Applications of Statistical Learning in Asset Allocation (1)

Hongshu Chen

4/21/2021

Topics

- Asset Allocation Frameworks
  - Mean-Variance
  - Mean-CVaR
  - Risk Parity
- Statistical Learning Methods
  - Supervised Learning
    - Time Series Models (AR, MA)
    - Linear Regression (Lasso, Gradient Boosting)
  - Unsupervised Learning
    - Gaussian Mixture Model (GMM)
    - Hidden Markov Model (HMM)
    - K-means clustering
    - Hierarchical clustering
- Case Studies

Asset Allocation

What Is Asset Allocation

- Analyze trade-offs between risk and reward
  - e.g. return vs. volatility
- Make decisions on asset classes weights
  - e.g. US Equity, Non-US Equity, Bonds, Cash
- Strategic vs. Tactic

	Stage of Decision Making	Forecasting Time Horizon	Active Management
Strategic	Early	Long	No
Tactic	Middle/Late	Medium/Short	Yes

Strategic Asset Allocation

- One of the most important investment decisions (Brinson, Hood, and Beebower 1986)
- Inputs:
  - Capital market assumptions
  - Investor characteristics: time horizon, objective, risk tolerance, etc.
- Output:
  - Benchmark asset weights
- Some examples:
  - **Endowment**:
    - Extremely long/infinite horizon, return need to meet spending policy requirements, relatively high risk tolerance
  - **Individuals saving for retirement**:
    - Time horizon up to end of life, return objective depends on financial conditions and projected retirement spending, relatively low risk tolerance.

Tactic Asset Allocation

- The goal is to generate positive return relative to benchmark
- Can be implemented alone or in combination with security selection
- The breadth tends to be smaller relative to security selection
- Signals usually come from macro and/or quantitative analysis and models
- **Not** a focus of this lecture

Mean-Variance Optimization

Problem Statement

$$
\begin{aligned}
arg \min_{\mathbf{w}} \mathbf{w}^T \Sigma \mathbf{w} \\
s.t. \\
\mu^T \mathbf{w} \geq r, \\
\mathbf{1}^T \mathbf{w} = 1, \\
\mathbf{lb} \leq \mathbf{w} \leq \mathbf{ub}
\end{aligned}
$$

where μ is the asset class expected return, Σ is the variance-covariance matrix of asset class returns, r is the return target, $\mathbf{lb}$ and $\mathbf{ub}$ are the lower bounds and upper bounds of asset weights respectively.

μ and Σ (Capital Market Assumptions) should be forward looking and match the investment investment horizon.

Challenge #1: Smoothed Asset Return

Why Smoothed Return Is a Problem

- In practice, Σ is often estimated based on historical asset returns
- The simple historical variance-covariance matrix is problematic if there is spurious autocorrelation in the historical data because of return smoothing
- Reported assets returns may be smoothed for varieties of reasons, such as lack of trading activities (a common problem for illiquid assets)
- Return smoothing leads to underestimation of volatility and over allocation to the asset

Lower Volatility $\Rightarrow$ Higher Allocation

Let's simplify the previous mean-variance optimization problem as maximizing the the utility $L = \mu^T \mathbf{w} - \frac{\lambda}{2}\mathbf{w}^T \Sigma \mathbf{w}$ where λ is the risk aversion parameter. To maximize L, we take the derivative to $\mathbf{w}$ and let it equals to zero:

$$\frac{\partial L}{\partial \mathbf{w}} = \mu - \lambda \Sigma \mathbf{w} = 0$$

then we get:

$$\mathbf{w} = \frac{1}{\lambda}\Sigma^{-1}\mu$$

If we further assume that Σ is a diagonal matrix, i.e., the cross assets correlations are zero, we can see the optimal asset weight is inversely proportional to asset variance.

Geltner Unsmoothing

Geltner (1993) assumes that the reported return r_t^* is a weighted sum of actual return r_t and previous reported return r_{t-1}^*:

$$r_t^* = (1-\alpha)r_t + \alpha r_{t-1}^*$$

If

$$r_t = \mu + \epsilon_t$$

where ϵ_t is the $i.i.d.$ noise term, then

r_t^* follows an AR(1) process;

α can be estimated as the first order autocorrelation of r_t^*, ρ.

