	–0.00*	–0.05*	–0.36*	0.07*
Mom	**3.32**	–0.03*	0.75*	0.09*	0.08*	0.08*	0.16*	–9.72*	–23.24*	–34.70*	–0.08*	–0.03*	–0.48*	0.16*
Str	**2.87**	0.20*	0.28*	1.91*	1.90*	1.85*	0.76*	–4.45*	–4.71*	–32.74*	1.71*	1.72*	1.36*	0.76*
Size-beme	**15.86**	0.88*	2.17*	0.90*	0.92*	0.78*	0.17*	–27.55*	–104.38*	–46.85*	0.51*	0.67*	–0.30*	0.17*
Size-mom	**18.61**	2.99*	2.95*	1.24*	1.23*	1.22*	0.37*	–60.70*	–171.60*	–125.31*	0.78*	0.95*	–0.17*	0.31*
Panel C: Ken French's equal-weighted daily data														
Ind	**16.64**	1.44*	5.78*	1.18*	1.19*	1.12*	0.61*	–3.59*	–53.83*	–28.50*	0.52*	0.57*	–0.03*	0.61*
Size	**17.49**	2.26*	3.70*	3.22*	3.22*	2.99*	0.90*	–1.50*	–44.28*	–7.53*	2.88*	2.90*	2.58*	0.90*
Beme	**16.42**	1.47*	4.37*	3.27*	3.27*	3.21*	0.92*	–1.31*	–79.34*	–11.62*	3.04*	3.07*	2.83*	0.92*
Ltr	**15.12**	1.37*	4.52*	2.85*	2.85*	2.81*	1.07*	–0.39*	–30.67*	–8.12*	2.66*	2.68*	2.41*	1.07*
Mom	**15.80**	1.17*	5.00*	1.47*	1.49*	1.43*	0.71*	–1.87*	–40.95*	–8.74*	1.25*	1.28*	0.90*	0.73*
Str	63.90	5.13*	26.00*	19.86*	19.69*	20.04*	13.84*	**98.73**	–84.20*	81.97	22.96*	23.42*	21.69*	13.84*
Size-beme	**22.72**	2.79*	4.71*	4.50*	4.52*	4.23*	1.31*	–43.71*	–86.80*	–45.13*	4.26*	4.38*	3.55*	1.31*
Size-mom	**23.36**	5.47*	5.99*	3.02*	3.14*	2.92*	1.61*	–54.47*	–79.01*	–114.86*	2.51*	2.71*	1.52*	1.60*

Notes. This table displays daily perceived certainty equivalents on 20 daily data sets for the robust and nonrobust BA-HVAR (Bayesian-averaging heterogeneous vector autoregression), B-VAR1 (first-order Bayesian vector autoregression), BA (Bayesian averaging), JOR (Jorion), KZ (Kan-Zhou), ROLL (rolling), and HIST (historical) methods. We set risk aversion to one, and the robust methods all use the optimal value of model uncertainty when the possible values are $\Theta = \{0, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16\}$. The prior for BA-HVAR is the empirical prior, and the explanatory variables include a constant and average excess returns over the past day, week, month, year, and 10 years. The window sizes for the JOR, KZ, and ROLL methods are 1,260 days. The highest certainty equivalent in each row is in boldface. Certainty equivalents are expressed in percent and are reported for all days between 1977 and 2016. The highest certainty equivalent in each row is in boldface. An asterisk indicates that the certainty equivalent is significantly different from the certainty equivalent for robust BA-HVAR at the 95% confidence level.

Table 4. Standard Certainty Equivalent Quartiles

	Robust strategies							Nonrobust strategies							Other strategies	
Quartile	BA-HVAR	B-VAR1	BA	JOR	KZ	ROLL	HIST	BA-HVAR	B-VAR1	BA	JOR	KZ	ROLL	HIST	MKT	1/N
Panel A: Monthly data																
Min	**1.24**	0.11	0.22	0.20	0.31	0.00	0.37	−10.72*	−38.54*	−88.20*	−85.68*	−11.11*	−622.00*	−0.56*	0.88	0.89
25%	**3.04**	1.88*	1.74	0.83*	0.84*	0.83*	0.95*	−1.00*	−5.87*	−25.16*	−10.01*	−3.32*	−51.13*	0.74*	0.90*	1.01*
Median	**9.10**	3.75*	4.23*	2.75*	2.17*	2.77*	3.08*	1.23	0.64*	−14.89*	−6.55*	−1.51*	−32.55*	2.44*	0.90*	1.05*
75%	**11.65**	6.04*	6.90*	4.28*	4.23*	4.47*	4.63*	6.72	3.96*	−0.16*	−5.25*	0.16*	−27.08*	4.41*	0.90*	1.16*
Max	**18.32**	10.08*	10.50*	13.17*	13.21*	12.96*	12.12*	11.12	8.98*	3.78*	0.52*	6.44*	0.28*	12.12*	0.90*	1.42*
Panel B: Weekly data																
Min	**0.84**	−2.33*	0.08*	0.17	0.16	0.05	0.16	−6,335.29*	−71.40*	−209.36*	−185.30*	−160.56*	−221.68*	0.03*	0.20	0.21
25%	**6.96**	0.96*	1.23*	0.41*	0.41*	0.35*	0.58*	−58.28*	−17.23*	−45.67*	−0.67*	−0.28*	−5.67*	0.60*	0.21*	0.23*
Median	**12.34**	4.24*	3.15*	3.40*	3.35*	3.19*	2.35*	−30.25*	−7.11*	−19.51*	0.05*	1.42*	−2.87*	2.31*	0.21*	0.27*
75%	**15.31**	5.42*	6.86*	7.00*	7.06*	6.85*	3.48*	−0.82*	0.08*	−10.56*	2.15*	3.43*	−0.02*	3.50*	0.21*	0.32*
Max	**45.99**	21.09*	38.40	39.29	41.59	39.42	18.94*	5.90*	4.35*	−3.25*	4.96*	5.46*	1.90*	18.94*	0.21*	0.39*
Panel C: Daily data																
Min	**2.37**	−2.61*	0.23*	0.06*	0.07*	0.03*	0.03*	−251.77*	−255.87*	−97.31*	−39.73*	−38.17*	−42.14*	0.00*	0.04*	0.04*
25%	**6.41**	−0.14*	0.83*	0.14*	0.14*	0.13*	0.15*	−101.18*	−120.88*	−31.57*	−0.09*	−0.03*	−0.45*	0.15*	0.04*	0.05*
Median	**12.73**	0.67*	3.20*	1.20*	1.19*	1.16*	0.60*	−37.02*	−62.02*	−20.39*	0.60*	0.69*	0.16*	0.58*	0.04*	0.06*
75%	**14.49**	1.47*	4.42*	2.76*	2.79*	2.71*	0.97*	−12.74*	−31.91*	−10.40*	1.85*	1.97*	1.18*	0.98*	0.04*	0.07*
Max	**36.42**	3.73*	23.18*	18.61*	18.28*	18.89*	14.65*	5.42*	−3.94*	−4.27*	3.76*	3.89*	3.05*	14.65*	0.04*	0.08*

