OPTIMAL ETF SELECTION FOR PASSIVE INVESTING

DAVID PUELZ, CARLOS M. CARVALHO AND P. RICHARD HAHN

ABSTRACT. This paper considers the problem of isolating a small number of exchange traded funds (ETFs) that suffice to capture the fundamental dimensions of variation in U.S. financial markets. First, the data is fit to a vector-valued Bayesian regression model, which is a matrix-variate generalization of the well known stochastic search variable selection (SSVS) of George and McCulloch (1993). ETF selection is then performed using the "decoupled shrinkage and selection" procedure described in Hahn and Carvalho (2015), adapted in two ways: to the vector-response setting and to incorporate stochastic covariates. The selected set of ETFs is obtained under a number of different penalty and modeling choices. Optimal portfolios are constructed from selected ETFs by maximizing the Sharpe ratio posterior mean, and they are compared to the (unknown) optimal portfolio based on the full Bayesian model. We compare our selection results to popular ETF advisor Wealthfront.com. Additionally, we consider selecting ETFs by modeling a large set of mutual funds.

Keywords: benchmarking; dimension reduction; exchange traded funds; factor models; personal finance; variable selection.

1. INTRODUCTION

Exchange traded funds (ETFs) have emerged in recent years as a low-fee way for individuals to invest in the stock market. The growth of ETF popularity the past 20 years stemmed from investors' desire to participate passively in the returns of stocks in the overall market. The first ETF began trading in January 1993 and was called the S&P 500 Depository Receipt, also known as SPDR. Since then, the size of the ETF market has grown to over $1 trillion, and SPDR, a company derived from State Street Global Advisors, is the world's second largest ETF provider with assets of nearly $350 billion. ETF investing spans a large variety of asset classes, with funds holding currencies, foreign equity, bonds, real estate, and commodities. The explosive growth of the ETF industry underscores the desire of the average investor to hold a diversified and broadly exposed portfolio for a cheap fee.

Although there are many fewer ETFs than there are generic tradable assets, an individual investor who has decided to invest entirely in ETFs still has decisions to make. Should one hold a variety of specialty ETFs, such as funds with a real-estate focus or a biotech focus? Or is it adequate to hold a single broad-spectrum "market" fund, such as the Russell 4000, which holds positions in thousands of individual stocks? In this paper, we perform variable selection on ETFs to reduce the options a (long-term) investor faces to just a handful of distinct funds.

To determine a small subset of ETFs most suitable for individual investing, our strategy will be to isolate those ETFs which capture the vast majority of variability in the stock market. Specifically, our analysis focuses on eight "financial anomalies" from the asset pricing literature. These assets are themselves formed as linear combinations of many individual stocks (according to an established recipe which involves pre-sorting the stocks by various criteria). Our working premise is that these eight assets represent a desirable cross-section of market risk for investors to be exposed to, see Fama and French (1992) and Fama and French (2015). Granting this premise, an investor still cannot invest in these eight factors directly for practical reasons; trading costs from thousands of buy-sell transactions prohibit this strategy (although several mutual fund providers such as Dimensional Fund Advisors sell products that attempt to mimic these theoretical strategies). For this reason, we refer to our response vector as the "unattainable or target assets." With this as background, our goal is simply to find a small number of ETFs that replicate the covariance structure of the unattainable assets to a reasonable practical tolerance. Once these ETFs are selected, various portfolio optimization strategies can be implemented. We compare the performance of these portfolios to the inferred performance of the (unknown) optimal portfolio implied by our statistical model.

Methodologically, our analysis combines and extends two previous techniques. First, we extend the decision-theoretic variable selection (DSS) approach of Hahn and Carvalho (2015) to the vector-valued response setting. The DSS approach consists of two phases, a model fitting phase and a variable selection phase. In the model-fitting phase, we adapt and extend the stochastic search variable selection (SSVS) (George and McCulloch, 1993; Brown and Vannucci, 1998) for a vector-valued response. This model differs from a naive application of SSVS to a vector-valued response in that variable inclusion is determined simultaneously across the individual univariate regressions; that is, the variable either appears in all of the regressions or none of them. Also, in the selection phase, we consider a stochastic design matrix; Hahn and Carvalho (2015) consider only a fixed-design utility function which draws a natural connection to model selection in gaussian graphical models as we will seek to explore the conditional independence relationships between the ETFs and the target assets (Jones et al., 2005; Wang et al., 2011; Wang, 2015). This modification is important in the context of investing, because the future returns of the ETFs are unknown at the time of selection.

1.1. **Previous ETF research.** Recent research has focused on evaluating ETFs as single investments. Poterba and Shoven (2002) examine the operation of ETFs from a tax efficiency perspective. They conclude that ETFs are more tax efficient than equity mutual funds by noting that taxable gains on ETFs are smaller than comparable mutual funds, suggesting they are a reasonable low-cost investment for taxable investors. Agapova (2011) compares passive, index tracking mutual funds and ETFs. She examines fund flows using a pooled OLS model and finds that ETFs are almost perfect substitutes for passive mutual funds. DiLellio and Jakob (2011)

look at whether published ETF trading strategies outperform the market. They found many strategies outperform the S&P 500 but with weak statistical significance. Several other papers study investment characteristics of ETFs, including Huang and Lin (2011), Shin and Soydemir (2010), Pennathur et al. (2002), Ackert and Tian (2008) and Kostovetsky (2005). We contribute to this diverse body of research by proposing an investment methodology for the average investor using ETFs as the sole financial product.

The construction of passive portfolios is a separate area of the literature but also relevant to our research. These problems are typically framed in a variable selection framework in which regularization and optimization become important tools. Index tracking is one approach to forming such a portfolio. This is done by determining which subset of index components can be invested in while maintaining similar performance to the index - commonly known as index tracking. Rockafellar and Uryasev (2002) present a conditional value-at-risk (CVaR) constrained optimization and apply it to tracking the S&P 100 index. Fastrich et al. (2013) consider penalized optimization to construct sparse optimal portfolios. They review empirical performance of portfolios built from several penalties, including l_q-regularizers, and develop a new penalty that leads to high Sharpe ratio (risk-adjusted return) tracking portfolios. In two separate papers, Wu et al. (2014) consider the special cases of l_1 and l_2 penalties in their optimization, known as the *lasso* and *elastic net*, respectively (Wu et al., 2014; Wu and Yang, 2014). They develop algorithms to solve a nonnegative optimization problem where the decision variables are the long-only weights on assets in a tracking portfolio. In other words, they do not allow the short-selling of an asset. A Nonnegative Irrepresentable (NIR) condition is also shown to guarantee variable selection consistency. Since picking a subset of countably many assets is the goal, exploring all possible combinations of assets can be undertaken with mixed-integer programming (MIP). Canakgoz and Beasley (2009) develop an MIP approach to index tracking as well as enhanced indexation, where the objective is to outperform the index. This approach includes transaction costs and linearization of the tracking portfolio returns for computational tractability. Similarly, Chen and Kwon (2012) consider a robust MIP formulation by incorporating estimation error into the objective quantities. They develop a fast algorithm by maximizing pairwise similarities between assets in the tracking portfolio and target index. Beasley et al. (2003) develop an evolutionary heuristic for the index tracking problem which incorporates transaction costs. They consider a minimization problem involving the tracking error and excess return and discuss in sample and out of sample performance.

Uncertainty is central to the problem at hand. It rears its head through parameters in the statistical model we specify and the asset returns we use for estimation. Jacquier and Polson (2010) review Bayesian tools in finance used to deal with uncertainty. They discuss a framework for evaluating predictive distributions of unknown returns and parameters and how one deals with financial quantities, such as the Sharpe ratio, from a Bayesian perspective. Pastor and Veronesi

(2009) survey recent literature focused on "learning in financial markets." The executive summary: acknowledging parameter uncertainty leads to easier interpretation of common models used in finance. This is the mantra of our approach. We take into account the unknown future by integrating over parameter and return uncertainty before the selection of ETFs is made.

2. AN ETF FACTOR MODEL OF MARKET COVARIATION

Our analysis revolves around eight financial "anomalies" from the finance literature, which go by the names: size, value, market, direct profitability, investment, short-term reversal, long-term reversal, and momentum. Each anomaly is a portfolio, constructed by cross-sectionally sorting stocks by various characteristics of a company and forming linear combinations based on these sorts. For example, the value anomaly is constructed using the book-to-market (the value of a company "on paper" divided by the market's perception of its value) ratio of a company. A high ratio indicates the company's stock is a "value stock" while a low ratio leads to a "growth stock" assessment. Essentially, the value anomaly is a portfolio built by going long stocks with high book-to-market ratio and shorting stocks with low book-to-market ratio. For detailed definitions of the first five factors, see Fama and French (2015). The data we use in our analysis was obtained from Ken French's website[1]. It is widely believed that these eight anomaly portfolios (or some subset of them) reflect all dimensions of independent variation in the stock market as reflected in Fama and French (1992) and Fama and French (2015).

Although these strategies cannot be readily implemented by the average investor, we might hope to determine a set of ETFs that recapitulates their covariance structure. To this end, we consider the 25 most highly traded (i.e., most liquid) equity funds from ETFdb.com. For more recent data periods, we are able to increase our number of ETFs to 46. Specifically, we use monthly ETF data from the Center for Research in Security Prices (CRSP) database from February 1992 through February 2015 (CRSP, 2015).

In the next section, we lay out a regression model that ties these attainable assets (ETFs, which are easy to invest in) to the target assets (the eight anomalies, with returns we may readily observe, but not readily invest in).

2.1. **The regression model.** Arbitrage pricing theory (APT) (Ross, 1976) expresses expected returns as a linear combination of systematic factors and sensitivity parameters:

$$(2.1) \qquad \mathbb{E}[R_j] = r_f + \beta_{1j}F_1 + \cdots + \beta_{pj}F_p,$$

where R_j denotes returns and F_p represent (possibly unobservable) sources of undiversifiable risk — the unavoidable risk inherent to putting one's money in the market. The theory derives its name from its assumption that all asset prices divergent from the model will be corrected by arbitrage.

---

[1]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

Our model will assume that we we may define our systematic factors in terms of ETFs. That is, given a set of target asset returns, $\{R_j\}_{j=1}^q$, and ETFs, $\{X_i\}_{i=1}^p$, we model the target returns as:

$$(2.2) \qquad R_j = \beta_{j1} X_1 + \cdots + \beta_{jp} X_p + \epsilon_j, \quad \epsilon_j \sim \mathrm{N}(0, \sigma^2).$$

We argue that it is reasonable to fix the ETFs as factors in an APT model because there are many such funds and that they trade across multiple asset classes and markets. In this formulation, the right-hand side of model 2.1 represents the set of assets attainable for the average investor. The left-hand side are unattainable but desirable assets — the target assets. The linear model provides a mapping between the attainable and unattainable spaces that can be rigorously studied. The remaining challenge is to determine a small number of ETFs that simultaneously well-approximate all eight of the target returns.

For T periods, our linear model can be expressed compactly as a matrix normal distribution (Dawid, 1981). Define the matrix of target assets as $\mathbf{R} \in \mathbb{R}^{T \times q}$ and the matrix of ETFs as $\mathbf{X} \in \mathbb{R}^{T \times p}$. Additionally, let $\gamma \in \{0,1\}^p$ be a binary vector identifying a particular ETF model where the nonzero entries specify which ETFs are included. We write the model M_γ as:

$$(2.3) \qquad M_\gamma: \quad \mathbf{R} \sim \mathrm{Matrix\ Normal}_{T,q}\left(\mathbf{X}_\gamma \boldsymbol{\beta}_\gamma,\ \sigma^2 \mathbb{I}_{T \times T},\ \mathbb{I}_{q \times q}\right).$$

Note the row and column covariances are diagonal by the APT assumptions.