After substituting ρ for α and rearrange the equation above, we get:

$$r_t = \frac{r_t^* - \rho r_{t-1}^*}{1-\rho}$$

GLM Unsmoothing

Getmansky, Lo, and Makarov (2004) suggested a different approach (henceforth referred to as GLM) to unsmooth illiquid assets return based on the assumption that the reported return is the moving average (MA) of actual returns:

$$r_t^* = \sum_{i=0}^{k} \theta_i r_{t-i},$$

and

$$\sum_{i=0}^{k} \theta_i = 1.$$

Once $\theta_{0,1,\ldots,k}$ are estimated, we can recursively estimate the actual returns as:

$$r_t = \frac{r_t^* - \sum_{i=1}^{k} \theta_i r_{t-i}}{\theta_0}$$

A Simulated Example

Assets A, B, C and D returns follows an $i.i.d.$ multivariate normal distribution.

Return and Volatility

	return	volatility
A	0.10	0.20
B	0.04	0.04
C	0.01	0.01
D	0.05	0.15

Correlation

	A	B	C	D
A	1.0	0.2	0	0.5
B	0.2	1.0	0	0.1
C	0.0	0.0	1	0.0
D	0.5	0.1	0	1.0

A Simulated Example (Cont.)

- We further assume asset D's reported returns follow the Geltner (1993) process with $\alpha = 0.5$
- Based on the assumptions above, we generated 50 years monthly simulated actual assets return $\mathbf{r}$ and reported returns $\mathbf{r}^*$
- $\mathbf{r}$ and $\mathbf{r}^*$ only differ in asset D returns

A Simulated Example (Cont.)

Volatility & correlation comparison:

```
Sample annulized volatility of actual and report return of asset D are:

     D     D*
0.1487 0.0877

Sample correlation of actual and report return of asset D with other assets are:

      A    B     C
D  0.47 0.15 -0.01
D* 0.41 0.12 -0.02
```

A Simulated Example (Cont.)

- Assume the minimal required return is 3%.
- Run mean-variance optimization using the same μ (population mean) and three different versions of Σ:

  1. The population variance-covariance matrix
  2. The sample variance-covariance matrix based on actual returns
  3. The sample variance-covariance matrix based on reported returns

A Simulated Example (Cont.)

We get the mean-variance optimal portfolio weights:

Portfolio Weights

	1	2	3
A	0.04	0.04	0.02
B	0.53	0.52	0.46
C	0.42	0.42	0.41
D	0.01	0.01	0.10

A Simulated Example (Cont.)

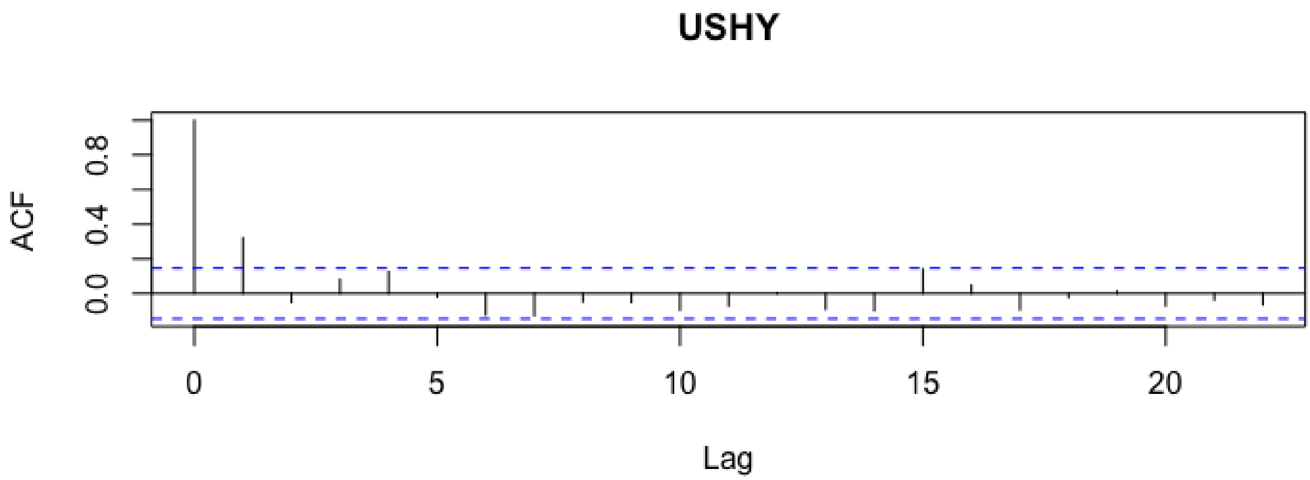
- Results from version 1 are the “true” optimal portfolio weights
- Results from version 2 deviate slightly from the “true” results because of not large enough sample size
- Results from version 3 are off quite a bit because of return smoothing, the allocation to asset D is much larger than the “true” optimal allocation.