Notes. This table displays standard certainty equivalent quartiles over 20 monthly, weekly, and daily data sets for the robust and nonrobust BA-HVAR (Bayesian-averaging heterogeneous vector autoregression), B-VAR1 (first-order Bayesian vector autoregression), BA (Bayesian averaging), JOR (Jorion), KZ (Kan-Zhou), ROLL (rolling), and HIST (historical) methods. Results for the MKT (market) and 1/N methods are also reported. We set risk aversion to one, and the robust methods all use the optimal value of model uncertainty when the possible values are $\Theta = \{0, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16\}$. The prior for BA-HVAR is the empirical prior, and the explanatory variables include a constant and average excess returns over the past day (only for daily data), week (only for daily and weekly data), month, year, and 10 years. The window sizes for the JOR, KZ, and ROLL methods are 60 months for monthly data, 260 weeks for weekly data, and 1,260 days for daily data. Certainty equivalents are expressed in percent and are reported for all periods between 1977 and 2016. The highest certainty equivalent in each row is in boldface. An asterisk indicates that the quantile is significantly different from the quantile for robust BA-HVAR at the 95% confidence level.

equivalents are almost always lower than the perceived certainty equivalents.

We have been assuming that investors update portfolios every period, which possibly can involve high transaction costs. As Online Appendix D describes, BA-HVAR can be adapted so that portfolios are periodically optimally rebalanced. Online Tables 3 and 4 show that periodically rebalanced portfolios perform very well. Online Table 3 reports perceived certainty equivalent quantiles for monthly and weekly optimally rebalanced portfolios when estimates are updated daily. Robust BA-HVAR earns the highest certainty equivalent on all five monthly quantiles and all five weekly quantiles. Using estimates at a higher frequency improves performance, and robust BA-HVAR achieves much higher monthly certainty equivalents when daily, rather than monthly, estimates are used. The median monthly certainty equivalent when monthly estimates are used is 9.86% (see Table 2) and the median monthly certainty equivalent when daily estimates are used is 15.73%, leading to a gain of about 6% per month. The benefits of using high-frequency estimates for weekly portfolio choices are smaller. The median weekly certainty equivalent when weekly estimates are used is 14.67%, and the median weekly certainty equivalent when daily estimates are used is 15.21%, leading to a gain of about 0.5% per week. Online Table 4 reports perceived certainty equivalent quantiles for quarterly and yearly optimally rebalanced portfolios when estimates are updated monthly. Robust BA-HVAR has a higher certainty equivalent, than other methods, on four of the five quarterly quantiles and four of the five yearly quantiles.

In unreported results, we find that the certainty equivalents of BA-HVAR are stable over subperiods. In daily data, robust BA-HVAR achieves higher certainty equivalents than other methods on all five quartiles, using data between 1997 and 2016, and four of five quartiles using data between 1977 and 1996. For the 1977–1996 subperiod, robust BA-HVAR loses the highest quantile to nonrobust BA-HVAR. We also find that robust BA-HVAR does well in both recessions and expansions. Somewhat surprisingly, the median certainty equivalent of robust BA-HVAR is slightly higher in recessions than in expansions.

7.2. Sharpe Ratios

Robust BA-HVAR almost always achieves higher Sharpe ratios than other methods. The Sharpe ratio of a portfolio is the sample mean of its excess return divided by the sample standard deviation of its excess return. Online Table 1 displays the Sharpe ratio quantiles of BA-HVAR and alternative methods. Robust BA-HVAR has higher Sharpe ratios, than every other method, at every quantile for weekly and daily frequencies. Robust BA-HVAR also earns the highest Sharpe ratio on four of the five monthly quantiles, losing the 25% percentile to nonrobust BA-HVAR. Online Table 2 displays Sharpe ratios on individual data sets and shows that BA-HVAR earns higher Sharpe ratios than other methods on all 20 daily data sets. In additional results available upon request, we find that robust BA-HVAR earns higher Sharpe ratios than other methods, on 19 of 20 weekly data sets and 16 of 20 monthly data sets.

The differences, in most cases, are statistically significant. At the daily frequency, robust BA-HVAR's Sharpe ratios at every quantile are statistically significantly better than the Sharpe ratios produced by other methods.[22] At the weekly frequency, BA-HVAR's Sharpe ratios are statistically significantly better than the Sharpe ratios of other methods at the 25th, 50th, and 75th percentiles. At the monthly frequency, BA-HVAR's Sharpe ratios are statistically significantly better than the Sharpe ratios of other methods at the 50th, 75th, and 100th percentiles.

8. Regimes

In order to understand better how BA-HVAR works, this section surveys some of the regimes detected by BA-HVAR. We discuss both ex post beliefs about the most likely historical regimes and real-time beliefs about the current regime. We also compare the number of regimes detected by BA-HVAR and BA.

8.1. Most Likely Regimes Since 1973

For each of our data sets, we determine the most likely regimes in recent U.S. history, using all available sample information. The most likely regimes are found by proceeding backward from the end of available data. We set the end date of the last regime to be the last period that data are available. We approximate the beginning date of the last regime by finding the model birthdate with the highest probability, using the model probabilities on the last period. We call the period after this model birthdate the beginning date of the last regime, because, using all the information available on the last period, it is most likely that the last regime started on this date.[23] At any date between this beginning date and the last period, the most likely birthdate could be different, but the probabilities formed on in-between dates are based on less information.

Now that the start and end dates of the most likely last regime are determined, we follow a similar procedure to find the start and end dates of the next-to-last regime. The end date of the next-to-last regime is set to the period before the beginning date of the last regime. We approximate the beginning date of the next-to-last regime by finding the model birthdate with the highest probability, using the model probabilities on the end date of the next-to-last regime. We use the