3. UTILITY-BASED ETF SELECTION

Our analysis adapts the model selection approach described in Hahn and Carvalho (2015), who cast model selection as a means to an end. They argue that if the goal is selection of a small subset of covariates, that desire should be reflected in a utility function rewarding sparsity (rather than via a prior). They propose a *DSS loss function (decoupled shrinkage and selection)* derived by integrating over the predictive and posterior distributions from standard model space sampling as in George and McCulloch (1993). Once integrating over posterior uncertainty, this loss function is used for covariate selection. Assuming the design matrix and prediction points are identical and given by $\mathbf{X}$, the DSS loss function is:

$$(3.1) \qquad \mathscr{L}(\gamma) = T^{-1} \|\mathbf{X}\bar{\beta} - \mathbf{X}\gamma\|_2^2 + \lambda \|\gamma\|_0,$$

where $\bar{\beta}$ is the posterior mean and γ is the choice variable. The loss function elegantly depends only on the posterior mean, $\bar{\beta}$. Approximating the penalty term in 3.1 with an L1-norm, the selection step amounts to solving the predictive loss minimization problem.

$$(3.2) \qquad \beta_\lambda := \operatorname*{arg\,min}_\gamma T^{-1} \|\mathbf{X}\bar{\beta} - \mathbf{X}\gamma\|_2^2 + \lambda \|\gamma\|_1,$$

where β_λ is sparse since the objective function is penalized. Thus, the nonzero elements of β_λ determine which covariates are selected. Hahn and Carvalho discuss approaches for choosing a tuning parameter λ along the solution path. We use the two step approach of this paradigm in our analysis, outlined as:

(1) Model fitting step: Modeling the marginal ETF distribution and sampling the conditional model space via Bayesian conditioning,

(2) Selection step: Integrate over posterior uncertainty and determine a sparse selection of covariates.

Many approaches can be used for the selection step including lasso optimization as in 3.2 or naive forward stepwise selection. Regardless of the approach, concerns of overfitting are sidestepped by working with a denoised target, $\mathbf{X}\bar{\beta}$. Given this pre-smoothed response, it is natural to think of the selection step as "fitting the fit." McCulloch (2015).

3.1. **Model fitting: The marginal and conditional distributions.** The future returns of the target assets and ETFs are unknown. Acknowledging this uncertainty is important in the overall decision of which ETFs to select. In fact, it is necessary for an honest ex ante selection of a subset of these assets. We account for this by modeling the marginal distribution of the ETFs (denoted by the matrix X) via a latent factor model. The target assets are modeled conditionally via the APT model, and this procedure is described in the next subsection. Using the compositional representation of the joint distribution:

$$(3.3) \qquad\qquad p(x, r) = p(r|x) p(x).$$

We specify the following model for the joint distribution:

$$(3.4) \qquad\qquad \begin{bmatrix} \mathbf{R} \\ \mathbf{X} \end{bmatrix} \sim N(\mu, \Sigma),$$

where Σ has a block covariance structure:

$$(3.5) \qquad\qquad \Sigma = \left[\begin{array}{c|c} \beta^T \Sigma_x \beta + \Psi & (\Sigma_x \beta)^T \\ \hline \Sigma_x \beta & \Sigma_x \end{array} \right].$$

Notice that the upper right block is the marginal variance of Y implied by the APT model. The lower right block is simply the marginal variance of X we must additionally model.

We obtain posterior samples of Σ by sampling the APT model parameters using a matrix-variate stochastic search algorithm (described below) and sampling the covariance of X from a latent factor model where it is marginally normally distributed. To reiterate our procedure is

- Σ_x is sampled from independent latent factor model,

- β is sampled from matrix-variate MCMC,
- Ψ is sampled from matrix-variate MCMC.

3.1.1. *Modeling the marginal distribution: A latent factor model.* We model ETFs via a latent factor model of the form:

$$
\begin{aligned}
\mathbf{X}_t &= \mu_x + \mathbf{B}\mathbf{f}_t + \mathbf{v}_t \\
\mathbf{v}_t &\sim \mathrm{N}(0, \Psi) \\
\mathbf{f}_t &\sim \mathrm{N}(0, \mathbf{I}_k), \\
\mu_x &\sim \mathrm{N}(0, \Phi)
\end{aligned}
$$

(3.6)

where Ψ is assumed diagonal and the set of k latent factors f_t are independent. The covariance of the ETFs is constrained by the factor decomposition and takes the form:

$$
\Sigma_x = \mathbf{B}\mathbf{B}^T + \Psi.
$$

(3.7)

To estimate this model, we use the R package `bfa` from Jared Murray (Murray, 2015). The software allows us to sample the marginal covariance as well as the marginal mean via a simple Gibbs step assuming a normal prior on μ_x.

3.1.2. *Modeling the conditional distribution: A Matrix-variate stochastic search.* We model the conditional distribution, $\mathbf{R}|\mathbf{X}$, by developing a novel variable selection algorithm and sample parameters via Bayesian conditioning. Recall that the conditional model is of the form 2.3, and we aim to explore the posterior on the model space, $\mathbf{P}\left(M_\gamma \mid \mathbf{R}\right)$. This is a generalization of stochastic search variable selection from George and McCulloch (1993) in that our response is vector-valued instead of a single random variable. Thus, the observed target asset data, $\mathbf{R}$, is a matrix.

Similar to George and McCulloch (1993), our algorithm explores the model space by calculating a Bayes factor for a paticular model M_γ. Given that the response $\mathbf{R}$ is matrix instead of a vector, we derive the Bayes factor as a product of vector response Bayes factors. This is done by separating the marginal likelihood of the target assets as a product of distinct vector response marginal likelihoods for each of the target assets separately. This derivation requires our priors to be independent across the target assets and is shown in the appendix. Our approach is novel precisely because we have adapted SSVS to a matrix-variate response. Note that we do not run standard SSVS on each target asset regression separately. Instead, we generalize George and McCulloch (1993) and require all covariates to be included or excluded from a model for all of the target assets *simultaneaously.*

The marginal likelihood requires priors for the parameters β and σ parameters in our model. We use the well known g-prior that is standard for linear models because it permits an analytical solution for the marginal likelihood integral (Zellner, 1986; Zellner and Siow, 1984; Liang et al., 2008a).

Our Gibbs sampling algorithm follows the standard stochastic search variable selection directly. The aim is to scan through all possible covariates and determine which ones to include in the model, and this is how the algorithm explores the model space. At each substep of the MCMC where we are looking at an individual covariate within a specific model, we compute the probability of covariate's inclusion as a function of the model's prior probability and the Bayes factors:

$$p_i = \frac{B_{a0}\mathbf{P}\left(M_{\gamma_a}\right)}{B_{a0}\mathbf{P}\left(M_{\gamma_a}\right) + B_{b0}\mathbf{P}\left(M_{\gamma_b}\right)}.$$

The prior on the model space, $\mathbf{P}\left(M_\gamma\right)$, can either be chosen to adjust for multiplicity or uniform - our results are robust to both specifications. In this setting, adjusting for multiplicity amounts to putting equal prior mass on different sizes of models. In contrast, the uniform prior for models involving p covariates puts higher probability mass on larger models, reaching a maximum $\binom{p}{2}$. The details of the priors on the model space and parameters, including an empirical Bayes choice of the g-prior hyperparameter, are discussed in the appendix.

Using this algorithm, we visit the most likely ETF factor models given our matrix of target assets. Under the model and prior specification, there are closed-form expressions for the posteriors of the model parameters β_γ and σ. Thus, we can easily sample any functional of these parameters, including the implied tangency portfolio returns and Sharpe ratio. Discussed in the empirical findings section, these metrics are useful in analyzing the selected ETF portfolios.

3.2. **Derivation of the conditional loss function.** Our goal is now to describe the relationship between the ETFs and the unattainable assets. While the parameters of our model do precisely this, we have only posterior samples of these parameters (not a simple point estimate) and, moreover, these parameters are potentially "larger" than we would like, in the sense that they involve all possible ETFs while perhaps a much smaller number accounts for the vast majority of the covariance structure of the target assets.

In order to find a parsimonious summary we consider a loss function motivated by the conditional distribution of $\mathbf{R}$ given $\mathbf{X}$. In particular, this likelihood takes the form:

$$(3.8) \qquad\qquad\qquad r|x \sim N(\gamma x, D^{-1}),$$

so the log-likelihood is:

$$(3.9) \qquad \log \det(D) - \frac{1}{2} \left(r^T D r - 2 x^T \gamma^T D r + x^T \gamma^T D \gamma x \right).$$

Using this as our loss function, we might ask for an "action" γ that summarizes our distribution; we consider D to be fixed. As we are not using this likelihood in a statistical capacity, we actually would like our γ summary to characterize future realizations $\tilde{R}$ and $\tilde{X}$. Because these future realizations are naturally unavailable to use, we cannot maximize (3.8) over $\tilde{R}$ and $\tilde{X}$. Instead, we first take expectations, yielding:

$$(3.10) \qquad \frac{1}{2} \mathrm{tr}[D \Sigma_r] - \frac{1}{2} \mathrm{tr}[\gamma^T D \gamma \Sigma_x] - \frac{1}{2} \mu_x^T \gamma^T D \gamma \mu_x + \mathrm{tr}[\gamma^T D \beta \Sigma_x] + \mu_x^T \gamma^T D \mu_r.$$

Define the integrated conditional loss function, $\mathscr{L}(\gamma, \Sigma, \mu_x, \mu_y)$, by dropping all terms that do not involve our choice variable, γ:

$$(3.11) \qquad \mathscr{L}(\gamma, \Sigma, \mu_x, \mu_r) = -\frac{1}{2} \mathrm{tr}[\gamma^T D \gamma \Sigma_x] - \frac{1}{2} \mu_x^T \gamma^T D \gamma \mu_x + \mathrm{tr}[\gamma^T D \beta \Sigma_x] + \mu_x^T \gamma^T D \mu_r.$$

Of course, the parameters appearing in this expression are also not known exactly, so we integrate once more over the posterior distribution of $\{\Sigma_x, \beta, \mu_x, \mu_y\}$:

$$(3.12) \qquad \mathscr{L}(\gamma) = -\frac{1}{2} \mathrm{tr}\left[D \gamma \left(\overline{\Sigma_x} + \Sigma_{\mu_x} + \overline{\mu_x}\, \overline{\mu_x}^T \right) \gamma^T \right] + \mathrm{tr}\left[D \left(\overline{\beta \Sigma_x} + \Sigma_{\mu_x \mu_r} + \overline{\mu_r}\, \overline{\mu_x}^T \right) \gamma^T \right].$$

