Portfolio Selection for Individual Passive Investing

David Puelz*
University of Chicago

P. Richard Hahn
Arizona State University

Carlos Carvalho
University of Texas

May 24, 2018

*david.puelz@chicagobooth.edu, 5807 S Woodlawn Ave, Chicago, IL 60637.

Portfolio Selection for
Individual Passive Investing

Abstract

This paper considers passive fund selection from an individual investor's perspective. The growth of the passive fund market over the past decade is staggering. Individual investors who wish to buy these funds for their retirement and brokerage accounts have many options and are faced with a difficult selection problem. Which funds do they invest in, and in what proportions? We develop a novel statistical methodology to address this problem. A Bayesian decision-theoretic approach is presented to construct optimal sparse portfolios for individual investors over time.

Keywords: Bayesian methods, decision theory, portfolio selection, model selection

1 Introduction

This paper considers the problem of portfolio construction from the perspective of an individual investor. These individuals range from recent college graduates to older retirees and share a desire to participate in domestic and global equity markets. Passive investing, which has seen tremendous growth in the past decade, provides an avenue for these individuals to invest. A passive fund is any investable product comprised of many stocks, and it is often mandated to track a broad stock market index. For example, SPY is the ticker symbol for a fund that is mandated to track the S&P 500 index. These funds are low cost for individual investors, and they provide an easy method for diversified equity market exposure.

Over the past two decades, firms have launched thousands of low-cost index-tracking funds, and today more than a third of all US assets are in passive funds. The main question

for individual investors: Which funds do I choose, and how do I construct my portfolio? This is further complicated by the fact that many companies offer funds which track similar indices, and it is unclear how to choose between them. As individuals increasingly allocate their retirement savings to these funds, this problem continues to gain significance.

Practitioners in industry have noticed and capitalized on this individual investing dilemma. Betterment and Wealthfront are two new financial technology companies that construct portfolios with a small number of passive funds for their millennial clients. Their methodology is based on diversifying among asset classes, and they hand select passive funds to provide exposure to different markets and sectors. This paper formalizes the individual investor's desire for portfolio simplicity. We develop a framework for individual fund selection that incorporates both investor utility and predictive uncertainty. Additionally, the selection approach is time-varying and we acknowledge statistical uncertainty using a dynamic model for fund returns.

Ultimately, we are interested in answering the question: *How does one invest optimally while keeping the simplicity (sparsity) of a portfolio in mind?* Our goal, in contrast to the traditional portfolio optimization literature, is not purely predictive or driven by constructing portfolios to maximize the out-of-sample Sharpe ratio. We care about these ideas insofar as they appear in the utility function and predictive distribution, but the framework presented is built around the concept of portfolio simplicity. In our application, the driving notion is that a manifestly simpler portfolio of a couple passive funds is intuitively valuable for an individual investor.

The desire for portfolio simplicity is represented by considering loss functions that balance utility and sparsity myopically for each time t:

$$\mathbf{L}_{\lambda_t}(w_t, \tilde{R}_t) = \mathcal{L}(w_t, \tilde{R}_t) + \Phi(\lambda_t, w_t), \tag{1}$$

where $\tilde{R}_t$ is a vector of future asset returns, $\mathcal{L}$ is the negative utility of the individual investor, w_t is a vector of portfolio weights, Φ is a function that encourages sparsity in w_t, and λ_t is a penalty parameter (or more generally, a penalty vector of equal length as w_t) governing the degree to which the complexity function Φ is influential in the overall loss function.

Note that Loss (1) is a random variable whose uncertainty is induced from the predictive distribution of asset returns $\tilde{R}_t$ – an innovation of this paper is the use of predictive uncertainty to compare simple portfolio decisions with one another. Also, special attention must be paid to λ_t, the parameter that governs the utility-sparsity tradeoff. If it is known a priori, the investor's optimal portfolio may be found for each time by routine minimization of the expectation of (1). By contrast, this paper considers the more challenging case where this parameter is unknown and may be thought of as part of the investor's decision.

The interplay between dynamics, utility and portfolio simplicity in the investor's portfolio decision are viewed through the lens of regret. Assuming the existence of a desirable target portfolio, we define regret as the difference in loss (or negative utility) between the simple portfolio and the target; our investor would like to hold a simple portfolio that is "almost as good as" a target. This paper distills a potentially intractable dynamic selection procedure into one which requires specification of only a single threshold of regret. At the outset, the investor need only answer the question: *With what degree of certainty do I want my simple portfolio to be no worse than the target portfolio?* Put differently: *What maximum probability of regretting my portfolio decision am I comfortable with?*

Once the regret threshold is specified, the investor's preference for portfolio simplicity will automatically adjust over time to accommodate this threshold. In other words, the penalty parameter λ_t continuously adjusts to satisfy the investor's regret tolerance. In one period, her portfolio may only need to be invested in a small number of assets to satisfy her regret threshold. However, that same portfolio may be far off from the target in the

4

next period, requiring her to invest in more assets to accommodate the same level of regret. This thought experiment illustrates that our procedure, although requiring a static regret tolerance to be specified at the outset, results in investor preferences for sparsity that are *dynamic*.

The regret-based selection approach presented in this paper is related to the *decoupling shrinkage and selection (DSS)* procedure from Hahn & Carvalho (2015) and Puelz et al. (2017). In both papers, loss functions with explicit penalties for sparsity are used to summarize complex posterior distributions. Posterior uncertainty is then used as a guide for variable selection in static settings. This paper extends these notions by developing a regret-based metric for selection and placing it within a dynamic framework. The novel contributions are (*i*) The use of investor's *regret* for selection and (*ii*) The development of a principled way to choose a *dynamic* penalty parameter λ_t (and thus select a portfolio) for a time-varying investment problem.

A key finding of this paper is that sparse portfolios and their more complex (or dense) counterparts are often very similar from an *ex ante* regret perspective. More surprisingly, this similarity often persists *ex post.* This gives credence to a common piece of investing advice: "Don't worry about investing in a variety of funds; just buy the market."

1.1   Previous research

Passive fund selection has received limited attention in the literature. Rudd (1980) provides a nice early discussion of the subject. They discuss the differences between passive and active funds and suggest stratification and optimization-based methods for portfolio construction. Sankaran & Patil (1999) develop an approach for mean-variance portfolio construction. They motivate portfolio simplicity by noting that diversification among many securities has diminishing benefits. In both papers, focus is placed on optimization as opposed to modeling the optimization inputs; i.e. characterizing the predictive distribution

of asset returns. The goal of this paper is to incorporate *both* optimization and predictive uncertainty into a principled dynamic selection procedure.

One area of relevant research highlighting Bayesian approaches to this problem may be found in Zhou et al. (2014). In this paper, the authors consider portfolio construction with sparse dynamic latent factor models for asset returns. They show that dynamic sparsity improves forecasting and portfolio performance. However, sparsity in their context is induced at the factor loading level, not the portfolio decision. In contrast, our methodology seeks to sparsify the portfolio decision directly for *any* generic dynamic model.

Additional areas of research focus on the portfolio selection problem, particularly in stock investing and index tracking. Polson & Tew (1999) consider the S&P 500 index and develop a Bayesian approach for large-scale stock selection and portfolio optimization from the index's constituents. The approach presented in this paper is distinct from index tracking. Instead of selecting non-diversified constituents of an index which are often single stocks, we are interested in building simple portfolios of diversified passive funds informed by a utility and predictive distribution. Other insightful Bayesian approaches to optimal portfolio choice include Johannes et al. (2014), Irie & West (2016), Zhao et al. (2016), Gron et al. (2012), Jacquier & Polson (2010), Puelz et al. (2015) and Pettenuzzo & Ravazzolo (2015). Methodological papers exploring high dimensional dynamic models relevant to this work are Carvalho et al. (2007) and Wang et al. (2011).

2   Overview

The focus will be loss functions of the following form:

$$\mathbf{L}_{\lambda_t}(w_t, \tilde{R}_t) = \mathcal{L}(w_t, \tilde{R}_t) + \Phi(\lambda_t, w_t) \tag{2}$$

where $\mathcal{L}$ is the negative utility of an investor, $\tilde{R}_t$ is a vector of N future asset returns, w_t is a vector of portfolio weights, and λ_t is a penalty parameter governing the degree to which the complexity function Φ is influential in the overall loss function. Let the future asset returns be generated from a model parameterized by Θ_t so that $\tilde{R}_t \sim \Pi(\Theta_t)$ and Π is a general probability distribution.

The time-varying preferences in (2) take into account an investor's negative utility as well as her desire for portfolio simplicity. Optimization of (2) in practice poses an interesting challenge since there is uncertainty in model parameters Θ_t and future asset returns $\tilde{R}_t$. Also, the penalty parameter λ_t is not known by the investor a priori, making her risk preferences ambiguous in portfolio complexity. A further obvious complication is that all of these unknowns are varying in time.

We propose a three-step approach to constructing a sequence of sparse dynamic portfolios. This procedure will be based on an investor's regret from investing in an alternative (target) portfolio defined by w_t^*. The three general steps are:

1. *Model specification:* Model the future asset returns $\tilde{R}_t \sim \Pi(\Theta_t)$.

2. *Loss specification:* Specify $\mathcal{L}$ and Φ in Loss (2). Then, the expected loss given by $\mathbf{L}_{\lambda_t}(w_t) = \mathbb{E}[\mathbf{L}_{\lambda_t}(w_t, \tilde{R}_t)]$ may be minimized for a sequence of $\lambda_t \, \forall t$. Define the collection of optimal portfolios in the cross-section as $\{w_{\lambda_t}^*\}$.

3. *Regret-based summarization:* Compare regret-based summaries of the optimal portfolios versus a target portfolio w_t^* by thresholding quantiles of a *regret* probability distribution, where regret as a random quantity is given by $\rho(w_{\lambda_t}^*, w_t^*, \tilde{R}_t)$. This random variable is a function of a sparse portfolio decision $w_{\lambda_t}^*$, the target portfolio w_t^*, and future asset returns.