Table 5. Robust and Reference Daily Log-Likelihoods

	Robust distributions							Reference distributions						
Portfolios	BA-HVAR	B-VAR1	BA	JOR	KZ	ROLL	HIST	BA-HVAR	B-VAR1	BA	JOR	KZ	ROLL	HIST
Panel A: CRSP daily data														
Beta	**0**	−36,621	−23,584	−10,802	−10,757	−10,861	−37,050	**−1**	−36,557	−24,244	−10,815	−10,768	−10,892	−37,054
Size	0	−31,229	−36,165	481	**594**	351	−32,352	100	−31,166	−36,809	462	**578**	303	−32,352
Std	**0**	−38,139	−25,764	−12,663	−12,599	−12,739	−38,954	**2**	−38,291	−26,342	−12,660	−12,596	−12,755	−38,954
Em	**0**	−5,552	−7,016	−4,737	−4,733	−4,739	−6,103	**16**	−5,522	−7,407	−4,744	−4,742	−4,749	−6,101
Panel B: Ken French's value-weighted daily data														
Ind	**0**	−42,185	−28,172	−14,237	−14,114	−14,385	−42,547	**−37**	−42,623	−30,212	−14,260	−14,130	−14,448	−42,548
Size	**0**	−39,502	−32,967	−5,691	−5,613	−5,788	−40,361	**41**	−39,927	−33,646	−5,711	−5,629	−5,837	−40,361
Beme	**0**	−28,223	−25,347	−13,186	−13,091	−13,300	−28,360	**−89**	−28,351	−26,739	−13,199	−13,100	−13,343	−28,360
Ltr	**0**	−26,866	−23,334	−14,548	−14,459	−14,654	−27,080	**−70**	−26,988	−24,850	−14,557	−14,463	−14,693	−27,080
Mom	**0**	−28,652	−22,988	−18,947	−18,839	−19,071	−29,004	**−50**	−28,914	−24,451	−18,958	−18,847	−19,110	−29,004
Str	**0**	−25,402	−24,897	−17,025	−16,931	−17,132	−25,788	**−53**	−25,554	−26,331	−17,033	−16,938	−17,160	−25,788
Size-beme	**0**	−121,702	−28,712	−22,314	−21,935	−22,793	−124,178	**−59**	−122,407	−30,414	−22,329	−21,944	−22,859	−124,178
Size-mom	**0**	−58,310	−27,095	−27,413	−27,007	−27,906	−59,497	**−25**	−58,849	−29,930	−27,420	−27,008	−27,964	−59,501
Panel C: Ken French's equal-weighted daily data														
Ind	**0**	−37,754	−20,009	−13,310	−13,219	−13,424	−38,451	**13**	−37,906	−21,108	−13,330	−13,237	−13,468	−38,451
Size	**0**	−39,293	−27,067	−13,709	−13,631	−13,817	−40,228	**82**	−39,720	−27,664	−13,708	−13,631	−13,818	−40,228
Beme	**0**	−34,279	−21,588	−11,492	−11,431	−11,571	−35,053	**46**	−34,401	−22,279	−11,493	−11,430	−11,587	−35,053
Ltr	**0**	−38,886	−22,926	−12,927	−12,874	−13,000	−39,600	**21**	−39,085	−23,529	−12,932	−12,877	−13,015	−39,600
Mom	**0**	−38,347	−22,602	−13,344	−13,276	−13,426	−39,146	**13**	−38,498	−23,174	−13,348	−13,279	−13,448	−39,145
Str	**0**	−33,943	−24,636	−15,075	−14,999	−15,137	−35,785	**62**	−33,752	−24,405	−14,293	−14,220	−14,388	−35,785
Size-beme	**0**	−123,243	−25,907	−26,999	−26,626	−27,488	−125,908	**4**	−123,754	−27,468	−26,992	−26,614	−27,505	−125,908
Size-mom	**0**	−59,465	−25,612	−24,791	−24,433	−25,230	−60,987	**28**	−59,712	−28,144	−24,802	−24,441	−25,290	−60,988

Notes. This table displays the differences of log-likelihoods of out-of-sample daily excess returns from the log-likelihood of the robust distribution of BA-HVAR (Bayesian-averaging heterogeneous vector autoregression), for the robust and reference distributions used by BA-HVAR, B-VAR1 (first-order Bayesian vector autoregression), BA (Bayesian averaging), JOR (Jorion), KZ (Kan-Zhou), ROLL (rolling), and HIST (historical) methods. For each data set, the log-likelihood of BA-HVAR's robust distribution is subtracted from the log-likelihoods of the robust and reference distributions of other methods. Positive numbers indicate a higher log-likelihood than BA-HVAR's robust distribution, and negative numbers indicate a lower log-likelihood than BA-HVAR's robust distribution. We set risk aversion to one, and the robust methods all use the optimal value of model uncertainty when the possible values are $\Theta = \{0, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16\}$. The prior for BA-HVAR is the empirical prior, and the explanatory variables include a constant and average excess returns over the past day, week, month, year, and 10 years. The window sizes for the JOR, KZ, and ROLL methods are 1,260 days. Among the robust distributions, the highest log-likelihood for each data set is in boldface. Also, among the nonrobust distributions, the highest log-likelihood for each data set is in boldface. The log-likelihoods are for returns between 1977 and 2016.

Table 6. Certainty Equivalent Predictions

	Robust optimal portfolios			Nonrobust optimal portfolios		
	Predicted certainty equivalents		Realized perceived certainty equivalent	Predicted certainty equivalents		Realized perceived certainty equivalent
Portfolios	Robust distribution	Reference distribution		Robust distribution	Reference distribution	
			Panel A: CRSP daily data			
Beta	7.31	12.29	9.93	−4.23	16.47	−3.94
Size	8.87	18.26	16.13	−47.13	30.53	−33.62
Std	9.63	16.40	14.01	−30.00	25.10	−21.17
Em	4.50	5.86	6.72	3.82	6.30	7.74
			Panel B: Ken French's value-weighted daily data			
Ind	6.87	11.88	8.70	−1.07	15.47	1.10
Size	7.08	14.75	12.58	−43.22	26.71	−25.72
Beme	2.72	5.57	2.61	−1.50	7.56	−3.41
Ltr	2.64	5.44	2.89	−2.76	7.68	−1.23
Mom	2.75	5.59	3.32	−1.18	7.47	−9.72
Str	2.92	5.32	2.87	1.26	6.36	−4.45
Size-beme	11.76	21.51	15.86	−36.36	34.74	−27.55
Size-mom	11.67	24.07	18.61	−53.21	42.19	−60.70
			Panel C: Ken French's equal-weighted daily data			
Ind	11.34	19.26	16.64	−10.08	26.61	−3.59
Size	10.26	19.07	17.49	−34.38	31.45	−1.50
Beme	10.43	17.56	16.42	−3.74	23.16	−1.31
Ltr	10.22	17.09	15.12	−4.40	22.55	−0.39
Mom	10.84	18.24	15.80	−5.19	24.47	−1.87
Str	36.81	65.60	63.90	−109.91	110.24	98.73
Size-beme	15.11	26.81	22.72	−65.27	44.01	−43.71
Size-mom	13.98	28.90	23.36	−74.21	52.28	−54.47

Notes. This table display the predicted daily certainty equivalents of robust and nonrobust BA-HVAR portfolios, made by BA-HVAR's robust and reference distributions. The realized perceived certainty equivalents are also reported. For the portfolios chosen by robust BA-HVAR, columns two and three display the certainty equivalent predictions made by BA-HVAR's robust and reference distributions. Column four displays the realized perceived certainty equivalent (which also were reported in column two of Table 3). For the portfolios chosen by nonrobust BA-HVAR, columns five and six display the certainty equivalent predictions made by BA-HVAR's robust and reference distributions. Column seven displays the realized perceived certainty equivalent (which also was reported in column 9 of Table 3). The prior for BA-HVAR is the empirical prior, and the explanatory variables include a constant and average excess returns over the past day, week, month, year, and 10 years. The predictions are for all days between 1977 and 2016.