The overlines are used to represent the posterior means of the model parameters. Defining $H = \overline{\Sigma_x} + \Sigma_{\mu_x} + \overline{\mu_x}\, \overline{\mu_x}^T$, $f = \overline{\beta \Sigma_x} + \Sigma_{\mu_x \mu_r} + \overline{\mu_r}\, \overline{\mu_x}^T$, and $H = L L^T$, we have:

$$(3.13) \qquad \begin{aligned} \mathscr{L}(\gamma) &= -\frac{1}{2} \mathrm{tr}\left(D \left[\gamma H \gamma^T - 2 f \gamma^T \right] \right) \\ &\propto -\frac{1}{2} \mathrm{tr}\left(D \left[(\gamma - f H^{-1}) H (\gamma - f H^{-1})^T \right] \right) \\ &= -\frac{1}{2} \mathrm{tr}\left((\tilde{\gamma} L - D^{\frac{1}{2}} f L^{-1})^T (\tilde{\gamma} L - D^{\frac{1}{2}} f L^{-1}) \right) \\ &= -\frac{1}{2} \mathbf{vec}\left(\tilde{\gamma} L - D^{\frac{1}{2}} f L^{-1} \right)^T \mathbf{vec}\left(\tilde{\gamma} L - D^{\frac{1}{2}} f L^{-1} \right). \end{aligned}$$

In 3.13, we complete the square with respect to γ and disregard constant terms that don't involve this action. We also redefine the action as $\tilde{\gamma} = D^{\frac{1}{2}} \gamma$. Finally, we convert the trace to an l_2 norm and distribute the vectorization operation across the expression using the Kroeneker product (where $\mathbb{I}$ is an identity matrix the same dimension as D):

$$(3.14) \qquad \mathscr{L}(\tilde{\gamma}) = -\frac{1}{2} \left\| \left[[L^T \otimes \mathbb{I}] \mathbf{vec}(\tilde{\gamma}) - \mathbf{vec}(D^{\frac{1}{2}} f L^{-1}) \right] \right\|_2^2 + \lambda \left\| \mathbf{vec}(\tilde{\gamma}) \right\|_1.$$

We include an l_1 penalty with parameter λ, which encourages the optimization solution to be sparse. This is emphasized in Hahn and Carvalho's DSS paper where l_1-regularization is accompanied with integration over uncertainty for model selection.

Expression 4.1 is now in the form of standard sparse regression loss functions (Tibshirani, 1996), with covariates L, "data" $D^{\frac{1}{2}}fL^{-1}$, and regression coefficients $\tilde{\gamma}$. Accordingly we may optimize (4.1) conveniently using existing software, such as the `lars` package of Efron et al. (2004).

Choice of the penalty parameter is a necessary practical concern. In this matter, we also follow the pragmatic Bayesian approach of Hahn and Carvalho (2015), who advocate choosing λ by scrutinizing plots that reflect the predictive deterioration attributable to λ-induced sparsification. Crucially, such plots convey posterior uncertainty in the chosen performance metric, allowing for intuitive criteria to be expressed along the lines of: chose λ such that, with posterior probability greater than 95%, the predictive error of the sparse predictor is no more than 10% worse than that of the unsparsified optimal prediction." In our application, we will use the log conditional distribution as our measure of predictive performance, which is simply our utility function without the sparsity penalty.

3.2.1. *Difference from lasso and original DSS.* The loss function 4.1 is distinct from original DSS loss function from Hahn and Carvalho (2015) in two important ways. First, its derivation relies on a statistical model represented in the compositional form of conditional and marginal distributions, as opposed to a standard linear regression with normal i.i.d. errors. Second, the loss metric is not explicitly squared error. Instead, our notion of accuracy is defined by the negative log-likelihood of the conditional distribution of the target assets given the ETFs. Our approach is different from the group lasso of Yuan and Lin (2006) (where grouped covariates enter the model simultaneously along the lasso solution path) for these same reasons Yuan and Lin (2006).

In the appendix, Hahn and Carvalho (2015) discusses covariance estimation in the context of a graphical lasso (Friedman et al., 2008) loss function, denoted as the "DSS graphical model posterior summary optimization problem." The goal is to find a parsimonious posterior summary of the covariance using the graphical DSS loss function. Our approach is similar, but our loss function is only focused on the off-diagonal block of the covariance quantifying the dependence between the target assets and the ETFs. That is, instead of considering the joint distribution, we focus on the implied conditional distribution. Unlike the graphical lasso optimization where all covariance components are penalized choice variables, our method allows for optimization only over coefficient matrix in the conditional distribution which represents the dependence between the ETFs and target assets. This difference is made explicit by a simple example in the appendix.

An important feature of our loss function are the posterior means of cross products of parameters (such as $\overline{\beta \Sigma_x}$) that appear in our formulation. They can have quite different distributions than the individual parameters as they are naturally dependent a posteriori via the model. It is interesting and notable that these moments, quantifying the relationship between across such parameters, appear in our final loss function.

4. EMPIRICAL FINDINGS

We now apply our model sampling and selection algorithm to ETF and financial anomaly data from February 1992 to February 2015. Using the parameters sampled in the matrix-variate MCMC, we calculate the value of our conditional loss function along the solution path of the lasso optimization and for several MCMC iterations at each solution. Recall that the loss function, written in "lasso form," is:

$$(4.1) \qquad \mathscr{L}(\tilde{\gamma}) = -\frac{1}{2} \left\| \left[[L^T \otimes \mathbb{I}]\mathbf{vec}(\tilde{\gamma}) - \mathbf{vec}(D^{\frac{1}{2}} f L^{-1}) \right] \right\|_2^2 + \lambda \left\| \mathbf{vec}(\tilde{\gamma}) \right\|_1.$$

Figure 4.1 shows the evaluations of this loss function for different values of $\tilde{\gamma}$ (and thus differing amounts of sparsity) and quantiles surrounding these evaluations. Note that model size is measured as the number of connections in the graph between the ETFs and target assets. As model size gets larger, more ETFs connect to the target assets (more components of the conditional dependence matrix between the ETFs and target assets are nonzero) and the value of the conditional loss function, as derived by the log-likelihood, increases. Every point along the "model fit" line should be thought of as a graph representing the dependence between the ETFs and target assets.

The loss function value plateaus at the dense model fit, ie: when all possible edges between the ETFs and target assets sampled in our Gibbs algorithm are included. The 40th to 60th quantile band of the dense model fit is shown in gray rectangle. We remove the scale on the y-axis since we only need to compare the model fit relative to the dense model fit to make a graph selection. Our heuristic for ETF selection is to choose the sparsest model on the solution such that the posterior mean of its fit is contained in the dense model quantile band. This selection heuristic is key to our approach and can only be done if uncertainty intervals about the model fit our known. Thus, a Bayesian fitting of our model provides these quantiles through which a selection can be made. The model can be made sparser or denser (more or fewer edge connections between the ETFs and target assets) by varying the size of the quantile band. This is a qualitative judgement to be made by the "ETF selector." However, since the first, model sampling, step tends to decrease the number of pertinent ETFs through exploration of only relevant models, we have found that a given ETF graph is relatively robust to changing the dense model fit quantiles.

Figure 4.2 shows the selected ETFs and their connection to the eight financial anomalies. In the lasso optimization, we unpenalize the connection between SPY and the market factor (Mkt.RF) since an investor intuitively desires to hold at least "the market." SPY appears in all models along the solution path as a result of this unpenalization. Note that IWM is connected to four of the eight anomalies. Given that IWM is a small blend ETF, its connection to the SMB (small minus big, or size) factor is intuitive. It is also connected to the LTR (long term reversal), HML (high minus low, or value), and RMW (robust minus weak, or profitability) factors. Similar to STR (short-term reversal), LTR is a trading strategy that buys stocks that have had a below average long term trend of returns and sell stocks that have had an above average trend. The intuition is that over time an outperforming (underperforming) stock will correct its above (below) average performance and systematically "trend reverse." STR and LTR capture the premium derived from this strategy over different length correction periods. IWM's connection to LTR suggests that small blend companies are exposed to the variation in LTR. Additionally, its connection to HML and RMW indicates that some of the companies in IWM may be trading below their book value (ie: are "value stocks") and that profitability is driving their return variation.

IWO is an ETF comprised of small growth companies. It is connected to the Mom (Momentum), CMA (conservative minus aggressive), LTR, RMW, and HML factors. The Momentum factor invests in stocks that have had sustained outperformance that is expected persist. IWO's connection to Mom is intuitive as the expansion of growing companies and their returns tend to be persistence through market cycles. Its connection CMA is equally appealing. CMA is a factor investing in companies who invest conservatively and selling companies who invest aggressively. Small companies that are growing quickly are intimately involved in investment; whether its increasing to fuel future growth or curbed to keep revenues high. Therefore, IWO must be tied to a factor capturing market variation of companies focused on investment.

The final ETF in our selected portfolio, IWV, is a blend of large market capitalization stocks. Since this ETF contains companies and are large and relatively mature, it is a good substitute for a "market-like" ETF. Nonetheless, it is included in our selection along with SPY which tracks the S&P500. As such, IWV and SPY are the only ETFs that are connected with the market factor. Additionally, IWV is connected to STR, suggesting that large cap companies trend reverse or correct over short periods of time. This makes sense when compared to IWM's connection to the LTR factor. In general, larger companies such as Apple are more closely followed by the media and public compared to smaller companies. Therefore, corrections to a large stock's above or below average performance should happen faster than a small stock.

4.1. **Benchmarking specific allocations.** In practice, individual investors want not only to know which funds to invest in, but also how much to invest in each. Portfolio optimization and its vast
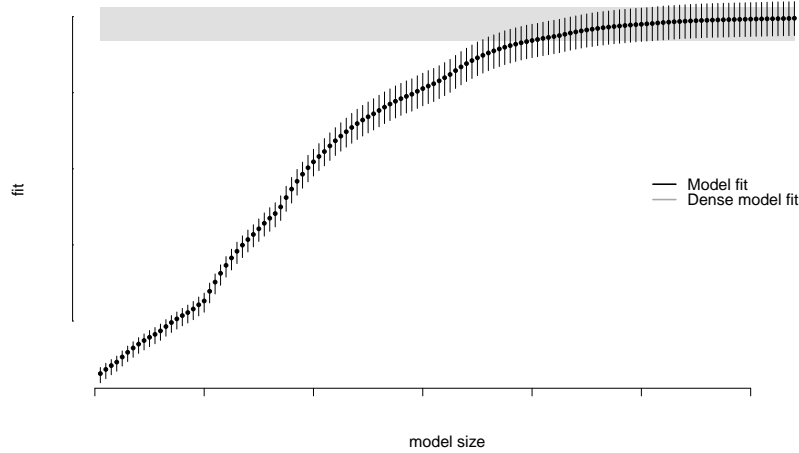
FIGURE 4.1. Model fits as measured by the conditional loss function. Allows for the selection of ETFs. Model size refers to number of edges in graph.

literature is beyond the scope of this paper. However, in this section we undertake an optimization approach that is Bayesian, intuitive, and simple. We construct the selected ETF portfolio by maximizing the posterior mean of its Sharpe ratio. The Sharpe ratio is a common financial metric characterizing the risk-adjusted return of an asset. Acknowledging the widely used and reasonable assumptions that investors like high returns and dislike risk (are risk adverse), the Sharpe ratio divides the first two moments of an asset's return, i.e.: A higher Sharpe ratio indicates more return per unit of standard deviation (risk). It was first mentioned in a paper written by nobel laureate William Sharpe in which he called it the "reward-to-variability ratio" (Sharpe, 1966).