The expectation and probability are both taken over the joint distribution of unknowns $(\tilde{R}_t, \Theta_t \mid \mathbf{R}_{t-1})$ where $\mathbf{R}_{t-1}$ is observed asset return data. Flexibly, the regret func-

tion $\rho(w^*_{\lambda_t}, w^*_t, \tilde{R}_t)$ can be any metric that is a function of the portfolio weights and unknowns and may be constructed using any target portfolio w^*_t. In this paper, we consider the difference in loss between the sparse portfolio decision and the target portfolio $\rho(w^*_{\lambda_t}, w^*_t, \tilde{R}_t) = \mathcal{L}(w^*_{\lambda_t}, \tilde{R}_t) - \mathcal{L}(w^*_t, \tilde{R}_t)$ as our measure of regret in keeping with the usual decision theoretic definition.

We can now see how portfolio sparsity appears in the dynamic setting. The dynamic model given by $\tilde{R}_t \sim \Pi(\Theta_t)$ interacts with the portfolio decision w_t via the expected loss minimization in step 2. Iterating step 3 over time gives a sequence of sparse portfolio summaries $\{w^*_{\lambda^*_t}\}$ where λ^*_t provides an index for the selected sparse decision. Ultimately, these sparse portfolio summaries select which subsets of assets are relevant for our "simple portfolio".

The details of this procedure will be fleshed out in the following subsection.

2.1 Details of Regret-based summarization

In following section, we discuss the specifics of the regret-based summarization procedure. We focus on expanding on step 3 of the methodology which represents the main innovation of this paper. Model specification and fitting as well as loss specification comprising steps 1 and 2 are only highlighted; a detailed formulation of these first two steps is presented in the Application section.

Suppose that we have inferred a model for future asset returns given observed data $\mathbf{R}_{t-1}$ that is parameterized by the vector Θ_t: $\tilde{R}_t \sim \Pi(\Theta_t \mid \mathbf{R}_{t-1})$. Let the resulting posterior distribution of parameters and future asset returns be $p(\Theta_t, \tilde{R}_t \mid \mathbf{R}_{t-1})$. For example, Π may be parameterized by dynamic mean and variance parameters $\Theta_t = (\mu_t, \Sigma_t)$. Further, suppose we have specified the utility and complexity components of the investor's loss function: $\mathbf{L}_{\lambda_t}(w_t, \tilde{R}_t) = \mathcal{L}(w_t, \tilde{R}_t) + \Phi(\lambda_t, w_t)$.

For each investing period t, we obtain a sequence of portfolio decisions indexed by λ_t

by optimizing the expected loss function $\mathbb{E}[\mathbf{L}_{\lambda_t}(w_t, \tilde{R}_t)]$. Regret-based summarization is an approach to select the appropriate optimal decision from the collection (i.e., select λ_t) for each time t, and this choice may be visualized by using sparsity-regret tradeoff plots.

Revisiting the regret, samples of the $\tilde{R}_t$ margin from posterior distribution $(\tilde{R}_t, \Theta_t \mid \mathbf{R}_{t-1})$ define the distribution for the regret random variable $\rho(w^*_{\lambda_t}, w^*_t, \tilde{R}_t)$ given by a difference in loss:

$$\rho(w^*_{\lambda_t}, w^*_t, \tilde{R}_t) = \mathcal{L}(w^*_{\lambda_t}, \tilde{R}_t) - \mathcal{L}(w^*_t, \tilde{R}_t), \tag{3}$$

where $w^*_{\lambda_t}$ are the optimal sparse portfolio weights for penalty parameter λ_t and w^*_t are the weights for a target portfolio – any portfolio decision the investor desires to benchmark against. Regret (3) is a random variable whose uncertainty is induced from the joint posterior distribution $(\tilde{R}_t, \Theta_t \mid \mathbf{R}_{t-1})$ from step 1 of the procedure.

We use the regret random variable as a tool for sparse portfolio selection. Each portfolio decision indexed by λ_t is assigned a number:

$$\pi_{\lambda_t} = \mathbb{P}[\rho(w^*_{\lambda_t}, w^*_t, \tilde{R}_t) < 0] \tag{4}$$

which is the probability that the sparse portfolio is no worse than the dense (target) portfolio. In other words, π_{λ_t} is the probability that I will not "regret" investing in the sparse λ_t-portfolio over the target portfolio. This may also be called the *satisfaction probability* for the sparse λ_t portfolio decision.
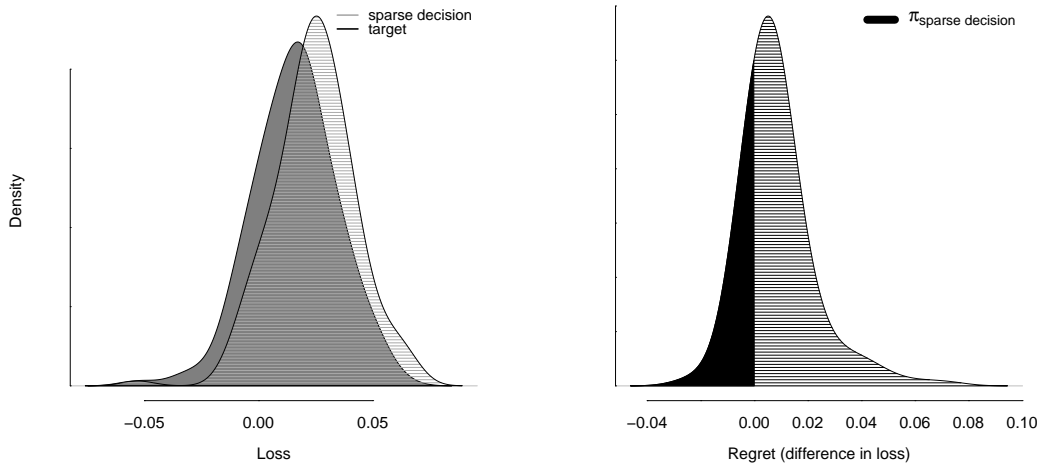
Figure 1: Loss (left) and regret (right) for an example using returns on 25 passive indices. The loss is defined by the log cumulative wealth. The sparse decision is a portfolio invested in 4 indices and is represented by the light shaded gray region. The target decision is a portfolio optimized over all 25 indices and is represented by the shaded black region. The regret distributions shown on the right represent the random variables constructed by subtracting the sparse decision loss from the target loss. Additionally, the black shaded region on the right shows $\pi_{\text{sparse decision}}$: The probability that the sparse decision is no worse than the target decision.

In Figure (1), we provide an illustration of the connection bewteen the loss and regret random variables. This figure is constructed using returns on 25 passive indices and a next period "log cumulative wealth" utility function. This is done for a snapshot in time with a focus on one sparse decision that is invested in 4 out of the 25 indices. The investor is considering this decision versus her target – a portfolio optimized over all 25 indices. The left figure displays the loss distributions of the sparse decision and target. The probability mass of the sparse loss is gathered at larger values compared with the target loss. It is "more costly" (higher loss potential) to neglect diversification benefits and invest in fewer assets.

The right plot in Figure (1) displays the regret distribution for the sparse decision. This is constructed by taking the difference between the sparse and target losses, as given in Equation (3), defining the regret random variable. With the regret distribution in hand, we can compute the probability that the sparse decision is no worse than the target portfolio

given by Equation (4) – this may be referred to as the "satisfaction probability" for the sparse decision. This is shown by the black shaded area on the right in Figure (1). The larger this probability, the "closer" the sparse decision's loss is to the target loss. By making a decision that satisfies a lower bound on this probability called κ, we are able to control our chance of being the same or better than a target portfolio. A lower bound (κ) on the probability of satisfaction (no regret) implies an upper bound ($1 - \kappa$) on the probability of regret.
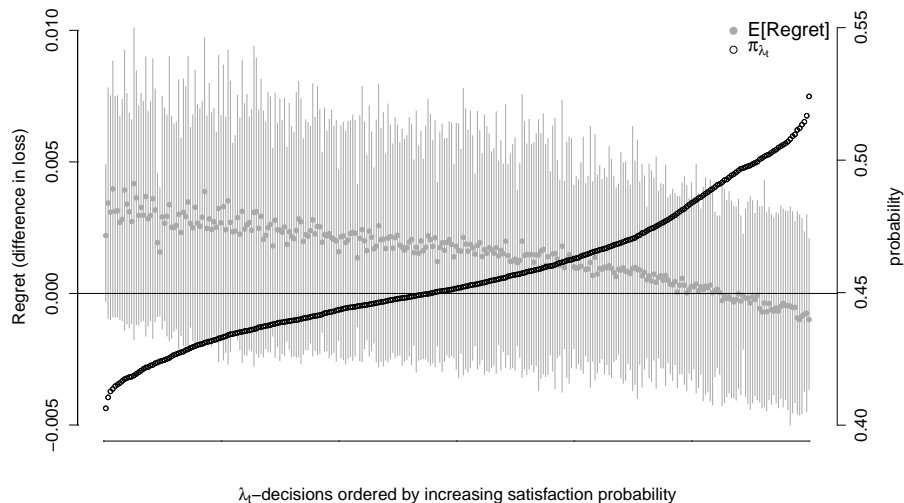

Figure 2: Regret distributions (left vertical axis) and π_{λ_t} (right vertical axis) for increasing π_{λ_t} values from left to right on the horizontal axis. Displayed are 300 sparse portfolio decisions indexed by λ_t. As the satisfaction probability (π_{λ_t}) increases, the mean regret represented by the gray dots will will typically trend downwards. Gray bands represent 20% centered posterior credible intervals for the regret.

The investor's portfolio decision boils down to answering the question first posed in the introduction: *With what degree of certainty do I want my simple portfolio to be no worse than the target portfolio?* As the investor moves through time, the loss and regret distributions will evolve and so will the probability associated with the sparse λ_t portfolio decisions. A dynamic sparse portfolio decision extends this probability thresholding approach to a time-varying framework. The investor chooses the a portfolio decision satisfying $\pi_{\lambda_t} > \kappa \quad \forall t$, ensuring she holds a portfolio that satisfies her regret tolerance in every

investing period.