An Example With Real Data

- Four assets: US Large Cap Equity (USLE), Emerging Markets Equity (EME), US Investment Grade Bonds (USIG) and US High Yield Bonds (USHY)
- Expected returns are assumed to be 10%, 12%, 4% and 8% respectively
- Variance-covariance matrix is estimated based on 15 years (200603 - 202102) monthly returns

An Example With Real Data

USHY returns exhibit strong autocorrelation:



An Example With Real Data (Cont.)

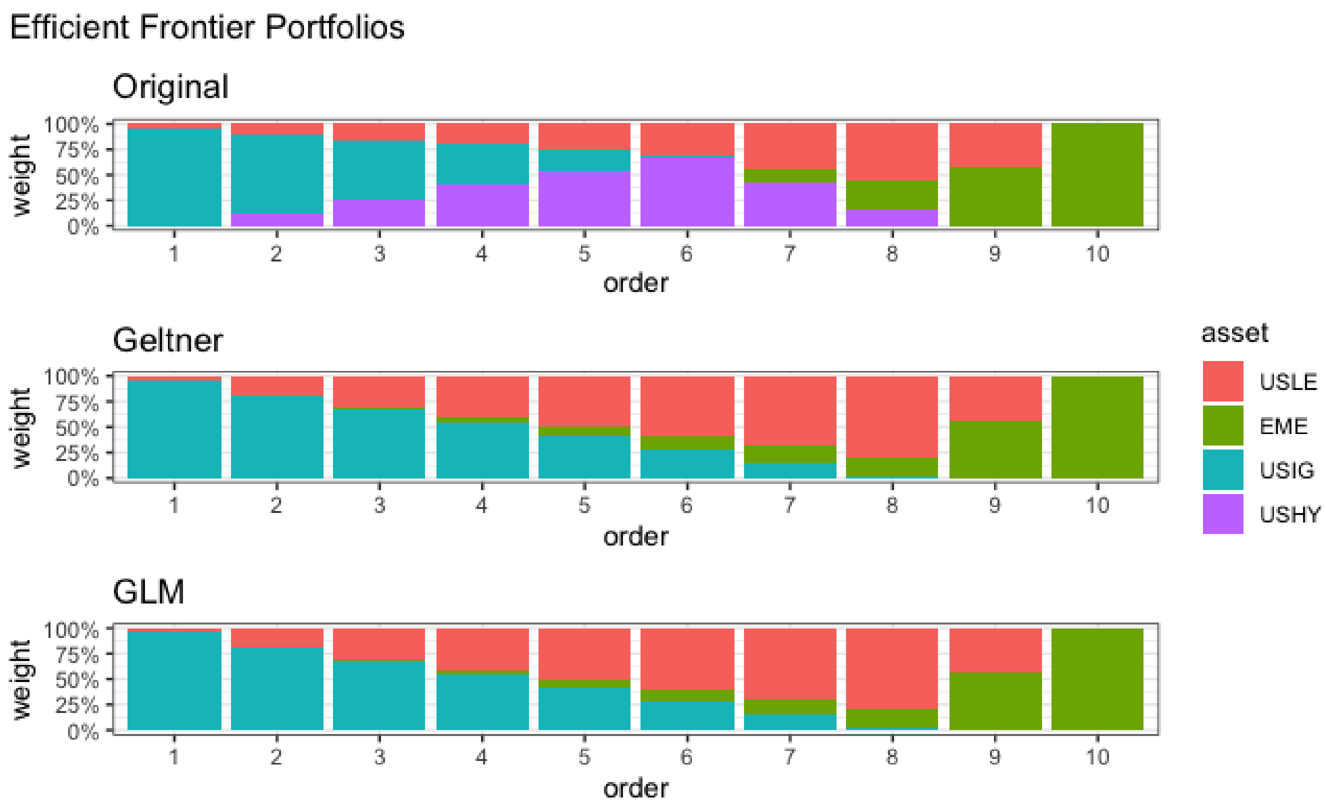
USHY Volatility Estimates

Original	Geltner	GLM
0.1	0.14	0.13

USHY Correlation with Other Assets

	Original	Geltner	GLM
USLE	0.76	0.76	0.75
EME	0.77	0.76	0.75
USIG	0.24	0.34	0.34

An Example With Real Data (Cont.)



An Example With Real Data (Cont.)

- Unsmoothing by Geltner and GLM result in higher volatility estimate for USHY, also higher correlation between USHY and USIG
- Geltner and GLM unsmoothing lead to similar MVO efficient frontier portfolio weights, USHY weights go to zero in all portfolios

Challenge #2: Missing Data

The Missing Data Problem

- A long return history is preferred for estimation of variance-covariance matrix
- Some asset classes have relatively short histories (e.g. TIPS)
- Asset return data may not be available for other reasons
- Ways to deal with missing data
  - Use complete observations
  - Use pairwise complete observations
  - **Impute missing data**

ImputeR

- Expectation-Maximization (EM) Algorithm
  - E-step: missing data are estimated
  - M-step: model parameters are estimated
- The *ImputeR* package (Feng et al. 2018) in R implement the EM algorithm with the flexibility of applying different supervised learning techniques in the M-step
- We are going to use Lasso (Tibshirani 1996) and Gradient Boosting (Hastie, Tibshirani, and Friedman 2009) in the M-step

ImputeR Algorithm

```