The space of weights is explored using a differential evolution optimization within the R package `DEoptimR` of Conceicao and Maechler (2015). We are able to constrain the weights to sum to 100% and restrict them to be positive (no short selling). This is a reasonable constraint for a layman investor and one that we are able to easily enforce. The maximum posterior mean Sharpe ratio portfolio is 96% market and large-cap ETFs with a 4% tilt towards a small-cap ETF. In sum, this portfolio includes ETFs that capture dominant sources of variation in the eight financial anomalies, and is allocated so that it risk-adjusted return is maximum. It is reasonable to expect that market and large-cap ETFs will have large allocations in our portfolios. These ETFs trade stocks that represent a large part of total market-cap of the U.S. financial markets. The tilt towards small-cap and no allocation to growth suggests that we can increase our Sharpe ratio further with additional exposure to variation that drives the small minus big (SMB) factor.

Crucially, what makes this orthodox Bayesian approach to allocation convenient and appealing was the initial reduction of the problem, first from all assets to only ETFs, and then, with our
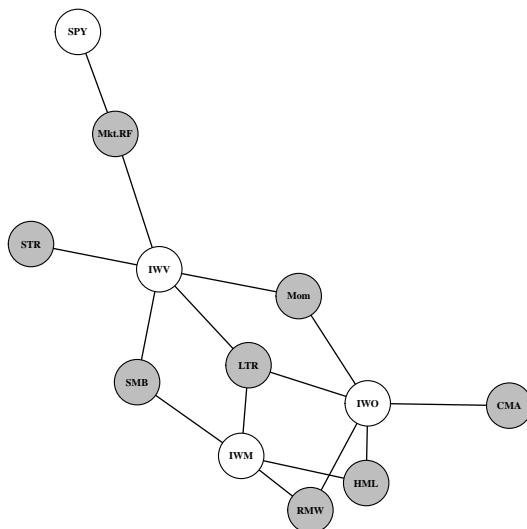
FIGURE 4.2. Selected ETFs and their edge connections to the unattainable assets.

ETF	SPY	IWM	IWO	IWV
weight	59.3 %	4.0 %	0.0 %	36.7 %
style	*market*	*small blend*	*small growth*	*large blend*

TABLE 4.1. February 1992 - February 2015: Selected ETF portfolio constructed by maximizing its Sharpe ratio's posterior mean (SPY is forced to be included through the lasso penalty).

contribution, to just a small subset of ETFs. Therefore, it is natural to ask how much the variable reduced approach gives up to the analogous approach which forgoes the ETF selection step. We see this comparison in Figure 4.3, which shows the distribution of sampled Sharpe ratios for the selected ETF portfolio and the full ETF porfolio (the maximum posterior mean Sharpe ratio portfolio of all 25 ETFs). We see that the distribution of Sharpe ratios for our reduced portfolio sits slightly below that of the distribution of Sharpe ratios of the portfolio investing in all ETFs (as one would expect). However, the two distributions overlap substantially and the price of an extremely parsimonious investing strategy is quite small.

Additionally, we consider the distribution of the Sharpe ratios corresponding to the model implied optimal portfolio if one could invest (both long and short positions) in the target assets themselves. This benchmark is unattainable in two ways. First, our analysis was predicated on the idea that the left-hand-side assets could not be directly invested in. More importantly,
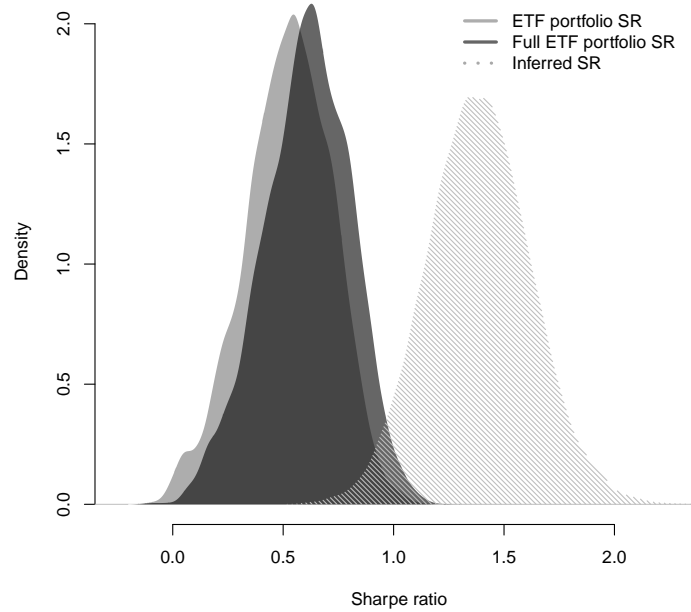
FIGURE 4.3. February 1992 - February 2015: Sampled Sharpe ratios for different portfolios

the Sharpe ratios obtained are from an optimal model that is allowed to change iteration-by-iteration during posterior sampling. That is, we are looking at the distribution of the performance of the optimal portfolio and not the performance of the optimal portfolio in expectation; it is not a single portfolio we are looking at, but many conditionally optimal ones. So, while the inferred SR optimal portfolio cannot be achieved, it does provide a natural scale for our comparisons. In figure 4.4, we show the same sampled Sharpe ratios with the inferred distribution removed.
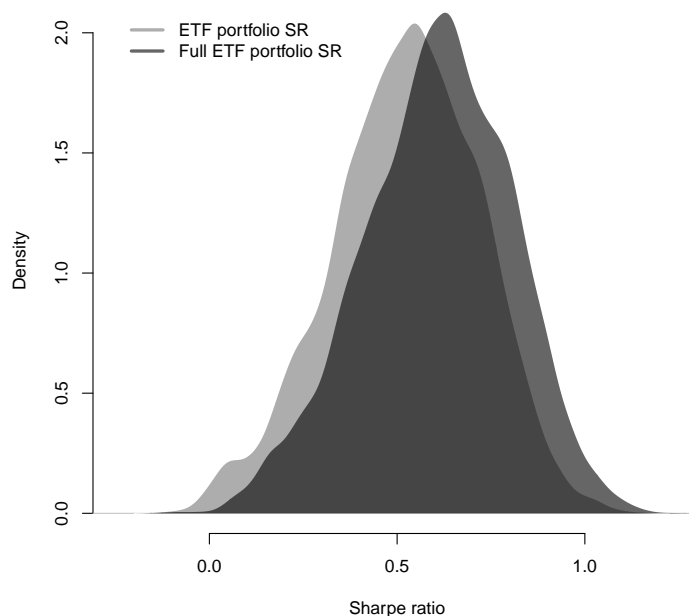
FIGURE 4.4. February 1992 - February 2015: Sampled Sharpe ratios for different portfolios

Lastly, we consider the case where the market ETF (SPY) is treated equal to any other ETF, and so can be dropped from the select set. The resulting ETF graph is shown in figure 4.5. This graph is the same as figure 4.2 except that SPY falls out due to penalization. The maximum posterior mean Sharpe ratio portfolio the contains only one ETF, IWV. This is a broad market ETF tracking the Russell 3000 index. After integrating over uncertainty in future returns and parameters, the chosen ETF portfolio is simply a market-like ETF. This provides rigorous intuition for the folk wisdom "just buy the market." After taking all unknowns into account, this result suggests that a broad portfolio of large stocks is not too bad.

FIGURE 4.5. Selected ETFs and their edge connections to the unattainable assets.

ETF	IWV
weight	100 %
style	*large blend*

TABLE 4.2. February 1992 - February 2015: Selected ETF portfolio constructed by maximizing its Sharpe ratio's posterior mean (SPY is not forced to be included through the lasso penalty).

Although confirming one widely help position, this result raises the question as to why there are other camps who espouse alternative investing advice. In our next section, we see that non-stationarity can explain some of this discrepancy.

4.2. **Rolling analysis.** Our analysis so far, built on an APT model, has assumed stationarity of the returns process, meaning that the distribution of returns are stable across time. It is natural to question this assumption. To investigate the possibility that the returns vary in time, and to examine how this might impact our ETF selection method, in this section we apply our approach separately to overlapping time periods.. We show overlapping 10-year periods from March 1995 through February 2015 in figure 4.2. More ETFs become available in the different time periods and we allow the algorithm to consider the larger set. The first time period has 25 ETFs and the final time period has 46 ETFs. Each ETF set is a subset of the future time periods' data. For the first time period (March 1995 - February 2005), we see value appear in IWD

and IVE as well as small-cap in IJR. Otherwise, the usual large-stock blend appears in IWV and small-stock blends in IWO and IWM. Notice how momentum is detached from the broad market ETFs. This suggests that momentum-based trading had isolated covariation among only a couple ETFs. Just before the beginning of this time period, Jegadeesh and Titman (1993) published their famous paper on the momentum strategy and a formalization of those ideas were in their infancy. Also, note that the small growth ETF, IWO, is connected to the market factor during this time period. March 1995 to February 2005 includes the tech boom and burst where many small technology companies grew at enormous rates. This ultimate bubble caused significant variation in the markets, and we see this manifested through IWO's connection to the market factor.
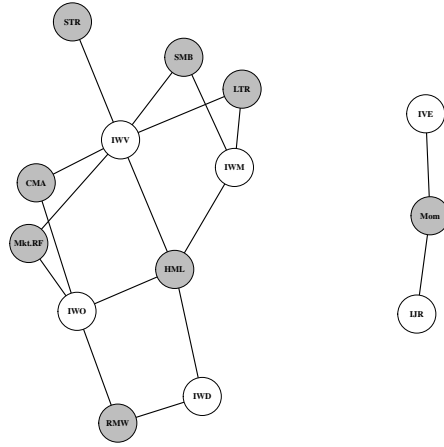
(a) March 1995 - February 2005
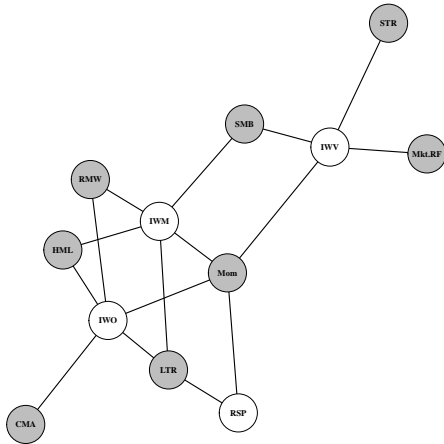
(b) March 2000 - February 2010
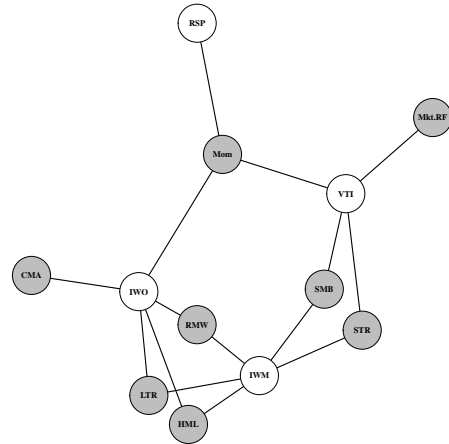
(c) March 2005 - February 2015

The two more recent time periods are similar. Both have three in common (IWO, IWM, RSP) and each has a broad market ETF (IWV for March 2000 - February 2010 and VTI for March 2005 - February 2015). The factors' connections to the ETFs do change through the periods, highlighting the reasonable fact that covariation within a given ETF may be driven by different factors over time. Specifically, we see that IWO loses its connection to the market factor and IWM gains connections to RMW and Mom. Also, the momentum factor joins the larger graph with connections to all selected ETFs suggesting that its variation is matured and exists in the

broader market. The S&P500 equal weight ETF, RSP, enters through the momentum and long-term reversal factors during the second time period.