In Figure (2), we show an example sequence of regret distributions (right vertical axis and gray bars) indexed by λ_t as well as the satisfaction probability π_{λ_t} (left vertical axis and open circles). Specifically, we show the regret distributions for 300 sparse decisions under the log cumulative wealth utility and versus a target portfolio optimized over all available assets. The highest regret decisions (by satisfaction probability) are on the left, and they become less regretful as one moves to the right on the horizontal axis. These sparse portfolio decision are all fairly close in terms of loss to the target decision. Therefore, the corresponding regret distributions hover around zero, and the π_{λ_t}'s hover around 0.5. Exploratory plots like Figure (2) aid in choosing a proper value for the κ threshold. In this case $\pi_{\lambda_t} > 0.5$ is quite high.

Once the time invariant threshold κ is specified, a dynamic selection strategy is easily implementable. At each time t, we are presented with a set of decisions such as those displayed in Figure (2) and choose the sparse portfolio decision such that its $\pi_{\lambda_t} > \kappa$. If there are several sparse decisions satisfying the threshold, we may choose the one whose π_{λ_t} is closest to κ. Of course, there is flexibility in the final step of selecting among admissible sparse decisions. For example, one may select a sparse decision at time t that is "close" (in terms of a norm or assets held) to the previous sparse decision at time $t - 1$ to reduce transaction costs related to the buying and selling of assets. These features will be discussed in the Application section.

3 Application

This section is divided into three parts. First, we describe a dynamic linear model used to infer time-varying moments of asset returns and fulfill step 1 of the methodology. Second, we specify loss and resulting regret functions used for selection (steps 2 and 3). Third,

we demonstrate the methodology using a set of 25 passive funds. In the demonstration sections, we consider a simple example and a more practical case study.

3.1   Model specification and data

The general model we infer parameterizes the distribution of future asset returns with a mean and covariance indexed by time: $\tilde{R}_t \sim \Pi(\mu_t, \Sigma_t)$. An important feature of our proposed methodology is that any model providing estimations of these time-varying moments may be used.

To demonstrate our methodology, we estimate a dynamic linear model (DLM) motivated by Fama & French (2015) who detail five "risk factors" with future returns $\tilde{R}_t^F$ as relevant to asset pricing. Specifically, we model the joint distribution of future asset returns and factor returns $p(\tilde{R}_t, \tilde{R}_t^F)$ compositionally by modeling $p(\tilde{R}_t \mid \tilde{R}_t^F)$ and $p(\tilde{R}_t^F)$. Following the dynamic linear model setup from Harrison & West (1999), the future return of asset i is a linear combination of future factor returns $(\tilde{R}_t^i \mid \tilde{R}_t^F)$:

$$\tilde{R}_t^i = (\beta_t^i)^T \tilde{R}_t^F + \epsilon_t^i, \qquad \epsilon_t^i \sim N(0, 1/\phi_t^i),$$

$$\beta_t^i = \beta_{t-1}^i + w_t^i, \qquad w_t^i \sim \mathrm{T}_{n_{t-1}^i}(0, W_t^i),$$

$$\beta_0^i \mid D_0 \sim \mathrm{T}_{n_0^i}(m_0^i, C_0^i),$$

$$\phi_0^i \mid D_0 \sim \mathrm{Ga}(n_0^i/2, d_0^i/2), \tag{5}$$

$$\beta_t^i \mid D_{t-1} \sim \mathrm{T}_{n_{t-1}^i}(m_{t-1}^i, R_t^i), \qquad R_t^i = C_{t-1}^i/\delta_\beta,$$

$$\phi_t^i \mid D_{t-1} \sim \mathrm{Ga}(\delta_\epsilon n_{t-1}^i/2, \delta_\epsilon d_{t-1}^i/2),$$

where $W_t^i = \frac{1-\delta_\beta}{\delta_\beta} C_{t-1}^i$ and D_t is all information up to time t. This model permits the coefficients on the factors as well as the observation and state level variances to vary in time.

Pre-specified discount factors δ_ϵ and $\delta_\beta \in (0.8, 1)$ accomplish this goal for the observation and state level variances, respectively. Also, note that C_t^i (the posterior variance of the state equation for β_t^i) is updated through moments of the prior $\beta_t^i \mid D_{t-1}$ and the one-step ahead forecast distribution $\tilde{R}_t^i \mid D_{t-1}$. Theorem 4.1 in Harrison & West (1999) provides the general updating equations for the univariate DLM. Table 10.4 in the book summarizes the recurrence relationships in the special case of variance discounting, providing the moments of the posteriors of the parameters $\{m_t^i, C_t^i, n_t^i, d_t^i\}$ for all t and each asset i.

We model the five factor future returns $\tilde{R}_t^F$ with a full residual covariance matrix using the following matrix normal dynamic linear model:

$$\tilde{R}_t^F = \mu_t^F + \nu_t \qquad \nu_t \sim \mathrm{N}(0, \Sigma_t^F),$$

$$\mu_t^F = \mu_{t-1}^F + \Omega_t \quad \Omega_t \sim \mathrm{N}(0, W_t, \Sigma_t^F),$$

$$(\mu_0^F, \Sigma_0^F \mid D_0) \sim \mathrm{NW}_{n_0}^{-1}(m_0, C_0, S_0), \tag{6}$$

$$(\mu_t^F, \Sigma_t^F \mid D_{t-1}) \sim \mathrm{NW}_{\delta_F n_{t-1}}^{-1}(m_{t-1}, R_t, S_{t-1}),$$

$$R_t = C_{t-1}/\delta_c,$$

where $W_t = \frac{1-\delta_c}{\delta_c} C_{t-1}$. Analogous to Model (5), the discount factors δ_F and δ_c in Model (6) serve the same purpose of permitting time variation in the observation and state level variances, respectively. An added benefit of (6) is that Σ_t^F is a full residual covariance matrix.

Elaborating on the intuition behind Models (5) and (6) and guided by Harrison & West (1999), the purpose of variance discounting is to provide a natural way to evolve the variance from the posterior to the prior while maintaining conjugacy for sequential

updating. For example, consider the posterior of the precision in Model (5):

$$\phi_{t-1}^i \mid D_{t-1} \sim \mathrm{Ga}(n_{t-1}^i/2, d_{t-1}^i/2). \tag{7}$$

To construct $p(\phi_t^i \mid D_{t-1})$, we wish to maintain a Gamma form so it is conjugate with the likelihood function for $\tilde{R}_t^i$ given by the one-step ahead forecast distribution. One reasonable approach is to preserve the mean of Distribution (7), but inflate the variance by discounting the degrees of freedom parameter $n_{t-1} \to \delta_\epsilon n_{t-1}$. The prior distribution then becomes:

$$\phi_t^i \mid D_{t-1} \sim \mathrm{Ga}(\delta_\epsilon n_{t-1}^i/2, \delta_\epsilon d_{t-1}^i/2). \tag{8}$$

Moving from Distribution (7) to (8) increases the dispersion of the prior to represent a "loss of information" characteristic of moving forward to time t with a lack of complete information in D_{t-1}. The remaining discount factors $\delta_\beta, \delta_C,$ and δ_F in Models (5) and (6) serve the analogous purpose of *variance inflation* for their respective stochastic processes.

The limiting behavior of variance-discounted learning corresponds to exponentially weighting historical data with decreasingly smaller weights given to values further in the past (see sections 10.8.2-3 in Harrison & West (1999)). Larger discount factors correspond to slower decaying weights and suggest the time series of parameters is slower-fluctuating. Smaller discount factors intrinsically mean we have less data with which to estimate the parameters because the sequence is believed to be more rapidly fluctuating. Thus, the choice of discount factors amounts to choosing decaying weights for previous data that are relevant for predicting the parameters today.

Models (5) and (6) together constitute a time-varying model for the joint distribution of future asset and factor returns: $p(\tilde{R}_t, \tilde{R}_t^F) = p(\tilde{R}_t \mid \tilde{R}_t^F)p(\tilde{R}_t^F)$. As detailed in Harrison & West (1999), they are Bayesian models that have closed form posterior distributions of all parameters at each time t, and the absence of MCMC is convenient for fast updating and

15

uncertainty characterization – a necessary ingredient for our regret-based portfolio selection procedure. Under these models, we obtain the following mean and covariance structure:

$$\mu_t = \beta_t^T \mu_t^F,$$

$$\Sigma_t = \beta_t \Sigma_t^F \beta_t^T + \Psi_t,$$

(9)

where column i of β_t are the coefficients on the factors for asset i, β_t^i. Also, Ψ_t is a diagonal matrix with ith element $\Psi_{tii} = 1/\phi_t^i$. The parameters $\Theta_t = (\mu_t, \Sigma_t)$ are inputs to step 2 of the procedure.

3.1.1   Data and choice of discount factors

We use data on the 25 most highly traded (i.e., most liquid) equity funds from ETFdb.com as our investable assets. This is monthly data from the Center for Research in Security Prices (CRSP) database from February 1992 through May 2016 (CRSP, 1992-2016). Since the ETF market is relatively young, we backfill the ETF return data with returns on the index each fund is mandated to track. This allows us to expand the length of our data set back to 1992.

The fund names, tickers, and sample statistics are displayed in Table (1). The returns on the Fama-French five factors are obtained from the publicly available data library on Ken French's website.[1] We start with 10 years of data to train the model and begin forming portfolios in February 2002.