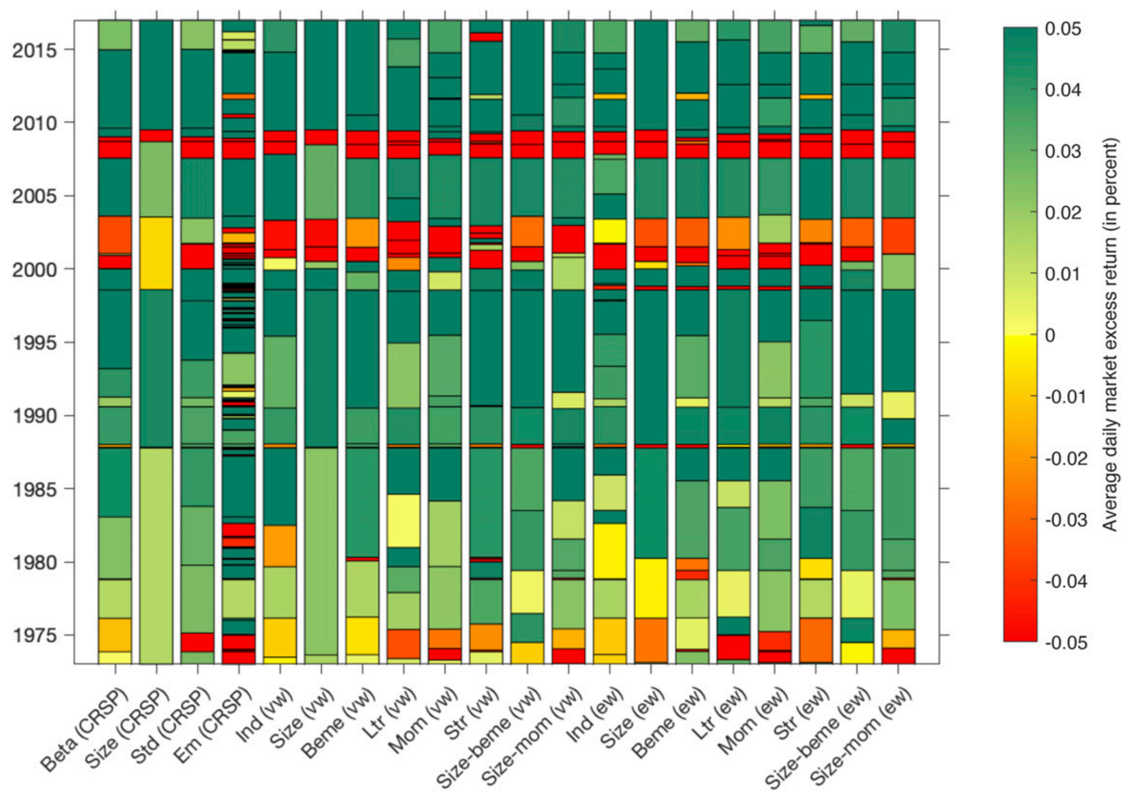
probabilities on this end date, because the probabilities on any subsequent date are about the real-time probabilities of the current regime, which ex post we believe is part of the last regime (not the next-to-last regime). The probabilities on any previous period use less information and are not as informative as the probabilities on this end date. After finding the beginning date of the next-to-last regime, we repeat this procedure to find earlier regimes.[24]

Figure 1 plots ex post beliefs about the most likely regimes in daily data using all sample information. Although the beginning and ending dates of regimes vary, there are many similarities in the 20 daily data sets. For example, 14 of 20 data sets indicate that there was exactly one regime that began in 1973, which we call the 1973 regime. The 1973 regime ended between 1974 and 1976 in 12 data sets and ended in 1987 in two data sets. Three additional data sets had more than one regime that started in 1973. In the other three data sets, no regimes began in 1973. Each of these three data sets was in the middle of a regime that ended in 1974.

All 20 daily data sets experienced short-term regimes that began just before Black Monday. (Black Monday occurred on October 19, 1987). The Black Monday regime started on October 14 in four data sets and on October 16 in 16 data sets. The Black Monday regime ended on October 20 in two data sets, on October 30 in nine data sets, in early to mid-November in five data sets, and in January 1988 in four data sets. During Black Monday regimes, asset returns were very low. For example, in regimes that lasted between October 16 and October 30, the *annualized* average daily value-weighted excess market return (as reported by CRSP) was about −98%.

During the rise in stock prices associated with the dot-com bubble, all 20 data sets had one or more new regimes. The starting dates for 18 of these dot-com regimes range from June to November 1998, and the ending dates, for 14 of the dot-com regimes, range from October 1999 to early January 2000. For example, in the value-weighted industry data set, the dot-com regime lasted between July 31, 1998, and November 26, 1999. During the dot-com regime, the high-tech sector earned a 69.7% annualized average daily excess return and had a market beta of 1.41. Before the bubble burst, there was an intermediate regime in some data sets. For example, between November 29, 1999, and October 3, 2000, there was a regime in the

Figure 1. Ex Post Most Likely Regimes and Market Returns



Notes. This figure displays the most likely regimes, as determined by BA-HVAR, in the United States between 1973 and 2016, for 20 daily data sets. The horizontal black lines separate regimes, and the colors indicate the average daily market excess return within regimes, using the color map on the right. Large positive returns are shaded green, large negative returns shaded red, and returns close to zero are shaded yellow. The datasets are identified by their short names listed in Table 1 and the abbreviations "vw" and "ew" stand for value- and equal-weighted data (respectively) from Ken French's website.

value-weighted industry data set, during which the high-tech sector earned a 10.6% annualized average daily excess return and had a market beta of 1.66. When the bubble burst, there were new regimes detected in most data sets. For example, when the bubble burst, the value-weighted industry data set was in a regime between October 4, 2000, and April 23, 2001, during which the high-tech sector had a −55.4% annualized average daily excess return and a market beta of 2.17.[25]

During the 2007–2008 financial crisis, all of the data sets indicate that there were one or more regimes that started between June and September 2008. In one data set, the crisis regime started in June, in three it started in July, and in 16 it started in September. Eight of the crisis regimes started on September 15, the day that Lehman Brothers collapsed. The end dates of the crisis regimes range from October 2008 to June 2009, with four ending between October and December, two ending in January, three ending in March, three ending in May, and eight ending in June.

Online Figure 1 displays real-time daily beliefs about regimes. On a majority of days, there was one regime with more than 50% probability and no other regimes with significant probability. Although, there were some days during which no regime had more than 50% probability and a few days with a large amount of regime uncertainty, during which many regimes had at least a 1% probability. There also were several false alarms where investors temporarily thought that they were in a new regime, only to later believe that a regime change never happened.

8.2. Number of Regimes

Table 7 displays the number of regimes in monthly, weekly, and daily data, according to BA-HVAR and Bayesian averaging (BA). We see that, in daily data, many more regimes are detected than in the monthly and weekly data. It is very possible that the additional daily regimes also affect monthly and weekly returns, but since monthly and weekly investors observe fewer data points, it is more difficult for them to discern regimes.