One interesting comparison is the March 2005 - February 2015 portfolio (table 4.3) and the portfolio formed over the longer time period, February 1992 - February 2015, in table 4.2. The former results in a portfolio of a single ETF of a blend of large stocks. In the latter, there is a roughly even split between broad market and market equal-weight ETFs. This latter portfolio is tilted toward smaller stocks giving the equal weighting of RSP and is due to the shorter and more recent data used.

ETF	VTI	IWM	IWO	RSP
weight	43.7 %	0.0 %	0.0 %	56.3 %
style	*broad market*	*small blend*	*small growth*	*market equal-weight*

TABLE 4.3. March 2005 - February 2015: Selected ETF portfolio constructed by maximizing its Sharpe ratio's posterior mean

4.3. **Further Applications.**

4.3.1. *Comparison to Wealthfront.* Over the shorter data period (March 2005 - February 2015), we are able to compare our selected portfolio to Wealthfront.com, an ETF investing firm that currently has $2.6bn in assets under management.
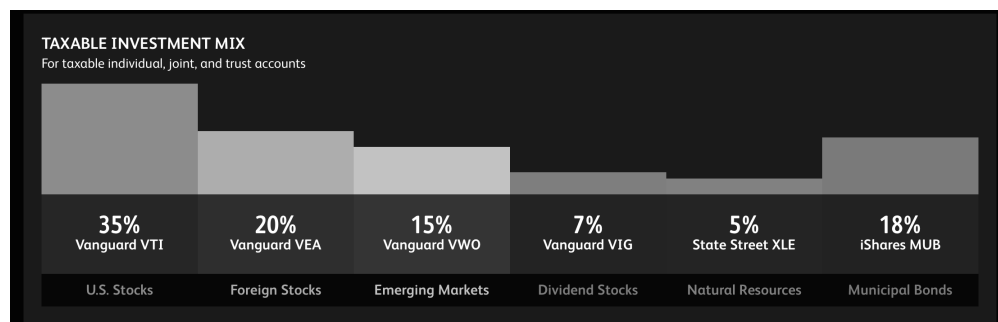

FIGURE 4.6. Source: wealthfront.com

In the past five years as ETFs have become increasingly popular, many other firms and products similar to Wealthfront's have emerged, including: Betterment, Charles Schwab Intelligent Portfolios, WiseBanyan, and LearnVest. Each company is marketed to the average investor. Given a level of risk tolerance determined by a series of questions answered by the investor, each company will generate a corresponding portfolio of ETFs. Figures 4.6 shows an example allocation from Wealthfront's website. Note the presence of only ETFs in each portfolio; these are portfolios of this one passive instrument. The growth of these companies is driven by the desire of the layman investor to not just index invest, but to index invest in an optimal way. Even though these companies make a reasonable first pass at providing a solution - to our

ETF	VWO	VEA	VTI	VIG	XLE
weight	18.3 %	24.4 %	42.7 %	8.6 %	6 %
style	*EM*	*Non-US*	*market*	*dividend*	*energy*

TABLE 4.4. Wealthfront portfolio.

knowledge, their product falls short in two key areas. (1) Their selection of ETFs is based on a qualitative analysis of expense ratios, liquidity, general market popularity, and predefined asset class buckets. (2) The optimal weights are calculated from traditional and potentially unstable mean-variance optimization. In this paper, we attempt to provide an improvement to step (1).
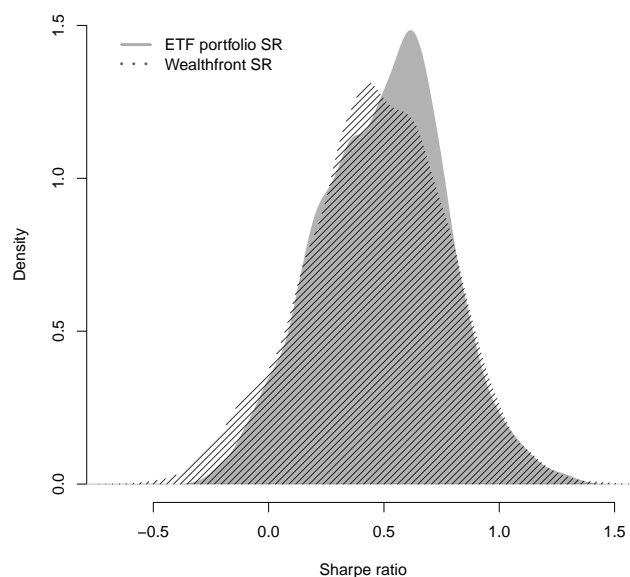

FIGURE 4.7. March 2005 - February 2015: Sampled Sharpe ratios for Wealthfront portfolio and our proposed ETF portfolio given in table 4.3.

We display the weights of the equity-only portfolio given by Wealthfront in table 4.4. In figure 4.7, we display the sampled Sharpe ratios. Both portfolios have similar upside potential, but note that the Wealthfront portfolios left tail is substantially larger than the ETF portfolio.

4.3.2. *Mutual funds as target assets.* As a final exercise, we consider the case of having mutual funds as target assets. We randomly sample 100 mutual funds from the CRSP Survivor-Bias-Free US Mutual Fund database and use these as our response matrix in our algorithm and our data is over the longer time period, February 1992 - February 2015. The solution path allowing us to select the appropriate model is shown in figure 4.8, and the selected graph is displayed in figure 4.9.

The selected graph shows the mutual funds that are connected to the chosen ETFs, and the result is quite remarkable. The three chosen ETFs have strategies precisely linked to the Fama and French three factors from their well known paper Fama and French (1992). This suggests that the covariation among mutual fund returns is largely encompassed by variation in the market, size, and value factors. However, the direction of causation is unknown. Either these three factors represent the true dimensions of the financial market, or mutual fund managers believe these are the dimensions and trade as such.
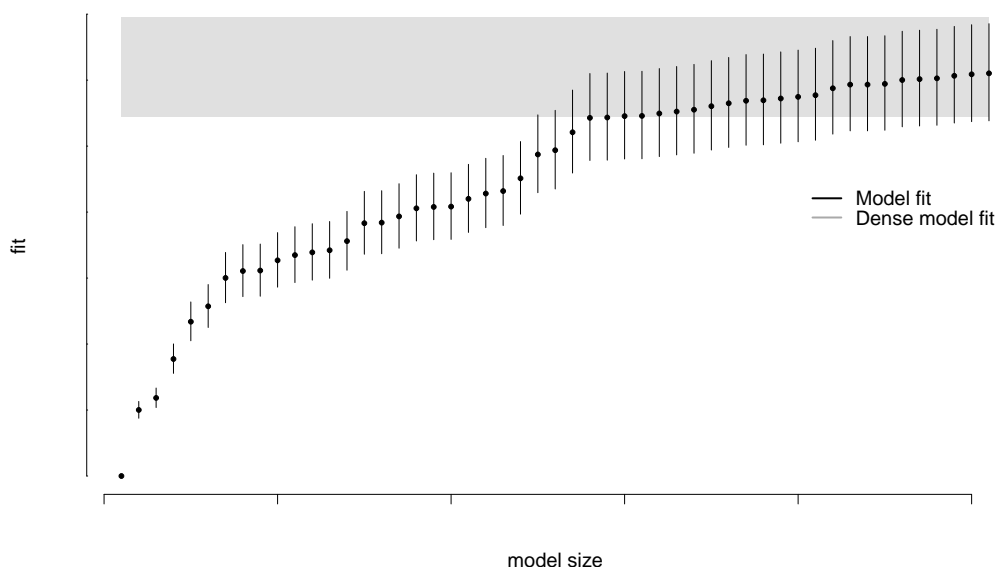

FIGURE 4.8. Model fits as measured by the conditional loss function. Allows for the selection of ETFs. Model size refers to number of edges in graph.

This example emphasizes the broader value and applicability of our algorithm. In today's world, there are thousands of mutual funds and ETFs one could invest in, and an investor is quickly overwhelmed with bank research, Morningstar ratings, and qualitative advice on which small subset of investments she should care about. Using modern technology in Bayesian estimation and regularized optimization combined with economic theory in the APT, our selection algorithm is able to sparsify this massive set of investment options for the average investor.
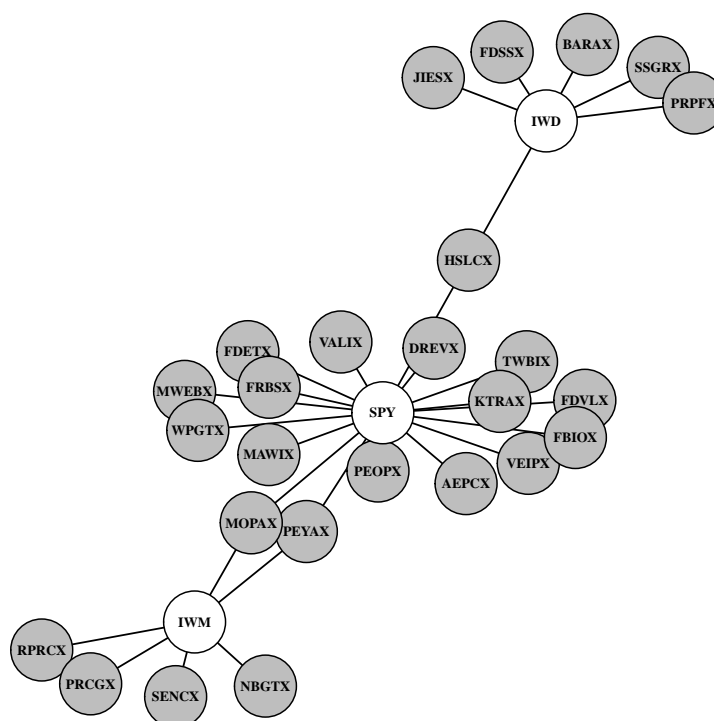
FIGURE 4.9. Selected ETFs and their edge connections to the set of mutual funds. Singleton mutual funds (with no edges) are not shown for clarity.

5. Conclusion

The investment universe is complicated. With the rise of multi-billion dollar money managers, active and passive mutual funds spanning every conceivable asset class, and several varieties of financial products and hedging instruments, it is challenging for the average investor to find the best place to invest. A common answer to this investment question is: "Just hold the market."

However, this simple solution might not be so simple to implement. How does one define the market? Is it all public equity? If so, it is impossible to hold each of these assets. The Standard and Poor's 500 index (S&P 500) comprised of the top 500 largest companies traded on the major exchanges might be a decent proxy for the United States equity market, but does it capture all of the market variation? Further, is there a way to identify which premia derived from this variation are most important? Answers to these complicated questions would greatly benefit the average investor.

An investment that gives broad exposure to many asset classes for low fees is the exchange traded fund (ETF). Are ETFs a working solution to the "just hold the market" dilemma? Perhaps. As the ETF universe expands into new financial markets and asset classes, there is no doubt that the average investor is exposed to more market variation than ever before. In fact, ETFs have introduced a secondary but important wrinkle in the investing decision: in which ETFs do I invest, and how much should I invest in each? The investment problem becomes a choice of ETFs and portfolio allocation problem. This index investing dilemma has led to the development of two notable firms in the past five years [2]: Betterment and Wealthfront. Each company is marketed to the average investor. Given a level of risk tolerance determined by a series of questions answered by the investor, each company will generate a corresponding portfolio of ETFs.