---

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

Fund name	Ticker	Return (%)	St. Dev. (%)
SPDR Dow Jones Industrial Average	DIA	8.06	14.15
SPDR S&P 500 ETF	SPY	7.46	14.34
SPDR Industrial Select Sector	XLI	9.19	16.49
Guggenheim S&P 500 Equal Weight ETF	RSP	9.37	15.92
iShares MSCI Emerging Markets	EEM	6.06	22.78
iShares MSCI EAFE	EFA	3.94	16.43
iShares MSCI Germany	EWG	6.82	22.23
iShares MSCI Japan	EWJ	0.22	19.28
iShares MSCI United Kingdom	EWU	4.63	15.88
iShares MSCI South Korea Capped	EWY	9.46	36.78
iShares MSCI Eurozone	EZU	6.00	19.85
iShares S&P 500 Value	IVE	7.36	14.95
iShares Core S&P 500	IVV	7.48	14.34
iShares Russell 1000	IWB	8.22	14.00
iShares Russell 1000 Value	IWD	8.10	14.26
iShares Russell 1000 Growth	IWF	7.63	14.56
iShares Russell 2000	IWM	8.00	18.76
iShares Russell 2000 Value	IWN	7.78	18.67
iShares Russell 2000 Growth	IWO	6.63	22.14
iShares Russell Mid-Cap Growth	IWP	8.52	19.98
iShares Russell Mid-Cap	IWR	9.52	15.96
iShares Russell 3000	IWV	7.52	14.61
iShares US Real Estate	IYR	9.32	19.12
iShares US Technology	IYW	11.64	26.09
iShares S&P 100	OEF	7.32	14.61

Table 1: List of exchange-traded funds (ETFs) used for the empirical study. Also displayed are the ticker symbols and realized return and standard deviation (annualized) over their sample period.

Step 1 in our procedure is specifying and inferring a model for asset returns. For our empirical analysis, we use Models (5) and (6). The discount factors for the factor coefficient and factor mean processes are set to $\delta_c = \delta_\beta = 0.9925$, and we consider time-varying residual variances $\delta_F = \delta_\epsilon = 0.97$. Evidence of time-varying residual variance is well-documented in the finance literature (Ng, 1991). The discount factors are chosen to incorporate an adequate amount of data in the exponentially weighted moving window. When $\delta = 0.9925$ (0.97), data eight (three) years in the past receives half the weight of data today. We require slower decay for the factor coefficients and mean processes because more data is needed to learn these parameters than residual volatility.

3.2 Loss and regret specification

We consider a loss function defined by the negative log cumulative return of a portfolio decision for N assets. Recalling general form of the loss function: $\mathbf{L}_{\lambda_t}(w_t, \tilde{R}_t) = \mathcal{L}(w_t, \tilde{R}_t) + \Phi(\lambda_t, w_t)$, define:

$$\mathcal{L}(w_t, \tilde{R}_t) = -\log\left(1 + \sum_{k=1}^{N} w_t^k \tilde{R}_t^k\right), \tag{10}$$

The utility in Loss (10) may be viewed as a version of the Kelly portfolio criterion (Kelly Jr, 1956) where the investor's preferences involve the portfolio growth rate. The complexity function $\Phi(\lambda_t, w_t)$ is separately specified in each of the two examples to follow.

Portfolio decisions w_t may now be evaluated using the negative log cumulative return preferences given by $\mathcal{L}(w_t, \tilde{R}_t)$. However, in order to find these portfolio decisions, we must first optimize the expectation of Loss (10). We do this in two steps: First, we approximate the loss using a second order Taylor expansion, and second, we take the expectation over all unknowns and optimize for each λ_t.

Following the work of Rising & Wyner (2012), we consider a convenient second order Taylor approximation $\mathcal{L}(w_t, \tilde{R}_t) \approx \mathring{\mathcal{L}}(w_t, \tilde{R}_t)$ of the original Loss (10) expanded about $w_0 = \vec{0}$:

$$\mathring{\mathcal{L}}(w_t, \tilde{R}_t) = \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} w_t^k w_t^j \tilde{R}_t^k \tilde{R}_t^j - \sum_{k=1}^{N} w_t^k \tilde{R}_t^k, \tag{11}$$

where we write the approximate loss including the penalty function as $\mathring{\mathbf{L}}_{\lambda_t}(w_t, \tilde{R}_t) = \mathring{\mathcal{L}}(w_t, \tilde{R}_t) + \Phi(\lambda_t, w_t)$. The approximate expected loss $\mathbb{E}[\mathring{\mathbf{L}}_{\lambda_t}(w_t, \tilde{R}_t)]$ is written as:

$$\mathbb{E}[\mathring{\mathbf{L}}_{\lambda_t}(w_t, \tilde{R}_t)] = \mathbb{E}\left[\frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} w_t^k w_t^j \tilde{R}_t^k \tilde{R}_t^j - \sum_{k=1}^{N} w_t^k \tilde{R}_t^k + \Phi(\lambda_t, w_t)\right]. \tag{12}$$

With the posterior distribution $(\Theta_t, \tilde{R}_t \mid \mathbf{R}_{t-1})$ in hand from step 1, we can take the expectation. We integrate over $(\tilde{R}_t \mid \Theta_t, \mathbf{R}_{t-1})$ followed by $(\Theta_t \mid \mathbf{R}_{t-1})$ to obtain the integrated approximate loss function:

$$
\begin{aligned}
\mathring{\mathbf{L}}_{\lambda_t}(w_t) = \mathbb{E}[\mathring{\mathbf{L}}_{\lambda_t}(w_t, \tilde{R}_t)] &= \mathbb{E}_{\Theta_t}\left[\mathbb{E}_{\tilde{R}_t \mid \Theta_t}\left[\mathring{\mathbf{L}}_{\lambda_t}(w_t, \tilde{R}_t)\right]\right] \\
&= \mathbb{E}_{\Theta_t}\left[\frac{1}{2}w_t^T \Sigma_t^{\mathrm{nc}} w_t - w_t^T \mu_t + \Phi(\lambda_t, w_t)\right] \\
&= \frac{1}{2}w_t^T \overline{\Sigma}_t^{\mathrm{nc}} w_t - w_t^T \overline{\mu}_t + \Phi(\lambda_t, w_t),
\end{aligned}
\tag{13}
$$

where the overlines denote posterior means of the mean μ_t and non-central second moment Σ_t^{nc}. The non-central second moment is calculated from the variance as $\Sigma_t^{\mathrm{nc}} = \Sigma_t + \mu_t \mu_t^T$. Loss function (13) may be minimized for a range of λ_t values at each time t.

In the subsections to follow, we will present two analyses using this model and data. First, we discuss a simple unsupervised example to demonstrate the regret-based selection procedure. Second, we present an in depth practical case study.

3.3 Example: Portfolio decisions with limited gross exposure

In this section, we present a simple example demonstrating the main components of regret-based selection. We complete Loss function (2) by specifying the complexity function as the ℓ_1 norm of the weight vector: $\Phi(\lambda_t, w_t) = \lambda_t \|w_t\|_1$. The complexity function measures *gross exposure* of a decision by summing the absolute value of each position: $\|w_t\|_1 = \Sigma_i |w_t^i|$. Decisions with larger absolute value components will evaluate to larger ℓ_1 norms. The penalty parameter λ_t corresponds directly to a single portfolio decision by amplifying the penalty in the loss function.

The approximate Loss function (13) is now convex and may be written as:

$$
\mathring{\mathbf{L}}_{\lambda_t}(w_t) = \frac{1}{2}w_t^T \overline{\Sigma}_t^{\mathrm{nc}} w_t - w_t^T \overline{\mu}_t + \lambda_t \|w_t\|_1.
\tag{14}
$$

19

Loss function (14) is readily optimized by a variety of software packages – please see the Appendix for details. Given its computational convenience, it possesses a couple important features worth noting.

First, Loss (14) requires no enumeration of decisions; it can be minimized quickly for a range of λ_t. In this way, it is an "unsupervised" approach to the sparse portfolio selection problem. Second, λ_t now has explicit meaning beyond indexing the decisions. Since it is multiplying the complexity function, larger (smaller) λ_t will correspond to sparser (denser) portfolio decisions. Conveniently, this displays the regret-based procedure's usefulness in selecting tuning parameters in penalized optimization problems with time-varying inputs. The dynamic nature of asset return data renders traditional cross validation approaches using i.i.d. sampled testing and training splits inappropriate.
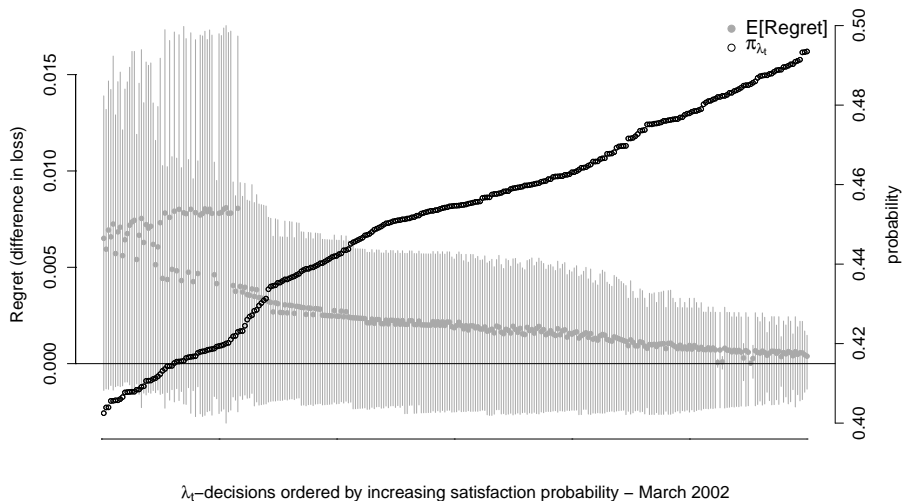
Figure 3: Regret distributions (left vertical axis) and π_{λ_t} (right vertical axis) for increasing π_{λ_t} values from left to right on the horizontal axis. Displayed are 300 of the sparse portfolio decisions indexed by λ_t for March 2002. As the satisfaction probability (π_{λ_t}) increases, the mean regret represented by the gray dots will will typically trend downwards. Gray bands represent 20% centered posterior credible intervals for the regret.