The BA method reports many more regimes than BA-HVAR. Although we do not know how many regimes there really are, we believe that BA is detecting too many regimes, for two reasons. One reason is that each of BA's models is too simple and not flexible enough to account for changing means. Each model in BA says that mean returns are constant over time, whereas each model in BA-HVAR allows means to vary linearly with recent returns. If the actual means returns really do vary linearly with past returns, then a single BA-HVAR model can capture this, whereas BA needs frequent regime changes to accommodate changes in means. Another reason why BA detects too many regimes is that older models transfer a large amount of probability to younger models each period (this is what Anderson and Cheng 2016 call the "sharing prior"), which causes investors to often believe in fictitious new regimes. For example, during the 2007–2008 financial crisis, BA (in the CRSP size data set) detected 29 different regimes, whereas BA-HVAR (also in the CRSP size data set) detected only one new regime. Unreported simulated results are consistent with BA finding too many regimes. We generate artificial daily data, using BA-HVAR's data-generating process (described in Section 2) and then estimate BA-HVAR and BA models. We find that BA-HVAR comes very close to predicting the correct number of regimes in every simulation, whereas BA finds about 40 times too many regimes. The multiplicative factor of 40 is roughly consistent with what we find with actual daily data in Table 7.

9. Predictions

In this section, we discuss the reasons for robust BA-HVAR's good performance. Robust BA-HVAR outperforms other methods because BA-HVAR predicts returns better than other methods and because robust optimizers choose portfolios with predictable certainty equivalents. We find that nonrobust portfolio choices perform poorly because they attempt to exploit deals that are too good to be true and that robust portfolio choices, by down-weighting seemingly good deals, choose portfolios with little downside risk. We begin by distinguishing reference and robust distributions. We then discuss return predictions and certainty equivalent predictions.

9.1. Reference and Robust Distributions

The reference distributions are the distributions used to make nonrobust and robust portfolio choices. The reference distribution used by BA-HVAR is the mixture of normals distribution in Equation (15b). The robust distributions are the endogenous alternative distributions that concern robust investors. When the reference distribution is a mixture of normals, the robust distribution is also a mixture of normals with altered means, variances, and model probabilities.[26]

Robust Distribution for a Mixture of Normals Reference Distribution. *Let the reference distribution be that excess returns have the mixture of normals distribution given in Equation (15b). The robust distribution is the mixture of normals distribution*

$$\overline{P}(\mathbf{r}_{t+1}|\mathcal{F}_t)\varrho_t^*(\mathbf{r}_{t+1}) = \sum_{m\in\mathcal{M}_{t+1}} \mathcal{N}\left(\boldsymbol{\mu}^*_{t+1|m,t}, \Sigma^*_{t+1|m,t}\right) \times P^*_{t+1}(m|\mathcal{F}_t)$$

Table 7. The Number of Regimes Detected

	BA-HVAR			BA		
Portfolios	Monthly data	Weekly data	Daily data	Monthly data	Weekly data	Daily data
			Panel A: CRSP daily data			
Beta	1	7	22	27	139	734
Size	1	3	8	25	114	689
Std	1	6	21	21	119	666
Em	43	72	402	189	834	3,984
			Panel B: Ken French's value-weighted daily data			
Ind	3	9	20	75	341	1,488
Size	1	4	12	24	113	638
Beme	3	9	18	50	211	1,003
Ltr	3	11	25	53	243	1,150
Mom	4	9	27	45	234	1,083
Str	3	8	33	51	218	1,144
Size-beme	3	7	17	21	91	427
Size-mom	3	8	26	26	140	668
			Panel C: Ken French's equal-weighted daily data			
Ind	3	12	34	59	279	1,315
Size	1	6	16	23	117	691
Beme	1	6	28	34	156	965
Ltr	1	8	24	31	144	858
Mom	3	9	30	31	150	843
Str	3	6	28	29	135	868
Size-beme	3	6	20	21	93	454
Size-mom	3	8	23	28	141	707

Notes. This table displays the number of regimes detected by the BA-HVAR (Bayesian-averaging heterogeneous vector autoregression) and BA (Bayesian averaging) methods in monthly, weekly, and daily data. The prior for BA-HVAR is the empirical prior, and the explanatory variables include a constant and average excess returns over the past day (only for daily data), week (only for daily and weekly data), month, year, and 10 years. The number of detected regimes between 1973 and 2016 are listed.

where formulas for the means $\boldsymbol{\mu}^*_{t+1|m,t}$, covariances $\Sigma^*_{t+1|m,t}$, and model probabilities $P^*_{t+1}(m|\mathcal{F}_t)$ are provided in Online Appendix E and depend upon the optimal robust portfolio choices, $\boldsymbol{\phi}_t$, model uncertainty, θ_t, and the reference distribution, $P_{t+1}(m|\mathcal{F}_t)$.

The robust means and variances reduce the attractiveness of good deals that are present in the reference distribution. It is computationally and intuitively useful to view the robust distribution as depending on the optimal robust portfolio choices. Although, since the optimal robust portfolio choices are determined by model uncertainty and the reference distribution, the robust distribution can alternatively be viewed as being entirely determined by model uncertainty and the reference distribution. To understand the roles of the reference and robust distributions, it is instructive to look at the relationship between the distributions and optimal portfolio choices. As described in Section 6.1, optimal nonrobust portfolio choices at time t are

$$\left(\frac{1}{\gamma}\right)\Sigma_{t+1|t}{}^{-1}\,\boldsymbol{\mu}_{t+1|t},$$

where $\boldsymbol{\mu}_{t+1|t}$ and $\Sigma_{t+1|t}$ are the means and covariances of asset returns in the reference distribution. Optimal robust portfolios solve the robust portfolio choice problem in Section 6.2 for the reference distribution. Optimal robust portfolios are mean-variance optimal when returns have the *robust* distribution and satisfy

$$\left(\frac{1}{\gamma}\right)\Sigma^*_{t+1|t}{}^{-1}\,\boldsymbol{\mu}^*_{t+1|t}, \tag{27}$$

where $\boldsymbol{\mu}^*_{t+1|t}$ and $\Sigma^*_{t+1|t}$ are the means and covariances of asset returns in the robust distribution.[27] Although robust investors worry that the assets have the robust distribution, they still believe that the reference distribution is the most likely distribution of asset returns. If a particular reference distribution correctly describes the world, then nonrobust portfolio choices that use that reference distribution should outperform other methods. Likewise, if a particular robust distribution correctly describes the world, then robust portfolio choices that use the corresponding reference distribution (and hence are concerned about this particular robust distribution) should outperform other methods.[28] However, in the more likely case that neither the reference nor the robust distribution correctly describes the world, we find that robust portfolio choices typically outperform nonrobust portfolio choices, even if out-of-sample returns have higher likelihood under the reference distribution than the robust distribution.

9.2. Likelihood of Returns

We compare the ability of reference and robust distributions to predict excess returns by examining the likelihood of returns. We find that BA-HVAR's reference distribution better predicts out-of-sample returns than other reference distributions, and that BA-HVAR's robust distribution better predicts returns than other robust distributions. Table 5 shows that the daily likelihood of returns is higher under BA-HVAR's reference distribution than other reference distributions on 19 of 20 data sets. Table 5 also shows that the daily likelihood of returns is higher under BA-HVAR's robust distribution than other robust distributions on 19 of 20 data sets. In results available upon request, we find that the weekly likelihood of returns is higher under BA-HVAR's reference distribution than other reference distributions on 18 of 20 data sets, and that the weekly likelihood of returns is higher under BA-HVAR's robust distribution than other robust distributions on 18 of 20 data sets. We also find that the monthly likelihood of returns is higher under BA-HVAR's reference distribution than other reference distributions on 17 of 20 data sets, and the monthly likelihood of returns is higher under BA-HVAR's robust distribution than other robust distributions on 17 of 20 data sets.