In this paper, we proposed a methodology to address these shortfalls of index investing. We formulate, from the investor perspective, the ETF choice problem as a Bayesian model selection problem where we select ETFs that most closely replicate a chosen set of target assets. We lean on the theoretical underpinnings of stochastic search variable selection from George and McCulloch (1993) and arbitrage pricing theory of Ross (1976) to develop our method. We then couple our statistical analysis with a practical variable selection approach based on decision theory, the end result being a handful of ETFs from which to build a portfolio. Crucially, our analysis does not stop there. We may continue to leverage the insights of our fully Bayesian statistical analysis to benchmark various portfolio allocations (among the selected ETFs) on the basis of widely-used criteria, such as the Sharpe ratio.

An important point to remember when considering asset allocation is that not only are future returns uncertain, but the distributions of those future returns are likewise uncertain. Additionally, while we might want to compare our ETF portfolio to the optimal portfolio, this optimal

[2]Others include Charles Schwab Intelligent Portfolios, WiseBanyan, and LearnVest.

portfolio is itself unknown (in addition to being impracticable). Fortunately, our Bayesian analysis permits us to compare the performance of any candidate portfolio to the unknown optimal portfolio, while accounting for all of these many sources of uncertainty. We make these comparisons manageable by first undertaking a principled variable selection step.

The upshot of our analysis is both expected and surprising. On the one hand, we find that, up to statistical uncertainty, our chosen ETF portfolios contain only a couple of broad-spectrum "market" ETFs. That is, as far as our data inform us, the most sensible ETF portfolios to hold are largely exposed to the market. Moreover, the broad market index the algorithm chooses is not SPY, but IWV: an ETF composed of large-cap stocks. We also find that our selected ETF portfolios have similar Sharpe ratio profiles to an unreasonable alternative of investing in all available ETFs. Over rolling time periods, our analysis routinely chooses a more diverse set of ETFs indicating that one should adjust their portfolios as markets change. Indeed, our analysis tilts heavily towards small-cap and value funds, with a dash of equal-weighting through RSP tied to the momentum factor. This finding lends statistical credence to prevailing folk-wisdom in investing circles.

APPENDIX A. MATRIX-VARIATE STOCHASTIC SEARCH

For model comparison, we calculate the Bayes factor with respect to the null model without any covariates. First, we calculate a marginal likelihood. This likelihood is obtained by integrating 2.3 over $\boldsymbol{\beta}_\gamma$ and σ multiplied by a prior for these parameters. A Bayes factor of a given model γ versus the null model, $B_{\gamma 0} = \frac{m_\gamma(\mathbf{R})}{m_0(\mathbf{R})}$ with:

$$(A.1) \qquad m_\gamma(\mathbf{R}) = \int \mathrm{MN}_{T,q}\left(\mathbf{R} \mid \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma, \ \sigma^2 \mathbb{I}_{T \times T}, \ \mathbb{I}_{q \times q}\right) \pi_\gamma\left(\boldsymbol{\beta}_\gamma, \sigma\right) d\boldsymbol{\beta}_\gamma d\sigma.$$

From the APT assumption, we have that the columns of $\mathbf{R}$ are independent. Additionally, we assume independence of the priors across columns of $\mathbf{R}$ so we can write the integrand in A.1 as a product across each individual target asset:

$$m_\gamma(\mathbf{R}) = \int \Pi_{i=1}^q \ \mathrm{N}_T\left(\mathbf{R}^i \mid \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma^i, \ \sigma^2 \mathbb{I}_{T \times T}\right) \pi_\gamma^i\left(\boldsymbol{\beta}_\gamma^i, \sigma\right) d\boldsymbol{\beta}_\gamma^i d\sigma$$

$$\Longleftrightarrow$$

$$m_\gamma(\mathbf{R}) = \int \ \mathrm{N}_T\left(\mathbf{R}^1 \mid \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma^1, \ \sigma^2 \mathbb{I}_{T \times T}\right) \pi_\gamma^1\left(\boldsymbol{\beta}_\gamma^1, \sigma\right) d\boldsymbol{\beta}_\gamma^1 d\sigma$$

$$\times \cdots \times \int \mathrm{N}_T\left(\mathbf{R}^q \mid \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma^q, \ \sigma^2 \mathbb{I}_{T \times T}\right) \pi_\gamma^q\left(\boldsymbol{\beta}_\gamma^q, \sigma\right) d\boldsymbol{\beta}_\gamma^q d\sigma$$

$$= m_\gamma\left(\mathbf{R}^1\right) \times \cdots \times m_\gamma\left(\mathbf{R}^q\right)$$

$$= \Pi_{i=1}^q m_\gamma\left(\mathbf{R}^i\right),$$

with:

$$(A.2) \qquad \mathbf{R}^i \sim \mathrm{N}_T\left(\mathbf{X}_\gamma \boldsymbol{\beta}_\gamma^i, \ \sigma^2 \mathbb{I}_{T \times T}\right).$$

Therefore, the Bayes factor for this matrix-variate model is just a product of Bayes factors for the individual multivariate normal models - a direct result of the APT model assumptions.

$$(A.3) \qquad B_{\gamma 0} = \widetilde{B}_{\gamma 0}^1 \times \cdots \times \widetilde{B}_{\gamma 0}^q$$

with:

$$(A.4) \qquad \widetilde{B}_{\gamma 0}^i = \frac{m_\gamma\left(\mathbf{R}^i\right)}{m_0\left(\mathbf{R}^i\right)}.$$

The simplification of the marginal likelihood calculation is crucial for analytical simplicity and for the resulting SSVS algorithm to rely on techniques already developed for vector response models. In order to calculate the integral for each Bayes factor, we need priors on the

parameters β_γ and σ. Since the priors are independent across the columns of $\mathbf{R}$, we aim to define $\pi_\gamma^i \left(\boldsymbol{\beta}_\gamma^i, \sigma \right) \ \forall i \in \{1, ..., q\}$, which we express as the product: $\pi_\gamma^i (\sigma) \, \pi_\gamma^i \left(\boldsymbol{\beta}_\gamma^i \mid \sigma \right)$. Motivated by the work on regression problems of Zellner, Jeffreys, and Siow, we choose a non-informative prior for σ and the popular g-prior for the conditional prior on $\boldsymbol{\beta}_\gamma^i$, (Zellner, 1986), (Zellner and Siow, 1980), (Zellner and Siow, 1984), (Jeffreys, 1961):

$$(A.5) \qquad \pi_\gamma^i \left(\boldsymbol{\beta}_\gamma^i, \sigma \mid g \right) = \sigma^{-1} \mathrm{N}_{k_\alpha} \left(\boldsymbol{\beta}_\gamma^i \mid \mathbf{0}, g_\gamma^i \sigma^2 (\mathbf{X}_\gamma^T (\mathbb{I} - T^{-1} \mathbb{1}\mathbb{1}^T) \mathbf{X}_\gamma)^{-1} \right).$$

Under this prior, we have an analytical form for the Bayes factor:

$$(A.6) \qquad B_{\gamma 0} = \widetilde{B}_{\gamma 0}^1 \times \cdots \times \widetilde{B}_{\gamma 0}^q$$

$$(A.7) \qquad = \Pi_{i=1}^q \frac{\left(1 + g_\gamma^i \right)^{(T - k_\gamma - 1)/2}}{\left(1 + g_\gamma^i \frac{SSE_\gamma^i}{SSE_0^i} \right)^{(T+1)/2}},$$

where SSE_γ^i and SSE_0^i are the sum of squared errors from the linear regression of column $\mathbf{R}^i$ on covariates $\mathbf{X}_\gamma$ and k_γ is the number of covariates in model M_γ. We allow the hyper parameter g to vary across columns of $\mathbf{R}$ and depend on the model, denoted by writing, g_γ^i.

We aim to explore the posterior of the model space, given our data:

$$(A.8) \qquad \mathbf{P} \left(M_\gamma \mid \mathbf{R} \right) = \frac{B_{\gamma 0} \mathbf{P} \left(M_\gamma \right)}{\Sigma_\gamma B_{\gamma 0} \mathbf{P} \left(M_\gamma \right)},$$

where the denominator is a normalization factor. In the spirit of traditional stochastic search variable selection Garcia-Donato and Martinez-Beneito (2013), we propose the following Gibbs sampler to sample this posterior.

A.1. **Gibbs Sampling Algorithm.** Once the parameters $\boldsymbol{\beta}_\gamma$ and σ are integrated out, we know the form of the full conditional distributions for $\gamma_i \mid \gamma_1, \cdots, \gamma_{i-1}, \gamma_{i+1}, \cdots, \gamma_p$. We sample from these distributions as follows:

(1) Choose column $\mathbf{R}^i$ and consider two models γa and γ_b such that:

$$\gamma_a = (\gamma_1, \cdots, \gamma_{i-1}, 1, \gamma_{i+1}, \cdots, \gamma_p)$$
$$\gamma_b = (\gamma_1, \cdots, \gamma_{i-1}, 0, \gamma_{i+1}, \cdots, \gamma_p)$$

(2) For each model, calculate B_{a0} and B_{b0} as defined by A.6.
(3) Sample

$$\gamma_i \mid \gamma_1, \cdots, \gamma_{i-1}, \gamma_{i+1}, \cdots, \gamma_p \sim Ber(p_i)$$

where

$$p_i = \frac{B_{a0}\mathbf{P}\left(M_{\gamma_a}\right)}{B_{a0}\mathbf{P}\left(M_{\gamma_a}\right) + B_{b0}\mathbf{P}\left(M_{\gamma_b}\right)},$$

Using this algorithm, we visit the most likely ETF factor models given our set of target assets. Under the model and prior specification, there are closed-form expressions for the posteriors of the model parameters β_γ and σ.

A.2. **Hyper Parameter for the g-prior.** We use a local empirical Bayes to choose the hyper parameter for the g-prior in A.5. Since we allow g to be a function of the columns of $\mathbf{R}$ as well as the model defined by γ, we calculate a separate g for each univariate Bayes factor in A.5 above. An empirical Bayes estimate of g maximizes the marginal likelihood and is constrained to be non-negative. From Liang et al. (2008b), we have:

(A.9) $$\hat{g}_\gamma^{EB(i)} = max\{F_\gamma^i - 1, 0\}$$

(A.10) $$F_\gamma^i = \frac{R_\gamma^{2i}/k_\gamma}{(1 - R_\gamma^{2i})/(T - 1 - k_\gamma)}.$$

For univariate stochastic search, the literature recommends choosing a fixed g as the number of data points Garcia-Donato and Martinez-Beneito (2013). However, the multivariate nature of our model induced by the multiple target assets makes this approach unreliable. Since each target asset has distinct statistical characteristics and correlations with the covariates, it is necessary to vary g among different sampled models and target assets. We find that this approach provides sufficiently stable estimation of the inclusion probabilities for the ETFs.