We optimize Loss function (14) for a range of λ_t and for each time t. Specifically, we consider 500 λ_t values spanning the sparsest "one asset decision" to the $\lambda_t = 0$ decision. The

latter decision is referred to as the unpenalized *Kelly optimal* portfolio and puts nonzero weight in all available assets. This is used as the target decision since there is no limit on gross exposure and will also be called the *dense portfolio*. We also normalize all decisions to sum to one, and allow long *and* short positions in assets. We normalize all decisions so that the investor is neither borrowing nor investing in the risk free (cash) rate. Instead, she is fully invested in a risky asset portfolio. In other words, denoting w_{cash} and w_{risky} as the percentage of wealth in cash and risky assets (the ETFs) respectively, $w_{\text{cash}} + w_{\text{risky}} = 1$ will always hold, and this paper considers the case when $w_{\text{risky}} = 1$.

At each point in time, the investor would like to choose among the 500 sparse decisions indexed by λ_t. This may be done by first computing the corresponding satisfaction probabilities π_{λ_t} for each of the 500 decisions under consideration and then choosing one that satisfies and a pre-specified threshold κ. Recall that π_{λ_t} is defined in (4) as the probability that the regret (versus the dense target) is less than zero. The utility, as specified in Equation (10), is the next period log cumulative wealth.

Figure (3) displays the cross-sectional regret distributions (left vertical axis) and satisfaction probabilities π_{λ_t} (right vertical axis) for 300 of the sparse decisions in March 2002. As prescribed by the regret-based procedure, the investor uses this information to select a portfolio. The satisfaction probabilities span 0.4 to 0.5 indicating that the decisions in this investing period are all quite similar. Guided by this figure, we choose a $\kappa = 42.5\%$ threshold to construct an example sequence of selected portfolios.

Once the static threshold is selected, we can iterate the selection procedure through time. At each time t, the investor is confronted with ex ante regret information provided by a cross-sectional plot like Figure (3) and selects a portfolio that satisfies the threshold κ. Once the sequence of decisions is constructed, we can look at how the regret distribution varies *over time*.

Figure (4) shows precisely how the regret of the selected decisions evolve over time.

21

This example demonstrates how both the mean (black line) and variance (surrounding shaded black regions) of regret can vary substantially. Notice that the regret is close to zero with small variance for most periods of time. However, surrounding the financial crisis in 2009, the mean increases and then drops below zero and the variance increases. When regret is negative, the utility of sparse portfolio decision exceeds that of the dense portfolio. During crisis periods shortly into 2009, sparse portfolio decisions appear to be preferred (as measured by ex ante investor utility) to the dense portfolio. Nonetheless, this drop in mean is accompanied by increased variance which informs the investor to be wary of the precision of her regret assessment.
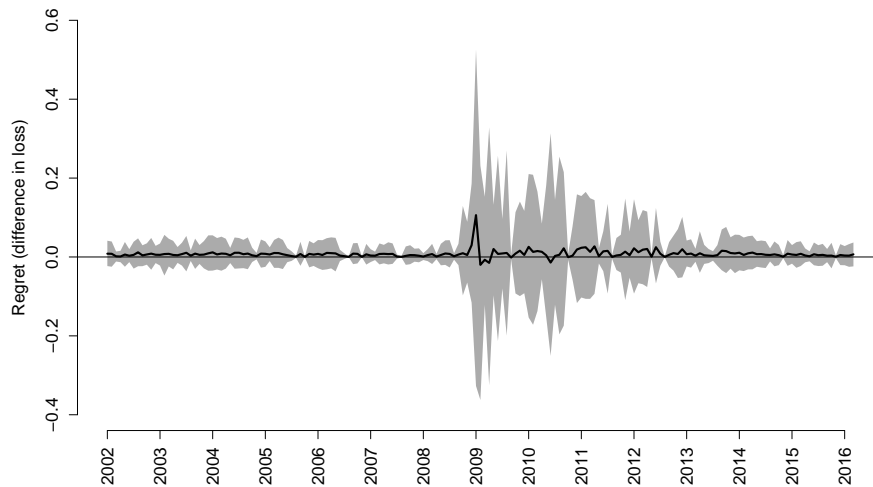
Figure 4: The evolution of the ex ante regret distributions for the sparse long/short portfolio decision given by a $\kappa = 42.5\%$ threshold and versus the unpenalized Kelly optimal target. The mean regret is shown by the lines and the surrounding areas represent the evolving centered 60% posterior credible intervals.

3.4   Case study: Selection among a large set of sparse decisions

The purpose of the following case study is to demonstrate regret-based portfolio selection for a layman investor. We assume our investor would like to hold a broad market fund (SPY) and a couple more diversifying positions in other funds. Additionally, we consider a scenario where the investor cannot hold negative positions in funds; i.e., short selling

22

is prohibited. Therefore, we consider decisions of only positive weights: $w_t \geq 0 \; \forall t$. We construct a set of portfolio decisions for a layman investor using the following rules for a sparse portfolio with $q < N$ funds:

1. $\geq 25\%$ of the portfolio is invested in SPY, a broad market fund tracking an index comprised of the 500 largest US companies by market capitalization.

2. $\geq 25\%$ of the portfolio is diversified across the $q-1$ non-market funds in the following way: The $q-1$ non-market funds each have weights $\geq \frac{25}{q-1}\%$.

We consider portfolios of two, three, four, and five funds, all of which include SPY. Each of these sparse portfolios are optimized using the unpenalized *Kelly optimal* loss as the objective (Loss (13) without the complexity function) and constraints defined as above. Since our data has 24 funds excluding SPY, enumeration of decisions in this way results in $\Sigma_{i=1}^4 \binom{24}{i} = 12,950$ sparse portfolios to select among. Enumeration of sparse decisions implies a complexity function that measures the *cardinality* or number of funds included in a portfolio. Since the complexity function is now implicit in the sparse enumeration, λ_t may be thought of as a convenient indexing of each possible portfolio decision.

As presented in the initial example, we must specify a target decision which then defines the regret random variable defined in Equation (3). We consider two targets at opposite ends of the sparsity spectrum for the empirical analysis.

1. **Dense target**: The unpenalized *Kelly optimal* decision; a portfolio optimized over all available assets. Define the Kelly optimal decision as $w_t^* = \arg\min_{w_t \geq 0} \mathring{\mathcal{L}}(w_t)$ where $\mathring{\mathcal{L}}(w_t) = \mathbb{E}[\mathring{\mathcal{L}}(w_t, \tilde{R}_t)] = \frac{1}{2} w_t^T \overline{\Sigma}_t^{\text{nc}} w_t - w_t^T \overline{\mu}_t$; the optimal decision in absence of the penalty function. This is the same target used in the "ℓ_1 penalty" example presented above, now with a positivity constraint on the weights.

2. **Sparse target**: The *market*; a portfolio composed of one asset representing broad exposure to the financial market. We choose SPY as the market fund.

The choice of each target will give an investor vastly different perspectives on sparse portfolio selection. In the dense target case, the investor desires a sparse portfolio decision that is close (in terms of regret) to a potentially unattainable decision involving all possible funds. The sparse target turns this approach on its head. In this case, the sparse target approach will inform the investor of the added benefit (if any) in diversifying away from a broad market fund.

Each of the 12,950 sparse decisions has a probability of satisfaction versus a target (π_{λ_t}) which can be readily calculated via simulation at each point in time using Equations (3) and (4) and the distribution of future returns given by Models (5) and (6). In Figure (5), we show histograms of the satisfaction probabilities for March 2002 across all 12,950 sparse decisions. It is related to Figure (2) in that the satisfaction probabilities corresponding to the right vertical axis are shown in histogram form, now for various targets. The probabilities versus the dense (SPY) target are shown in black (gray). The dense target is the *dense* portfolio optimized over all 25 funds. Satisfaction versus this dense portfolio decision are gathered at smaller probabilities when compared with the SPY portfolio decision. Of course, the satisfaction rate versus a diversified dense portfolio will intuitively be lower than versus a sparse portfolio of a single fund.
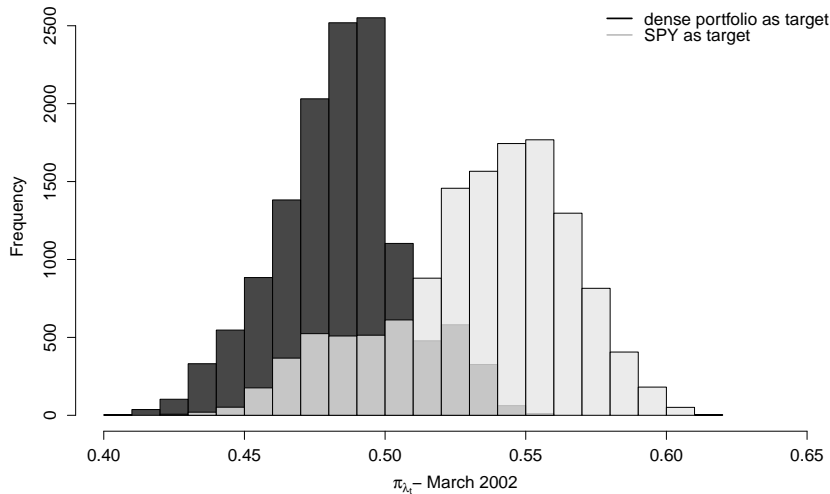
Figure 5: Histograms of the satisfaction probabilities (π_{λ_t}) for two target decisions: The dense portfolio (black) and SPY (i.e., sparse target, in gray). These are shown for the March 2002 investing period and are distributions across all 12,950 enumerated sparse decisions.

Figure (5) aids in the proper choice of the regret threshold κ. When evaluated under the next period cumulative wealth utility, all long only portfolio decisions are somewhat similar. Thus, the regret (difference in loss) will generally be gathered around values close to zero, and the satisfaction probabilities will be gathered around value close to 0.5. As a next step, we select a κ and present the resulting dynamic portfolio decision for the dense target and SPY target.