BA-HVAR's robust and reference likelihoods are very similar with the robust likelihood being higher for some data sets and the reference likelihood higher for other data sets. On most of the data sets, in which returns have a larger likelihood under the reference distribution, robust portfolio choices have a higher certainty equivalent than nonrobust portfolio choices. On all daily, weekly, and monthly data sets, in which returns have a larger likelihood under the robust distribution, robust portfolio choices have a higher certainty equivalent than nonrobust portfolio choices.

9.3. Certainty Equivalent Predictions

We show that the certainty equivalents of optimal robust portfolio choices are easier to predict than the certainty equivalents of optimal nonrobust portfolio choices.

The predicted certainty equivalent is the certainty equivalent that investors expect to receive from hypothetical portfolio choices. The predicted certainty equivalent uses the expectations of investors to forecast the one-period-ahead conditional means and variances of portfolios. The predicted certainty equivalent is similar to the perceived certainty equivalent, except that beliefs are used for all quantities, not just the conditional means that determine variances. Under the reference distribution, the predicted certainty equivalent is the sample mean

$$\bar{E}\left(\boldsymbol{\phi}_t'\boldsymbol{\mu}_{t+1|t} + r_{f,t+1} - \left(\frac{\gamma}{2}\right)\boldsymbol{\phi}_t'\Sigma_{t+1|t}\boldsymbol{\phi}_t\right),$$

where $\boldsymbol{\mu}_{t+1|t}$ and $\Sigma_{t+1|t}$ are the means and covariances of asset returns in the reference distribution, and ϕ_t are the hypothetical portfolio choices. Under the robust distribution, the predicted certainty equivalent is the sample mean

$$\bar{E}\left(\boldsymbol{\phi}_t'\boldsymbol{\mu}^*_{t+1|t} + r_{f,t+1} - \left(\frac{\gamma}{2}\right)\boldsymbol{\phi}_t'\Sigma^*_{t+1|t}\boldsymbol{\phi}_t\right),$$

where $\boldsymbol{\mu}^*_{t+1|t}$ and $\Sigma^*_{t+1|t}$ are the means and covariances of asset returns in the robust distribution.

Table 6 reports certainty equivalent predictions for optimal robust and nonrobust BA-HVAR portfolios. The most striking result is that the reference distributions predict very large (and positive) certainty equivalents for the optimal nonrobust portfolios on all 20 data sets, whereas the realized perceived certainty equivalents are *negative* on 17 of the 20 data sets.[29] The robust distribution does a better job of predicting the signs of the certainty equivalents for nonrobust portfolios and predicts negative certainty equivalents on 19 of 20 data sets. For example, on Ken French's value-weighted size and momentum data set, the reference distribution predicts that optimal nonrobust portfolio choices will have a daily certainty equivalent of 42%, but the realized perceived certainty equivalent is −60%. This is a daily prediction error of about 100%. For the same data set, the robust distribution predicts that nonrobust choices will have a certainty equivalent of −53%, for a prediction error of 7%. We also see that the robust distribution *under*predicts the certainty equivalent of optimal robust portfolio choices on 18 of 20 daily data sets, whereas the reference distributions *over*forecasts the certainty equivalent of optimal robust portfolio choices on 19 of 20 data sets. Both the robust and reference distributions can more accurately predict the certainty equivalents of optimal robust portfolio choices than optimal nonrobust portfolio choices.

9.4. Discussion

Nonrobust portfolio choices (for all methods) perform poorly because optimal nonrobust portfolios try to find the best deals present in estimates. This often involves taking highly leveraged positions that exploit imprecise estimates of mean returns and the correlation of returns. According to the reference distribution, these portfolios should achieve large certainty equivalents. In reality, these portfolios often earn very low certainty equivalents, sometimes lower than −50% per day. The reason is that these good deals are not actually present in the real world. Optimal robust portfolios, by taking a pessimistic view of the world, assume that any good deal is too good to be true, and endogenously downwardly adjust mean returns and alter correlations.

10. Conclusions

Most out-of-sample evaluations of portfolio strategies, such as DeMiguel et al. (2009) and Kan and Zhou (2007), focus on methods that assume means and covariances of stock returns are constant within a fixed window. Although more complicated approaches can perform much better in-sample, it is often thought that they perform poorly out-of-sample (Welch and Goyal 2008). This paper studies a portfolio choice strategy, with a large number of models and estimated coefficients, that allows mean returns to evolve over time, and it proposes a Bayesian-averaging heterogeneous vector autoregressive (BA-HVAR) portfolio strategy with many different vector autoregressions that differ on hyperparameters and on the amount of past data used to learn parameters. We show that robust BA-HVAR portfolio choices usually achieve higher certainty equivalents and Sharpe ratios than other portfolio choice strategies on 20 monthly, weekly, and daily data sets. The key reason for the excellent performance is that BA-HVAR gives data abundant flexibility to speak for itself, whereas our priors, though based only on available information, prevent unreasonable estimates that often cause other methods with a large number of parameters to fail out-of-sample.

In order to objectively evaluate the out-of-sample performance of portfolio choice strategies, it is essential that investors' choices do not depend on unknown information. The BA-HVAR portfolio choice strategy takes several steps to avoid using unavailable information, by choosing priors (and other parameters) based on a priori available knowledge. Unreported results show that the particular explanatory variables used are not that important and that BA-HVAR achieves excellent out-of-sample certainty equivalents with many different sets of explanatory variables. We generally find that investors can safely include as many explanatory variables as computationally possible and achieve high certainty equivalents and Sharpe ratios, because our priors on coefficients mitigate the possibility of spurious estimates contaminating out-of-sample performance. Using the sample means of past returns as explanatory variables is a computationally convenient way to capture the useful information from a large number of previous returns. It is not computationally feasible to include the same information in a conventional vector autoregression, because there would be too many explanatory variables.

This paper contributes to a large literature on detecting regime changes by proposing a tractable Bayesian approach that allows investors to form many new models each period, where each model takes a different stand on hyperparameters. In non-Bayesian setups, many alternative approaches have been developed to detect structural changes in a time series. Davis et al. (2006) provide a criterion that can be used to detect regimes and the order of an appropriate autoregressive process within each regime. Aue and Horváth (2013) discuss a procedure to detect structural changes in means and/or in covariances, by using cumulative sum statistics. It would be interesting to compare the ability of BA-HVAR and these methods to detect regimes.

This paper also contributes to the growing literature on robust portfolio choices by showing that robust versions of BA-HVAR and other methods almost always achieve higher out-of-sample certainty equivalents than their nonrobust counterparts. We suggest that robust portfolio choices do well because robust optimizers are able to find high-performing portfolios with predictable certainty equivalents. Nonrobust optimal portfolio choices perform poorly because they attempt to exploit deals that are too good to be true.