APPENDIX B. SIMULATION STUDY

In this section of the appendix, we show results of applying our sampling algorithm to simulated data. Recall that we model the conditional, $R|X$, with parameters Ψ and β and the marginal, X, independently with parameters μ_x and Σ_x. Using the posterior means of these parameters, we construct simulated target assets R_{sim} and ETFs X_{sim} under the data generating process:

$$(B.1) \qquad\qquad X_{sim} \sim N(\overline{\mu}_x, \overline{\Sigma}_x)$$

$$(B.2) \qquad\qquad R_{sim} \sim \text{Matrix Normal}_{T,q}\left(X_{sim}\overline{\beta},\ \overline{\Psi},\ \mathbb{I}_{q\times q}\right),$$

where the overlines represent the posterior means. In B.2, we show the true Sharpe ratio as well as its inferred value from our algorithm. The true value is calculated using and the known moments of the data generating process for the simulated returns. The Markov Chain Monte Carlo sampling does an excellent job at recovering the true Sharpe ratio as it is close to the posterior means from three separate simulated data sets.
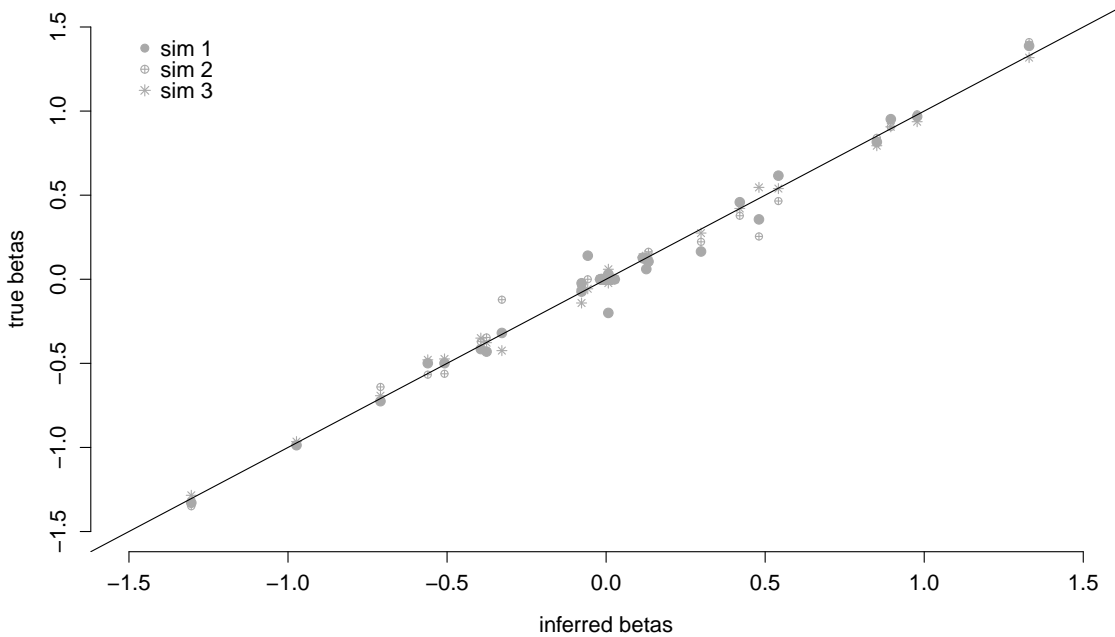

FIGURE B.1. True versus inferred β's for the three sets of simulated R and X.

Additionally, we compare the posterior means of the β coefficients with the true β's used in each of the simulations. The inferred β's line up well with their true values as shown by their proximity to the 45 degree line in figure B.1. Under reasonable data sets mimicking financial
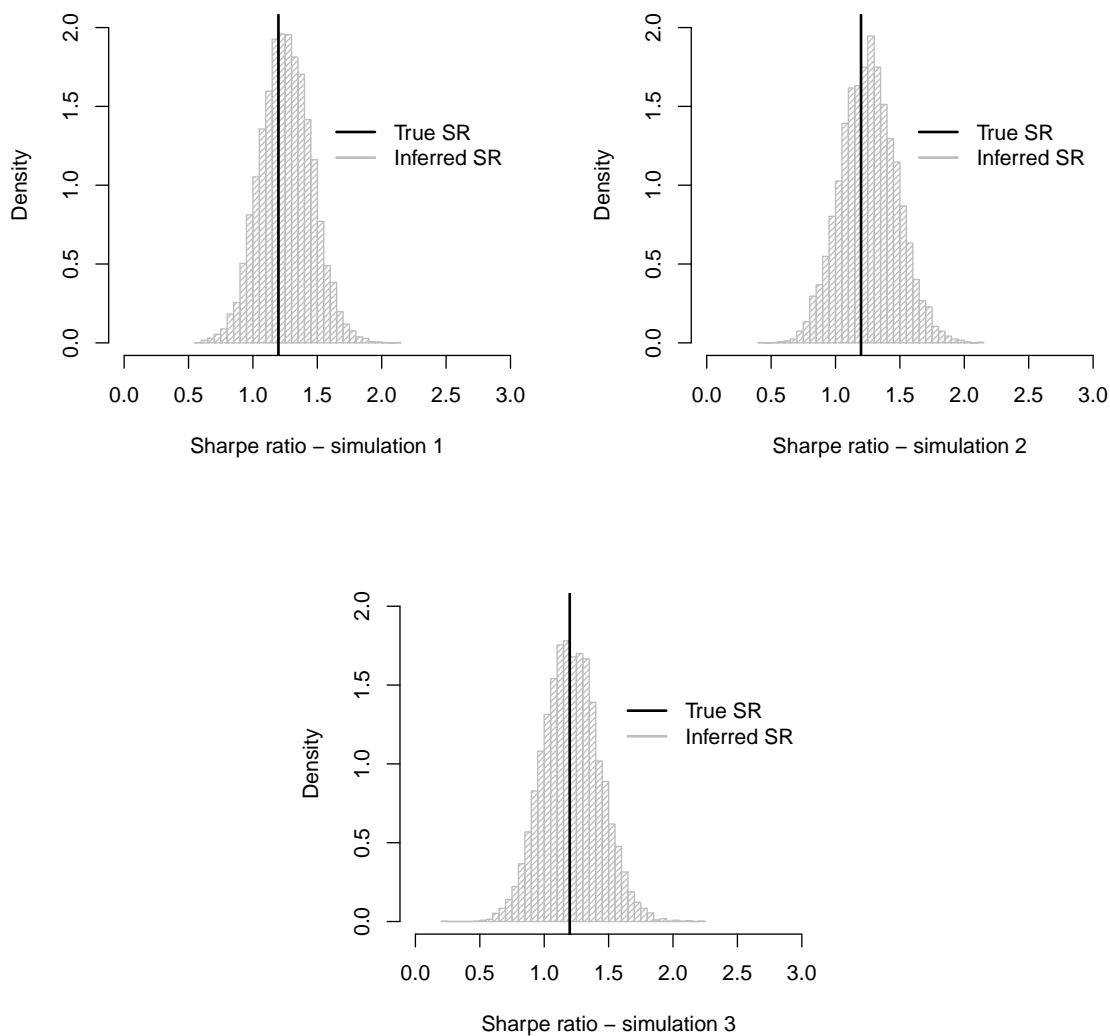
FIGURE B.2. Posterior distribution of Sharpe ratios of the tangency portfolio for three simulate realizations of R and X. The Sharpe ratio of the true tangency portfolio is shown as a vertical black line.

asset returns, our model sampling algorithm does well at recovering the parameters defining the data generating process.

APPENDIX C. COMPARISONS

C.1. **Difference between conditional loss function and graphical lasso.** To demonstrate the difference between the conditional loss function and graphical lasso (glasso) approaches to the selection problem, consider a simple bivariate mean-zero model with one target asset and ETF. Assume parameters a, b, and c have posterior means $\overline{a}, \overline{b}$, and $\overline{c}$.

(C.1)
$$\begin{pmatrix} r \\ x \end{pmatrix} \sim N(\vec{0}, \Sigma)$$
$$\Sigma = \begin{bmatrix} a & c \\ c & b \end{bmatrix}.$$

Analogous to the general setup, define the conditional model as:

(C.2)
$$r|x \sim N\left(\gamma x, d^{-1}\right).$$

The goal is to achieve a parsimonious posterior summary of the off-diagonal element of Σ. To achieve this, one may use the graphical lasso loss function discussed in the appendix of Hahn and Carvalho (2015) where the sparsification penalty only includes the off-diagonal element of our choice variable, which is the precision matrix, Γ:

(C.3)
$$\Gamma = \begin{bmatrix} \psi & g \\ g & \kappa \end{bmatrix}.$$

The glasso loss function from Hahn and Carvalho (2015) is: $\mathscr{L}(\Gamma) = \rho \|\Gamma\| - \log\det(\Gamma) + \text{tr}(\overline{\Sigma}\Gamma)$. Note that ρ is the parameter controlling the amount of penalization. Only penalizing the dependence between r and x, we simplify the loss function to:

(C.4)
$$\mathscr{L}_{\text{glasso}}(g, \psi, \kappa) = \rho|g| - \log(\psi\kappa - g^2) + (\overline{a}\psi + \overline{b}\kappa + 2\overline{c}g).$$

The conditional loss function employed in the paper analogous with equation 3.12 is:

(C.5)
$$\mathscr{L}(\gamma) = \lambda|\gamma| - \frac{1}{2}\gamma^2\overline{b} + \gamma\overline{c},$$

where λ is the penalization parameter. A very important point involves the comparison of C.4 and C.5. In glasso, g is an element in the precision matrix. Therefore, the implied covariance block for a choice of g (using the 2x2 matrix inversion formula) is: $\gamma^*_{\text{glasso}} = -g \det\overline{\Sigma}$. In contrast, the conditional loss function choice variable, γ, is directly the coefficient matrix on x in the

conditional distribution, $r|x$. Thus, comparisons of the two solutions will be made between γ_{glasso} and γ.

C.1.1. *Conditional loss function optimum.* The first order conditions give the optimal action for the conditional loss function C.5, $\gamma^*(\lambda)$:

(C.6)
$$
\gamma > 0 \implies \gamma^*(\lambda) = \frac{\overline{c} - \lambda}{\overline{b}}
$$
$$
\gamma < 0 \implies \gamma^*(\lambda) = \frac{\overline{c} + \lambda}{\overline{b}},
$$

where we divide the action space into γ positive and negative to account for the derivative of the absolute value in the penalty.

C.1.2. *Glasso loss function optimum.* There are three actions for the glasso optimization. Combining the first order conditions on ψ and κ, we conclude that:

(C.7)
$$
\psi = \frac{\overline{b}}{\overline{a}}\kappa.
$$

Substituting this ratio back into the first order conditions for ψ and κ and solving the resulting quadratic equation, we obtain ψ and κ as functions of the parameters and γ^2:

(C.8)
$$
\kappa = \frac{1}{2\overline{b}}\left(1 + \sqrt{1 + 4\overline{a}\overline{b}g^2}\right), \quad \psi = \frac{1}{2\overline{a}}\left(1 + \sqrt{1 + 4\overline{a}\overline{b}g^2}\right).
$$

We take the positive roots to ensure the diagonal elements of our action are positive. This is necessary since glasso seeks a positive definite matrix. The first order condition for g when $g < 0$ implies:

(C.9)
$$
-\rho + \frac{2g}{\psi\kappa - g^2} + 2\overline{c} = 0.
$$

The first order condition for the case of $g > 0$ is the same, but with a positive sign on ρ. Substituting C.8 into C.9, we obtain the optimal action, $g^*(\rho)$:

(C.10)
$$
g > 0 \implies g^*(\rho) = \frac{\frac{1}{2}\rho - \overline{c}}{\det\overline{\Sigma} + \overline{c}\rho - \frac{1}{4}\rho^2}
$$
$$
g < 0 \implies g^*(\rho) = \frac{-\frac{1}{2}\rho - \overline{c}}{\det\overline{\Sigma} - \overline{c}\rho - \frac{1}{4}\rho^2},
$$

where:

$$\text{(C.11)} \qquad \overline{\Sigma} = \left[\begin{array}{cc} \overline{a} & \overline{c} \\ \overline{c} & \overline{b} \end{array} \right].$$

The unpenalized solutions for our conditional loss function and the graphical lasso are:

$$\text{(C.12)} \qquad \begin{aligned} \gamma^*(0) &= \frac{\overline{c}}{\overline{b}} \\ \gamma^*_{\text{glasso}}(0) &= -g^*(0) \det \overline{\Sigma} = \overline{c} \end{aligned}$$

C.1.3. *Numerical demonstration.* Given the derived solutions for the conditional loss function and graphical lasso optimizations, we provide a numerical example of their difference. We set $\overline{a} = 12$, $\overline{b} = 1$, and $\overline{c} = 3$. Setting $\overline{b} = 1$ guarantees that the unpenalized solutions: $\gamma^*(0)$ and $\gamma^*_{\text{glasso}}(0)$ will be equal. Further, since Σ must be positive definite, we have that its determinant, $\overline{a}\overline{b} - \overline{c}^2 = 3$, is positive.