We show the selected sparse decisions with the $\kappa = 45\%$ threshold for the dense (portfolio optimized over all funds in white columns) and sparse (portfolio of only SPY in gray columns) target decisions in Table (2). Each of these decisions possesses the property that, at each point in time, the satisfaction probability of investing in this decision versus the respective target is at least 45%. The portfolios are updated on a monthly basis; the table displays annual weights for brevity. There are many decisions that will satisfy this threshold for each of the target (see for example the probability mass above 0.45 in Figure (5)). In this case, we have added flexibility in which sparse decision to choose. We construct

25

the sparse decisions so that at most one fund is selected or removed from month to month. For example, if the current portfolio at time t has SPY, OEF, and IVV, two admissible portfolios at $t+1$ could include the funds be SPY, OEF, IVV, and EWG *or* SPY and OEF assuming they both also satisfy the κ threshold.

Dates	SPY		EZU		EWU		EWY		EWG		EWJ		OEF		IVV		IVE		EFA		IWP	
2003	25	75	-	-	58	-	-	-	-	-	-	-	8.3	-	-	-	-	-	-	-	-	8.3
2004	25	75	-	-	43	-	-	-	-	-	20	-	-	-	6.2	-	-	-	-	-	-	-
2005	25	75	-	-	25	-	-	-	6.2	-	13	-	8.3	-	-	-	-	-	-	-	-	-
2006	62	75	-	-	-	-	-	-	6.2	-	19	-	12	-	-	-	-	-	-	-	-	-
2007	75	75	-	-	-	-	25	-	-	8.3	-	-	-	-	-	-	-	-	-	-	-	8.3
2008	44	75	-	-	-	-	-	12	8.3	13	21	-	-	-	-	-	-	-	26	-	-	-
2009	30	45	-	-	-	-	6.2	-	-	-	41	-	-	-	-	34	-	17	21	6.3	-	-
2010	75	55	-	-	-	-	8.3	-	-	-	-	-	-	-	-	26	-	-	11	8.3	-	-
2011	58	57	-	-	25	-	-	-	-	-	-	-	-	-	-	26	-	-	-	8.3	-	-
2012	29	25	8.3	-	-	-	-	-	-	-	54	-	56	-	6.2	-	-	-	-	-	-	-
2013	34	25	-	-	-	-	-	6.2	-	6.2	49	-	56	-	-	-	-	-	-	-	8.3	-
2014	25	75	-	-	-	-	-	-	-	-	37	-	26	-	-	-	-	-	6.2	-	-	-
2015	45	25	-	-	-	-	-	-	-	-	39	36	27	-	-	-	8.3	-	-	6.2	8.3	-
2016	35	75	-	-	-	-	-	-	-	-	40	-	-	-	-	17	-	8.3	-	8.3	8.3	8.3

Dates	IWR		IWF		IWN		IWM		IYW		IYR		RSP		EEM		IWO		IWV	
2003	-	-	-	-	8.3	-	-	-	-	8.3	-	-	8.3	-	-	-	-	-	-	-
2004	-	-	-	12	-	-	-	-	-	12	-	-	6.2	-	-	-	-	-	-	-
2005	-	-	-	8.3	-	-	-	-	-	8.3	30	-	-	-	-	-	-	-	-	-
2006	-	-	6.3	-	-	-	6.2	-	-	12	-	-	-	-	-	-	-	-	-	-
2007	-	-	-	-	-	-	-	-	-	8.3	-	-	-	-	-	-	-	-	-	-
2008	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2010	-	-	-	-	-	-	-	-	8.3	8.3	-	-	-	-	-	-	-	-	-	-
2011	-	-	-	-	-	-	-	-	8.3	8.3	-	-	-	-	-	-	-	-	-	8.3
2012	-	-	-	6.2	-	-	-	-	8.3	-	-	-	-	-	-	-	-	-	-	6.3
2013	-	-	-	-	-	-	-	-	8.3	-	-	-	-	-	6.3	-	-	-	-	-
2014	6.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12	-	12	-	-
2015	-	-	-	-	-	-	-	-	-	6.2	-	-	-	-	-	-	-	-	-	-
2016	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 2: Sparse portfolio decisions (in percent) for DLMs (5) and (6) with $\delta_F = \delta_\epsilon = 0.97$ and $\delta_c = \delta_\beta = 0.9925$. Shown are the selected portfolio decisions for the two targets: dense portfolio (left column in white) and SPY (right column in gray). Note that annual weights are shown for brevity although portfolios are updated monthly. In this dynamic portfolio selection, the regret threshold is $\kappa = 45\%$ for both targets.

In Table (2), the sparse decision for the SPY target has larger allocations to SPY over the trading period compared with the sparse decision for the dense portfolio target. Also, it possesses a consistent allocation to the US technology sector fund IYW. In contrast, the

sparse decision for the dense target often possesses a significant allocation to the Japanese equity specific fund EWJ.

Figure (6) displays the evolution of the regret distributions for the sparse decisions shown in Table (2). The lines are the expected regret, and the surrounding areas correspond to the centered 60% posterior credible intervals. The expected regret for both decisions remains close to zero and for most investing periods is slightly above zero; this is by construction since we choose the sparse decision that satisfies the $\kappa = 45\%$ threshold at each point in time. Overall, these decisions do not result in much regret. Indeed, many of the enumerated long only decisions appear similar in terms of the next period log wealth utility.
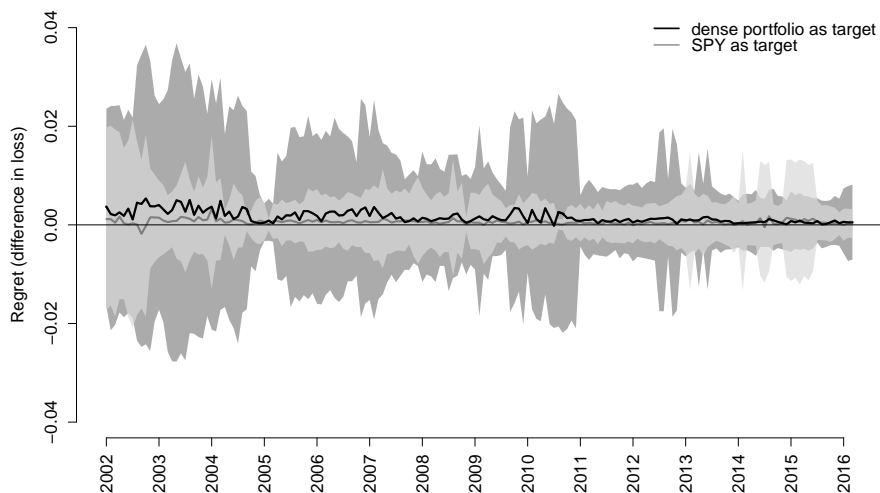

Figure 6: The evolution of the ex ante regret distributions for the sparse decisions in Table (2) versus the two targets: dense portfolio (black) and SPY (gray). The mean regret is shown by the lines and the surrounding areas represent the evolving centered 60% posterior credible intervals.

The variance of the regret distributions in Figure (6) changes substantially over the investing period. The range of log cumulative wealth difference for the "dense portfolio as target" at the beginning is large (~ 0.98 to 1.02 on the cumulative wealth scale). The sparse decision for the dense target collapses in variance around 2005 exactly when the

sparse decision is very close to the dense portfolio. Notice also that the variance of regret for the sparse decision with SPY as the target is, in general, smaller than the dense target decision. Since all of the enumerated decisions have at least 25% of the portfolio allocated to SPY (the target itself) with other diversifying positions, it is intuitive that the uncertainty in regret should be smaller.

The evolution of regret for the sparse decision with SPY as the target sheds light on another question: Are there diversification benefits of allocating to other funds in consideration? Selecting among the 12,950 sparse decisions including up to four non-SPY funds, the expected regret appears to be essentially zero (see the gray line in Figure (6)). This analysis suggests that under the log cumulative wealth utility and considering the large set of enumerated decisions defined at the beginning of this section, the best sparse decision from an *ex ante* perspective may be SPY itself!

The ex ante evolution of other metrics, such as the portfolio Sharpe ratio, may be studied for the sparse decisions displayed in Table (2). The Sharpe ratio is not a utility since it is not a function of future returns $\tilde{R}_t$. However, it is a function of our model parameters whose uncertainty is characterized by the posterior distribution. Specifically, define the predictive portfolio Sharpe ratio:

$$
\begin{aligned}
\mathcal{SR}(w_t, \Theta_t) &= w_t^T \mu_t / (w_t^T \Sigma_t w_t)^{1/2}, \\
\rho_{\mathcal{SR}}(w_{\lambda_t}^*, w_t^*, \Theta_t) &= \mathcal{SR}(w_t^*, \Theta_t) - \mathcal{SR}(w_{\lambda_t}^*, \Theta_t),
\end{aligned}
\tag{15}
$$

where $\rho_{\mathcal{SR}}(\cdot)$ is predictive in the sense that future returns $\tilde{R}_t$ conditional on the model parameters are integrated out. This portfolio metric differs from the Kelly criterion loss in that it focuses on a ratio of the portfolio expected return and variance. It may be used as an exploratory tool to accompany selection from regret-based portfolio selection.