This paper studies Bayesian averaging specifications with simple explanatory variables. It would be interesting to investigate alternative explanatory variables with our approach, such as the covariates suggested by Gargano et al. (2019) and Cochrane (2007, 2011). All of the portfolio allocation strategies discussed in this paper are myopic and not dynamically optimal. Investors choose portfolios that maximize the expected value of the utility of next period's returns and ignore all other future consequences of today's decisions. Following many other prominent recent papers on out-of-sample performance, transaction costs and the advantages of dynamic hedging are ignored. It would be interesting to include both of these aspects of optimal dynamic decision making, but because of the complexity of our large number of models, such an undertaking is not currently computationally feasible.

Acknowledgments

The authors thank Ronald Gallant, Mohammad Jahan-Parvar, Hening Liu, and Ke Tang for helpful comments. The authors also thank seminar participants at the Econometric Conference on Big Data and AI in Honor of Ronald Gallant, held on December 21, 2018, in Beijing, China. Wharton Research Data Services (WRDS) and data from the Center for Research in Security Prices (CRSP) was used in this research. This service and the data available thereon constitute valuable intellectual

property and trade secrets of WRDS, CRSP, and/or its third-party suppliers.

Endnotes

[1] See Zweig (2007, p. 4).

[2] See Meucci (2005), Brandt (2010), and Guidolin and Rinaldi (2013) for surveys of portfolio choice methods.

[3] Many other papers have made important contributions using Bayesian methods. Winkler and Barry (1975) and Stambaugh (1997) made progress when there is one model; and Cremers (2002) and Tu and Zhou (2004) advanced the literature when there are many models. Black and Litterman (1992), Kandel and Stambaugh (1996), Pastor (2000), Pástor and Stambaugh (2000, 2002), Avramov (2004), and Tu and Zhou (2010) proposed various informative priors.

[4] Readers interested in a brief technical summary of BA-HVAR may wish to consult Online Appendix A.

[5] An alternative approach is to scale the constant term in $\mathbf{x}_{t-1}$ so that the same hyperparameters could be used for all elements of Φ.

[6] In our applications, $h = 9$, since there are nine possible combinations of the values for α_i and β_i.

[7] In general, we remove $\max(0, M_t + h - M)$ during period t, where M_t is the number of models in $\mathcal{M}_t$. In early periods, $M_t \leq (M - h)$ and no models are removed. After the burn-in period is over, $M_t = M$ and h models are removed. If M/h is not an integer, then before the burn-in period is over, there will be one period in which $(M - h) < M_t < M$ and fewer than h models are removed.

[8] The notation $\mathcal{T}(\boldsymbol{\mu}, S, \bar{\nu})$ denotes the multivariate t distribution with mean $\boldsymbol{\mu}$, scale matrix S, and degrees of freedom $\bar{\nu}$.

[9] See Online Appendix D for a discussion of multiperiod-ahead forecasts.

[10] If parameter uncertainty in coefficients (but not covariances) is included, then the reference distribution is a mixture of normals. The estimation error, in coefficients, adjusts the covariances in each model from $\Sigma_{t+1|m,t}$ to $V_{t+1|m,t}$. Solutions to the robust portfolio problem (described in later sections) are very similar to the solutions without uncertain coefficients (when portfolios are rebalanced every period).

[11] Relative entropy is not a formal measure of distance because it is not symmetric in the reference and alternative distributions.

[12] When $\theta_t = 0$, we interpret preferences as nonrobust with optimal portfolio choices given by Equation (17).

[13] See Online Appendix B for more details.

[14] Note that $\boldsymbol{\phi}_{s,\theta}$ is the solution to the robust/risk-sensitive problem in Equation (21), where θ_t is replaced with θ.

[15] The notation $\bar{E}^t_{s=b}\, y_s$ denotes the sample mean of y_s between periods b and t.

[16] As discussed in Section 7.1.1, we could compute certainty equivalents using conditional expectations for some quantities, rather than sample means. We obtain similar model uncertainty calibrations for both ways of computing historical certainty equivalents.

[17] Robust versions of the rolling, historical, Jorion, and Kan-Zhou methods make the additional assumption that excess returns are normally distributed. See Online Appendix C for a detailed review of the alternative methods.

[18] We use 20 of the 24 data sets studied by Anderson and Cheng (2016). Four of the data sets do not have enough data at the daily frequency, without gaps (missing values), and are not used in our study.

[19] Following DeMiguel et al. (2009), we set $\gamma = 1$ and use mean-variance preferences to compute certainty equivalents, regardless of the preferences of investors. Anderson and Cheng (2016, appendix C) argue that this is a sensible criterion for robust preferences. Both DeMiguel et al. (2009) and Anderson and Cheng (2016) use what we call the *standard certainty equivalent*. As we discuss below, the standard certainty equivalent is appropriate when the means and variances of portfolios are time-invariant.

[20] See Online Appendix F for further discussion and a proof.

[21] We compute standard errors of certainty equivalent differences using a stationary bootstrapping method (Politis and Romano 1994), with an average block size of five years. The standard errors are justified when the certainty equivalent differences between robust BA-HVAR and any other algorithm is stationary. Although we cannot guarantee that the differences are stationary, it is more plausible that the differences are stationary than that the certainty equivalents themselves are stationary.

[22] We compute standard errors of Sharpe ratio differences using a stationary bootstrapping method (Politis and Romano 1994), with an average block size of five years.

[23] The beginning date is the period after the birthdate with the highest probability, because models born at time t are designed to account for asset returns at time $t+1$ and later dates. To compute the model birthdate with the highest probability, we aggregate the probability of all models born on the same period.

[24] Online Appendix G formally describes the algorithm used to date regimes.

[25] For comparison, the market beta of the high-tech sector in regimes between 2007 and 2016 has been close to one.

[26] The reference distribution used by BA is also a mixture of normals distribution. The reference distributions used by the first-order Bayesian vector autoregression, Jorion, Kan-Zhou, rolling, and historical methods are normal and are described in Online Appendix C. The market and $1/N$ methods do not attempt to forecast returns and are not discussed in this section. When the reference distribution is normal, the robust distribution is normal. Online Appendix E provides formulas for the means and variances of the robust distribution, when the reference distribution is normal or a mixture of normals.

[27] The formula in Equation (27) is useful for interpretation but is not useful for computing robust portfolios. It is numerically convenient to first find the optimal robust portfolio choices and to then find the robust distribution.

[28] These statements follow from the first-order conditions for optimality.

[29] In this paragraph and in Table 6 we refer to the "perceived certainty equivalent" defined in Section 7, as the realized perceived certainty equivalent to distinguish it from the predicted certainty equivalent.

References

Anderson EW, Cheng A-R (2016) Robust Bayesian portfolio choices. *Rev. Financial Stud.* 29(5):1330–1375.

Anderson EW, Hansen LP, Sargent TJ (2003) A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *J. Eur. Econom. Assoc.* 1(1):68–123.