Figure C.1 displays how the optimal solutions for the conditional loss function and graphical lasso change with their penalty parameters - known as the solution paths. For simplicity, we plot both penalty parameters on the same axis. The left part x-axis is the beginning of the solution path where the penalties are large enough to send the solutions to zero. This occurs when $\rho = 2\overline{c}$ and $\lambda = \overline{c}$. The right part of the x-axis shows the unpenalized solutions, and they are designed to be equal in our example.

We see in figure C.1 that the solution paths are very different. The graphical lasso solution depends nonlinearly on its penalty parameter, ρ. The concavity of its solution path can be increased by decreasing $\overline{a}$ towards its constrained value required by $\det \overline{\Sigma} > 0$, necessary for positive definiteness. When $\overline{a} = 200$ as in figure C.2, the graphical lasso solution path becomes much more linear, and the two paths begin to coincide (they are, however, not the same due to the different penalty scales and a choice of $\overline{b}$ other than 1 would affect the slopes). This occurs when the $\det \overline{\Sigma}$ dominates the numerator and denominator terms in $\gamma^*_{\text{glasso}}(\rho) = -g^*(\rho) \det \overline{\Sigma}$. Intuitively, this can also be understood through a correlation argument. As $\overline{a}$ gets large, the correlation between r and x squared: $\frac{\overline{c}^2}{\overline{a}\overline{b}}$ goes to zero. This increased "independence" between r and x results in the penalized graphical lasso objective function becoming exactly the conditional loss objective function. In fact, one could substitute the optimal solutions for κ and ψ (C.8) into the glasso objective function (C.4) and Taylor expand about g (since $g^*(\rho)$ is small when $\overline{a}$ is large) to directly see the similarity between the two optimizations.
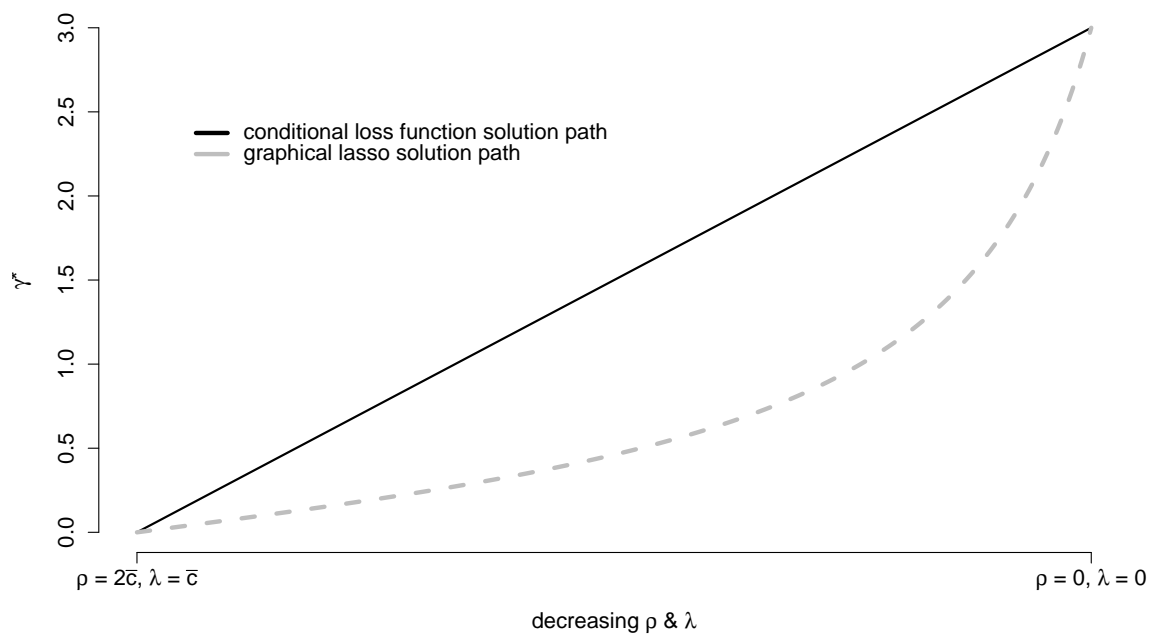
FIGURE C.1. $\overline{a} = 12$
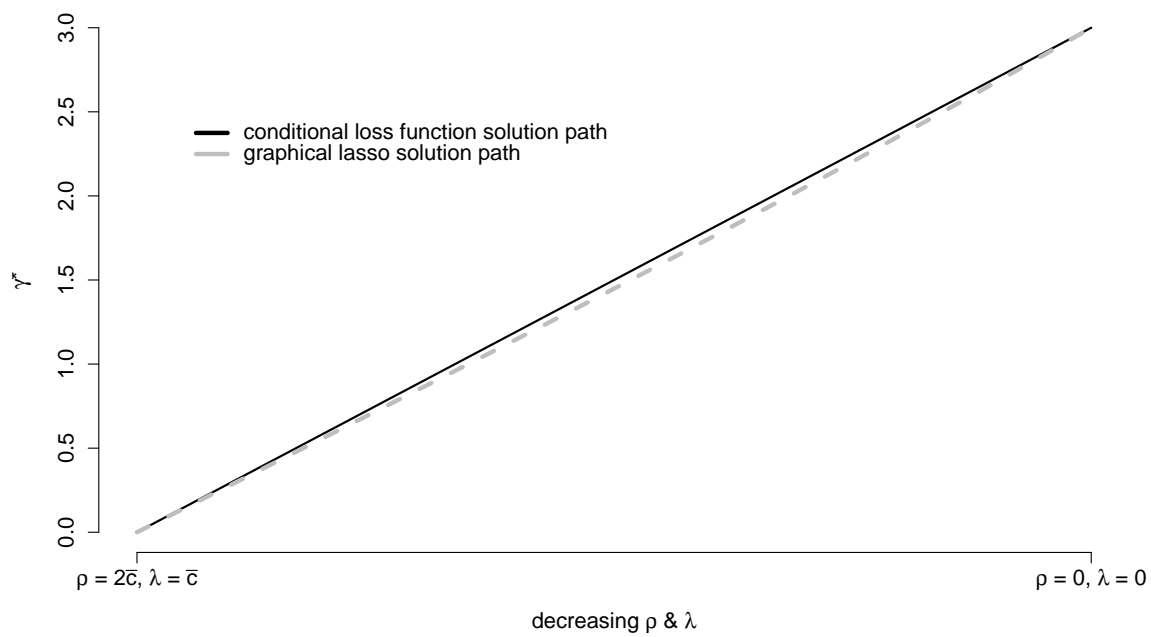
FIGURE C.2. $\overline{a} = 200$

REFERENCES

Ackert, L. F. and Tian, Y. S. (2008). Arbitrage, liquidity, and the valuation of exchange traded funds. *Financial markets, institutions & instruments*, 17(5):331–362.

Agapova, A. (2011). Conventional mutual index funds versus exchange-traded funds. *Journal of Financial Markets*, 14(2):323–343.

Beasley, J. E., Meade, N., and Chang, T.-J. (2003). An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research*, 148(3):621–643.

Brown, P. and Vannucci, M. (1998). Multivariate bayesian variable selection and prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 627–641.

Canakgoz, N. A. and Beasley, J. E. (2009). Mixed-integer programming approaches for index tracking and enhanced indexation. *European Journal of Operational Research*, 196(1):384–399.

Chen, C. and Kwon, R. H. (2012). Robust portfolio selection for index tracking. *Computers & Operations Research*, 39(4):829–837.

Conceicao and Maechler (2015). Deoptimr.

CRSP (1992-2015). The center for research in security prices. Wharton Research Data Services.

Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a bayesian application. *Biometrika*, 68(1):265–274.

DiLellio, J. A. and Jakob, K. (2011). Etf trading strategies to enhance client wealth maximization. *Financial Services Review*, 20(2):145.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fastrich, B., Paterlini, S., and Winker, P. (2013). Constructing optimal sparse portfolios using regularization methods. *Computational Management Science*, pages 1–18.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Garcia-Donato, G. and Martinez-Beneito, M. (2013). On sampling strategies in bayesian variable selection problems with large model spaces. *Journal of the American Statistical Association*, 108(501):340–352.

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Hahn, P. R. and Carvalho, C. M. (2015). Decoupling shrinkage and selection in bayesian linear models: a posterior summary perspective. *Journal of the American Statistical Association*,

110(509):435–448.

Huang, M.-Y. and Lin, J.-B. (2011). Do etfs provide effective international diversification? *Research in International Business and Finance*, 25(3):335–344.

Jacquier, E. and Polson, N. (2010). Bayesian econometrics in finance.

Jeffreys, H. (1961). Theory of probability (3rd edt.) oxford university press.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91.

Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., and West, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, 20(4):pp. 388–400.

Kostovetsky, L. (2005). Index mutual funds and exchange-traded funds. *ETF and Indexing*, 2005(1):88–99.

Liang, F., Paulo, R., Molina, G., Clyde, M., and Berger, J. (2008a). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103:410–423.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008b). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481).

McCulloch, R. (2015). Utility based model selection for bayesian nonparametric modeling using trees. SBIES 2015.

Murray, J. (2015). bfa.

Pastor, L. and Veronesi, P. (2009). Learning in financial markets. Technical report, National Bureau of Economic Research.

Pennathur, A. K., Delcoure, N., and Anderson, D. (2002). Diversification benefits of iShares and closed-end country funds. *Journal of Financial Research*, 25(4):541–557.

Poterba, J. M. and Shoven, J. B. (2002). Exchange traded funds: A new investment option for taxable investors. Technical report, National Bureau of Economic Research.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of economic theory*, 13(3):341–360.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of business*, pages 119–138.

Shin, S. and Soydemir, G. (2010). Exchange-traded funds, persistence in tracking errors and information dissemination. *Journal of Multinational Financial Management*, 20(4):214–234.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Wang, H. (2015). Scaling it up: Stochastic search structure learning in graphical models. *Bayesian Anal.*, 10(2):351–377.

Wang, H., Reeson, C., and Carvalho, C. M. (2011). Dynamic financial index models: Modeling conditional dependencies via graphs. *Bayesian Anal.*, 6(4):639–664.

Wu, L. and Yang, Y. (2014). Nonnegative elastic net and application in index tracking. *Applied Mathematics and Computation*, 227:541–552.

Wu, L., Yang, Y., and Liu, H. (2014). Nonnegative-lasso and application in index tracking. *Computational Statistics & Data Analysis*, 70:116–126.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*, 6:233–243.

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. *Trabajos de estadística y de investigación operativa*, 31(1):585–603.

Zellner, A. and Siow, A. (1984). *Basic issues in econometrics*. University of Chicago Press Chicago.