We utilize this "difference in Sharpe ratio" distribution in an exploratory fashion in Figure (7) shown on an annualized scale. The evolution of the difference in Sharpe ratio

is similar to the regret in Figure (6). In this case, a larger positive difference in Sharpe ratio means the selected sparse decision possesses a smaller return-risk tradeoff compared to the target decision. The sparse decision for the dense target is larger variance and trends around larger positive values compared with the sparse decision for the SPY target. The rationale for these features is similar: The enumerated sparse decisions are constructed to contain SPY, so the Sharpe ratios (like the loss) of the sparse decision and the SPY target decision will often be close. Following the financial crisis around 2009, the difference in Sharpe ratio stabilizes at lower values for both sparse decisions.
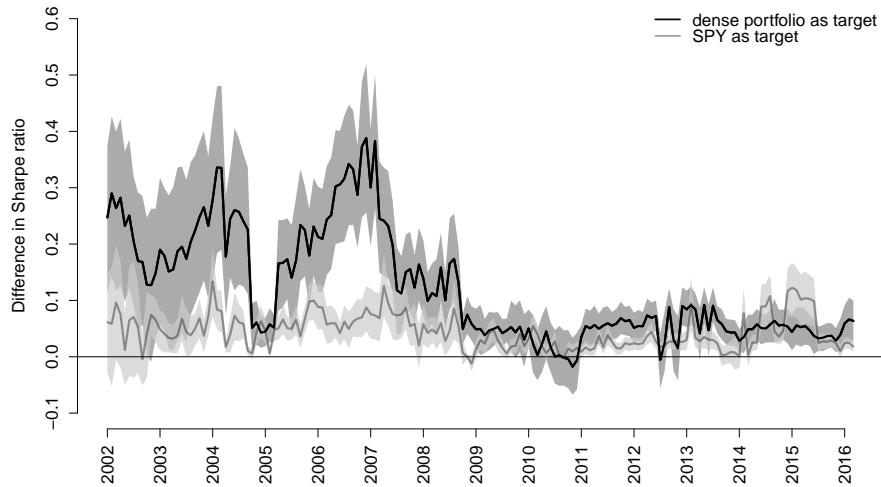

Figure 7: The evolution of the ex ante "Difference in annualized Sharpe ratio" distributions for the sparse decisions in Table (2) versus the two targets: dense portfolio (black) and SPY (gray). The mean regret is shown by the lines and the surrounding areas represent the evolving centered 60% posterior credible intervals.

3.4.1 What happens when κ is varied?

The selection of dynamic portfolio decisions will change based on the regret threshold κ. In Figure (8), we show how expected regret and difference in Sharpe ratio (on an annualized scale) change for selected sparse decisions using the SPY target. The evolution of these metrics is shown for sparse decisions constructed using three κ thresholds: $\kappa = 45\%$ (black),

50% (dark gray), and 55% (light gray). The black lines in both figures correspond to the "SPY as target" paths in Figures (6) and (7), now compared to other κ choices.

Since κ is a lower bound on the satisfaction probability, increasing this lower bound should lead to dynamic sparse decisions with generally smaller regret. In other words, if the investor would like to be satisfied with higher probability, a "lower regret"-sequence of sparse decisions should be selected. Figure (8) demonstrates this when SPY is the target. Larger κ generally lead to smaller expected regret and difference in Sharpe ratio paths. The $\kappa = 55\%$ sparse decision leads to expected regret and difference in Sharpe ratio that are mostly negative from 2002 through 2016, indicating that portfolios with SPY and diversifying funds may be preferred to just SPY alone at this high satisfaction threshold. However, these differences in expectation are still close to zero and small, especially for the evolution of expected regret.
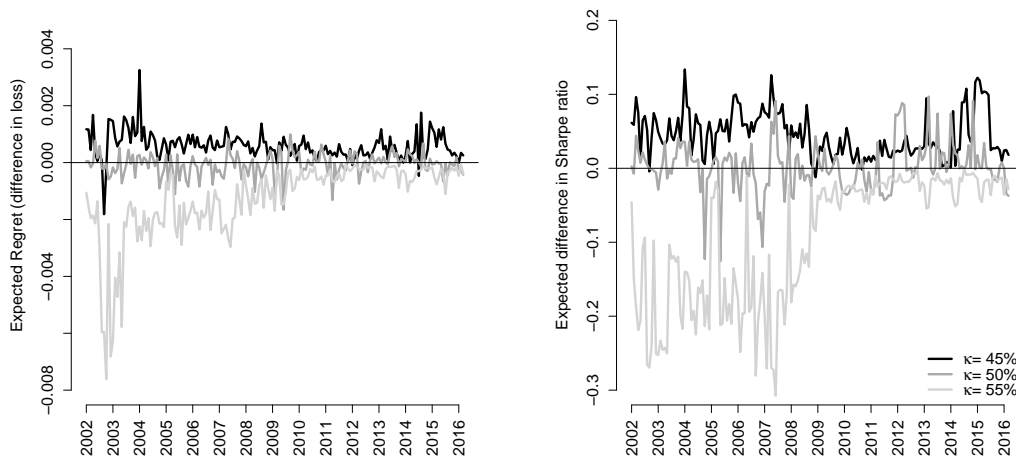

Figure 8: Expected regret (left) and expected difference in Sharpe ratio (right) for sparse decisions with SPY as the target. These metrics are shown for three regret thresholds (the lower bound on probability of satisfaction): $\kappa = 45\%$ (black), 50% (dark gray), and 55% (light gray). Note that as the lower bound on the probability of satisfaction increases, both the expected regret and difference in Sharpe ratio tend to decrease for the selected sparse decisions.

3.4.2 Enumerated decisions without using the utility

The enumerated decisions considered up to this point are constructed by optimizing the integrated approximate loss. An investor might prefer to construct decisions without any consideration for utility and statistical model. An equal-weighted portfolio (where each of the N assets is given weight $1/N$) is one such example of a "utility agnostic" decision. Financial practitioners often advocate for this decision because of its out of sample performance and purity in not involving errors in the statistical model and optimization procedure (DeMiguel et al., 2007). The regret-based procedure can readily accommodate a set of decisions with these characteristics as well.

In the following analysis we consider the set of sparse enumerated equal-weighted portfolios with up to four funds. This amounts to $\Sigma_{i=1}^{4} \binom{25}{i} = 15,275$ decisions to choose among at each point in time. We choose the "dense $1/N$" portfolio as the target decision. This target has 4% invested in each of the 25 funds. To remain consistent with the previous analysis, we consider selection when $\kappa = 45\%$.

The weights for the selected portfolio decision are shown in Table (3). At the $\kappa = 45\%$ threshold, all portfolios have either three or four funds included, and the portfolio decisions have sustained exposures to EWJ (Japanese equity) and IYW (technology) throughout the investing period.

Dates	SPY	DIA	EZU	EWU	EWY	EWJ	OEF	EEM	IVE	EFA	IWP	IWR	IWF	IWO	IWM	IYW	XLI
2003	-	-	-	-	-	-	-	-	-	25	25	-	-	-	25	25	-
2004	-	-	-	-	33	33	-	-	-	-	-	-	-	-	-	33	-
2005	-	-	-	-	25	-	25	-	-	-	-	-	-	-	-	25	25
2006	-	-	-	33	-	-	-	-	-	-	-	33	-	-	-	33	-
2007	-	-	-	33	-	-	-	-	-	-	33	-	33	-	-	-	-
2008	-	-	25	-	-	-	25	-	-	-	25	-	25	-	-	-	-
2009	-	33	-	-	-	33	-	-	-	-	-	-	-	-	-	33	-
2010	-	-	-	-	-	33	-	33	-	-	-	-	-	-	-	33	-
2011	-	-	-	-	-	-	-	33	-	33	33	-	-	-	-	-	-
2012	-	-	-	33	33	-	33	-	-	-	-	-	-	-	-	-	-
2013	-	-	-	-	-	-	33	-	33	-	-	-	-	33	-	-	-
2014	33	-	-	-	-	-	-	-	-	33	-	-	-	33	-	-	-
2015	-	-	25	-	-	25	-	-	25	-	-	-	-	-	25	-	-
2016	-	-	-	33	-	33	-	-	33	-	-	-	-	-	-	-	-

Table 3: Sparse portfolio decision (in rounded percent) for DLMs (5) and (6) with $\delta_F = \delta_\epsilon = 0.97$ and $\delta_c = \delta_\beta = 0.9925$. Each point in time represents an equal-weighted portfolio and corresponding λ_t such that the decision satisfies the $\kappa = 45\%$ threshold. The target decision is the equal-weighted portfolio of all 25 funds – also known as the dense $1/N$ portfolio. Note that annual weights are shown for brevity although portfolios are updated monthly.

This approach to equal-weighted portfolio selection possesses innovative and important features that should be highlighted. While traditional "$1/N$" approaches avoid the investor's utility and model for future asset returns altogether, the regret-based procedure still accounts for both. The decisions themselves may be constructed without a utility or model in mind, but the characterization of regret must involve the utility and model. Regret is computed by the difference in *utility* between the target and sparse decisions, and it is a random variable that may be simulated using the *model* for future asset returns. Regardless of the set of decisions considered by the investor, the utility and model will always play a crucial role in a regret-based selection procedure.

3.4.3   Ex post decision analysis

In this section, we consider the realized performance of the three sparse portfolio decisions presented in Tables (2) and (3) relative to their target decisions. We present out of sample statistics for the six decisions in Table (4). Shown is the annualized Sharpe ratio, standard deviation of return, and mean return.

The sparse enumerated decision for the dense target performs similarly to the dense target. This is comforting – this sparse decision is a dynamic portfolio that is allocated to the market (SPY) and at most four other diversifying positions, and its out of sample performance is comparable to the dense portfolio optimized over all 25 funds at each time.

	out-of-sample statistics		
	Sharpe ratio	s.d.	mean return
sparse enumerated - dense as target	0.40	14.98	6.02
dense	0.45	14.41	6.47
sparse enumerated - SPY as target	0.43	14.65	6.28
SPY	0.43	14.63	6.28
sparse EW enumerated - dense $1/N$ as target	0.49	16.71	8.15
dense $1/N$	0.44	16.47	7.32

Table 4: Comparison of out of sample statistics for the six portfolio strategies considered over the investing period February 2002 to May 2016. The three solid lines correspond to the sparse portfolio decisions presented in Tables (2) and (3). The three dotted lines correspond to the target decisions used for the regret-based selection procedure. All statistics are presented on an annualized scale. "EW" refers to the equal-weighted portfolio decision.

The sparse enumerated decision for the SPY target is equally interesting. Since the SPY target is a sparse decision itself, comparison of it with selected sparse decisions provides insight into whether or not one should diversify away from investing in just the *Market*. The out of sample performance shown in rows three and four of Table (4) display similar performance of the sparse enumerated decision and SPY. Even after considering 12,950 sparse decisions containing up to four funds other than SPY – the diversification benefits of exposure beyond SPY are negligible. The decisions that are *ex ante* better than SPY with 45% probability turn out to help out little *ex post*. Note that this conclusion is with respect to the next period cumulative wealth utility. Future work will involve consideration of other utilities and compare how their selection and ex ante/post analyses differ.