Aue A, Horváth L (2013) Structural breaks in time series. *J. Time Series Anal.* 34(1):1–16.

Avramov D (2002) Stock return predictability and model uncertainty. *J. Financial Econom.* 64(3):423–458.

Avramov D (2004) Stock return predictability and asset pricing models. *Rev. Financial Stud.* 17(3):699–738.

Avramov D, Zhou G (2010) Bayesian portfolio analysis. *Ann. Rev. Financial Econom.* 2:25–47.

Bawa VS, Brown SJ, Klein RW (1979) *Estimation Risk and Optimal Portfolio Choice* (North-Holland, Amsterdam).

Black F, Litterman R (1992) Global Portfolio Optimization. *Financial Analysts J.* 48(5):28–43.

Bollerslev T, Patton AJ, Quaedvlieg R (2016) Exploiting the errors: A simple approach for improved volatility forecasting. *J. Econometrics* 192:1–18.

Brandt MW (2010) Portfolio choice problems. Aït-Sahalia Y, Hansen LP, eds. *Handbook of Financial Econometrics*, vol. 1 (Elsevier, Amsterdam), 269–336.

Cochrane JH (2007) The dog that did not bark: A defense of return predictability. *Rev. Financial Stud.* 21(4):1533–1575.

Cochrane JH (2011) Presidential address: Discount rates. *J. Finance* 66(4):1047–1108.

Corsi F (2009) A simple approximate long-memory model of realized volatility. *J. Financial Econometrics* 7(2):174–196.

Cremers KJM (2002) Stock return predictability: A Bayesian model selection perspective. *Rev. Financial Stud.* 15(4):1223–1249.

Davis RA, Lee TCM, Rodriguez-Yam GA (2006) Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* 101(473):223–239.

Del Negro M, Schorfheide F (2011) Bayesian macroeconometrics. Koop G, Geweke J, van Dijk HK, eds. *The Oxford Handbook of Bayesian Econometrics* (Oxford University Press, Oxford, UK).

DeMiguel V, Garlappi L, Uppal R (2009) Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Rev. Financial Stud.* 22(5):1915–1953.

DeMiguel V, Nogales FJ, Uppal R (2014) Stock return serial dependence and out-of-sample portfolio performance. *Rev. Financ. Stud.* 27(4):1031–1073.

Dow J, Werlang S (1992) Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica* 60(1):197–204.

Epstein LG, Schneider M (2010) Ambiguity and asset markets. *Ann. Rev. Financial Econom.* 2:315–346.

Frost PA, Savarino JE (1986) An empirical Bayes approach to efficient portfolio selection. *J. Financial Quant. Anal.* 21(3):293–305.

Gargano A, Pettenuzzo D, Timmermann A (2019) Bond return predictability: Economic value and links to the macroeconomy. *Management Sci.* 65(2):508–540.

Garlappi L, Uppal R, Wang T (2007) Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Rev. Financial Stud.* 20(1):41–81.

Guidolin M, Rinaldi F (2013) Ambiguity in asset pricing and portfolio choice: A review of the literature. *Theory and Decision* 74:183–217.

Hansen LP, Sargent TJ (1995) Discounted linear exponential quadratic Gaussian control. *IEEE Trans. Automation Control* 40(5):968–971.

Hansen LP, Sargent TJ (2007a) Recursive robust estimation and control without commitment. *J. Econom. Theory* 136(1):1–27.

Hansen LP, Sargent TJ (2007b) *Robustness* (Princeton University Press, Princeton, NJ).

Hansen LP, Sargent TJ, Tallarini TD Jr (1999) Robust permanent income and pricing.*Rev. Econom. Stud.* 66(4):873–907.

Jorion P (1986) Bayes-Stein estimation for portfolio analysis. *J. Financial Quant. Anal.* 21(3):279–292.

Kan R, Zhou G (2007) Optimal portfolio choice with parameter uncertainty. *J. Financial Quant. Anal.* 42(3):621–656.

Kandel S, Stambaugh RF (1996) On the predictability of stock returns: An asset-allocation perspective. *J. Finance* 51(2):385–424.

Litterman RB (1986) Forecasting with Bayesian vector autoregressions: Five years of experience. *J. Bus. Econom. Statist.* 4(1):25–38.

Liu H (2010) Robust consumption and portfolio choice for time varying investment opportunities. *Ann. Finance* 6(4):435–454.

Maenhout PJ (2004) Robust portfolio rules and asset pricing. *Rev. Financial Stud.* 17(4):951–983.

Maenhout PJ (2006) Robust portfolio rules and detection error probabilities for a mean-reverting risk premium. *J. Econom. Theory* 128(1):136–163.

Markowitz H (1952) Portfolio selection. *J. Finance* 7(1):77–91.

Meucci A (2005) *Risk and Asset Allocation* (Springer, New York).

Pastor L (2000) Portfolio selection and asset pricing models. *J. Finance* 55(1):179–223.

Pástor Ľ, Stambaugh RF (2000) Comparing asset pricing models: An investment perspective. *J. Financial Econom.* 56(3):335–381.

Pástor Ľ, Stambaugh RF (2002) Mutual fund performance and seemingly unrelated assets. *J. Financial Econom.* 63(3):315–349.

Pastor L, Veronesi P (2009) Learning in financial markets. *Ann. Rev. Financial Econom.* 1:361–381.

Politis DN, Romano JP (1994) The stationary bootstrap. *J. Amer. Statist. Assoc.* 89(428):1303–1313.

Sims CA, Zha T (1998) Bayesian methods for dynamic multivariate models. *Internat. Econom. Rev.* 39(4):949–968.

Stambaugh RF (1997) Analyzing investments whose histories differ in length. *J. Financial Econom.* 45(3):285–331.

Tiao GC, Zellner A (1964) On the Bayesian estimation of multivariate regression. *J. Royal Statist. Soc. B* 26(2):277–285.

Tu J (2010) Is regime switching in stock returns important in portfolio decisions? *Management Sci.* 56(7):1198–1215.

Tu J, Zhou G (2004) Data-generating process uncertainty: What difference does it make in portfolio decisions? *J. Financial Econom.* 72(2):385–421.

Tu J, Zhou G (2010) Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. *J. Financial Quant. Anal.* 45(4):959–986.

Uppal R, Wang T (2003) Model misspecification and underdiversification. *J. Finance* 58(6):2465–2486.

Welch I, Goyal A (2008) A comprehensive look at the empirical performance of equity premium prediction. *Rev. Financial Stud.* 21(4):1455–1508.

Winkler RL, Barry CB (1975) A Bayesian model for portfolio selection and revision. *J. Finance* 30(1):179–192.

Zellner A (1971) *An Introduction to Bayesian Inference in Econometrics* (Wiley, Hoboken, NJ).

Zellner A, Chetty VK (1965) Prediction and decision problems in regression models from the Bayesian point of view. *J. Amer. Statist. Assoc.* 60(310):608–616.

Zweig J (2007) *Your Money and Your Brain: How the New Science of Neuroeconomics Can Help Make You Rich* (Simon and Schuster, New York).