While the sparse optimal strategies both underperform their targets, the sparse equal-weighted strategy slightly outperforms its dense $1/N$ target. This is shown in rows five and six of Table (4). Interestingly, its out of sample performance even exceeds its sparse

optimal counterparts shown in rows one and three in the Table.

4 Discussion

This paper presents a new approach to portfolio selection based on an investor-specified regret tolerance. A loss function defined by the expected portfolio growth rate is used in tandem with a new procedure that separates statistical inference from selection of a sparse dynamic portfolio. We illustrate the procedure using a set of exchange-traded funds. After analyzing two target decisions: (i) A dense portfolio of all available assets, and (ii) A portfolio comprised of a single market fund, we find that selected sparse decisions differ little from their targets in terms of utility; especially after taking into account uncertainty. This finding persists ex post, and a variety of sparse decisions perform similarly to their target decisions on a risk adjusted return (or Sharpe ratio) basis.

The procedure offers a fresh approach to portfolio selection. While traditional approaches typically focus on either the careful modeling of parameters or the optimization procedure used to calculate portfolio weights, regret-based selection combines both through analysis of the regret random variable. Portfolio decisions that are *parsimonious* in nature are then evaluated in a framework that incorporates *uncertainty* in the investor's utility.

Areas of future research include alternative utility specifications. Two relevant examples are: (i) incorporation of fees and (ii) minimization of transaction costs. In each case, a variant of Loss function (10) may be considered. Fees of the funds can be incorporated directly into the vector of future returns. For example, suppose a vector of expense ratios (percentage fee charged of total assets managed) of all funds were given by τ. The vector of future returns within the loss function may be adjusted by τ to reflect an investor's

sensitivity to fees:

$$\mathcal{L}(w_t, \tilde{R}_t) = -\log\left(1 + \sum_{k=1}^{N} w_t^k(\tilde{R}_t^k - \tau_k)\right), \tag{16}$$

where $\tilde{R}_t^k - \tau_k$ is the net return on investment in for fund k.

Sensitivity to transaction costs can be similarly accounted for by modifying the complexity (penalty) function Φ. This can be accomplished by penalizing the difference in consecutive weight vectors through time, $w_t - w_{t-1}$. An example penalty function would look like:

$$\Phi(\lambda_t^1, \lambda_t^2, w_t) = \lambda_t^1 \|w_t\|_1 + \lambda_t^2 \|w_t - w_{t-1}\|_1. \tag{17}$$

This penalty is designed to encourage sparsity as well as slow movement in portfolio positions over time so as to avoid frequent buying and selling of assets. It poses an interesting challenge since there are two penalty parameters (λ_t^1 and λ_t^2) that must be chosen. This is precisely where the regret-based framework has merit. Portfolio decisions indexed by these two penalties can be mapped to a digestible single probability of regret. Then, selection of an appropriate $\{\lambda_t^1, \lambda_t^2\}$ pair can be done in this intuitive "regret space".

The remainder of this section takes a step back and discusses the modularity and important features of regret-based portfolio selection. The methodology is intended to be general – the particular loss, model and dataset used for the empirical analysis are only chosen to demonstrate how the procedure works in practice. The primitive components are: (i) a utility function characterizing investor preferences, (ii) a complexity function measuring how "simple" a portfolio decision is, (iii) a statistical model, and (iv) the investor's regret tolerance; where regret is defined as a difference in utility. The regret tolerance stitches together the first two primitives by answering the question: How does the investor view

the *tradeoff* between her utility and portfolio complexity? Using the utility and posterior distribution defined by the statistical model (primitive three), one can construct a mapping between a set of penalty parameters $\{\lambda_t\}$ and probabilities of regret (as displayed by the right vertical axis in Figure (2)). However, this is not enough. The collection of portfolio decisions indexed by λ_t must be distilled down to one. The fourth primitive accomplishes this by placing an upper bound on the probability of regret; a portfolio that satisfies this upper bound is selected. By incorporating the four primitives, the main (and surprising) feature of this methodology is that a *static* regret threshold produces a sequence of *dynamic* portfolio decisions, one for each investing period.

A Penalized optimization

In this supplement, we show how to minimize a penalized loss function of the following form:

$$\mathbf{L}_{\lambda_t}(w_t) = \frac{1}{2} w_t^T \overline{\Sigma}_t w_t - w_t^T \overline{\mu}_t + \lambda_t \left\| w_t \right\|_1. \tag{18}$$

Completing the square in w_t, dropping terms that do not involve the portfolio weights, and defining $\overline{\Sigma}_t = L_t L_t^T$, we rewrite the first two terms of the loss function as an l_2-norm:

$$
\begin{aligned}
\mathbf{L}_{\lambda_t}(w_t) &= \frac{1}{2} w_t^T \overline{\Sigma}_t w_t - w_t^T \overline{\mu}_t + \lambda_t \left\| w_t \right\|_1 \\
&\propto \frac{1}{2} (w_t - \overline{\Sigma}_t^{-1} \overline{\mu}_t)^T \overline{\Sigma}_t (w_t - \overline{\Sigma}_t^{-1} \overline{\mu}_t) + \lambda_t \left\| w_t \right\|_1 \\
&= \frac{1}{2} (w_t - L_t^{-T} L_t^{-1} \overline{\mu}_t)^T L_t L_t^T (w_t - L_t^{-T} L_t^{-1} \overline{\mu}_t) + \lambda_t \left\| w_t \right\|_1 \\
&= \frac{1}{2} \left[L_t^T (w_t - L_t^{-T} L_t^{-1} \overline{\mu}_t) \right]^T \left[L_t^T (w_t - L_t^{-T} L_t^{-1} \overline{\mu}_t) \right] + \lambda_t \left\| w_t \right\|_1 \\
&= \frac{1}{2} \left\| L_t^T w_t - L_t^{-1} \overline{\mu}_t \right\|_2^2 + \lambda_t \left\| w_t \right\|_1.
\end{aligned}
\tag{19}
$$

The optimization of interest is written as:

$$\min_{w_t \geq 0} \quad \frac{1}{2} \left\| L_t^T w_t - L_t^{-1} \overline{\mu}_t \right\|_2^2 + \lambda_t \left\| w_t \right\|_1 . \tag{20}$$

Optimization (20) is now in the form of standard sparse regression loss functions, (Tibshirani, 1996), with covariates, L_t^T, data, $L_t^{-1} \overline{\mu}_t$, and regression coefficients, w_t. We may optimize (20) conveniently using existing software, such as the `glmnet` package of Friedman et al. (2010).

References

Carvalho, C. M., West, M. et al. (2007), 'Dynamic matrix-variate graphical models', *Bayesian analysis* **2**(1), 69–97. 6

CRSP (1992-2016), 'The center for research in security prices', Wharton Research Data Services. 16

DeMiguel, V., Garlappi, L. and Uppal, R. (2007), 'Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?', *The review of Financial studies* **22**(5), 1915–1953. 31

Fama, E. F. and French, K. R. (2015), 'A five-factor asset pricing model', *Journal of Financial Economics* **116**(1), 1–22. 13

Friedman, J., Hastie, T. and Tibshirani, R. (2010), 'Regularization paths for generalized linear models via coordinate descent', *Journal of statistical software* **33**(1), 1. 37

Gron, A., Jørgensen, B. N. and Polson, N. G. (2012), 'Optimal portfolio choice and stochastic volatility', *Applied Stochastic Models in Business and Industry* **28**(1), 1–15. 6

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. 5

Harrison, J. and West, M. (1999), *Bayesian Forecasting & Dynamic Models*, Springer. 13, 14, 15

Irie, K. and West, M. (2016), 'Bayesian emulation for optimization in multi-step portfolio decisions', *arXiv preprint arXiv:1607.01631* . 6

Jacquier, E. and Polson, N. (2010), 'Simulation-based-estimation in portfolio selection'. 6

Johannes, M., Korteweg, A. and Polson, N. (2014), 'Sequential learning, predictability, and optimal portfolio returns', *The Journal of Finance* **69**(2), 611–644. 6

Kelly Jr, J. L. (1956), 'A new interpretation of information rate', *Information Theory, IRE Transactions on* **2**(3), 185–189. 18

Ng, L. (1991), 'Tests of the capm with time-varying covariances: A multivariate garch approach', *The Journal of Finance* **46**(4), 1507–1521. 17

Pettenuzzo, D. and Ravazzolo, F. (2015), 'Optimal portfolio choice under decision-based model combinations', *Journal of Applied Econometrics* . 6

Polson, N. and Tew, B. (1999), 'Bayesian portfolio selection: An analysis of the s&p 500 index 1970-1996', *Journal of Business and Economic Statistics* **18**, 164–173. 6

Puelz, D., Carvalho, C. M. and Hahn, P. R. (2015), 'Optimal etf selection for passive investing', *arXiv preprint arXiv:1510.03385* . 6

Puelz, D., Hahn, P. R. and Carvalho, C. (2017), 'Variable selection in seemingly unrelated regressions with random predictors', *Bayesian Analysis* . 5

Rising, J. K. and Wyner, A. J. (2012), Partial kelly portfolios and shrinkage estimators, *in* 'Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on', IEEE, pp. 1618–1622. 18

Rudd, A. (1980), 'Optimal selection of passive portfolios', *Financial Management* pp. 57–66. 5

Sankaran, J. K. and Patil, A. A. (1999), 'On the optimal selection of portfolios under limited diversification', *Journal of Banking & Finance* **23**(11), 1655–1666. 5

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288. 37

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

Zhao, Z. Y., Xie, M. and West, M. (2016), 'Dynamic dependence networks: Financial time series forecasting and portfolio decisions', *Applied Stochastic Models in Business and Industry* **32**(3), 311–332. 6

Zhou, X., Nakajima, J. and West, M. (2014), 'Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models', *International Journal of Forecasting* **30**(4), 963–980. 6