1. Sort the data columns based on percentage of missing data, with the first
   column missing the most.
2. Make initial estimations of missing data (E-Step).
3. While Not Converged:
      For j from 1 to p (p = number of columns with missing data):
        M-Step: fit a supervised learning model (y ~ x) to estimate the j-th column data
                based on non-missing j-th column observations (y) and other columns
                of the same entries (x);
        E-Step: update j-th column missing data estimates based on the model
                in previous step and data of other columns.
      End
  End
```

Lasso

- Lasso is a popular linear regression method with built-in feature selection
- It adds a L1 norm penalty term to objective function of Ordinary Least Squares (OLS) regression.

$$arg \min_{\beta} \frac{1}{N} ||\mathbf{y} - \mathbf{X}\beta||_2^2 + \lambda||\beta||_1$$

- Special cases:
  - When the penalty parameter λ equals zero, Lasso is equivalent to OLS,
  - When λ is large enough, all the regression coefficients become zero.

Gradient Boosting

Gradient Boosting builds a strong learner by sequentially adding weaker learners; The *mboost* (Hofner et al., n.d.) R package implements one version of Gradient Boosting algorithm:

```
1. Initialize learner f(x) with the offset term.
2. Specify a set of base-learners (e.g. univariate linear regression model).
3. While Stopping Criterion Not Reached:
      a) Calculate negative gradient vector of loss function (e.g. sum of squared error)
          to learner: -dL/df at f(x).
      b) Fit base-learners to the negative gradient vector.
      c) Choose the best base-learner u(x).
      d) Update learner: f(x) = f(x) + v * u(x), where v is the learning rate.
   End
```

A Case Study With Missing Data

- Eight assets: USLE, DME, EME, USIG, CORP, TIPS, COM, CASH
- Assume assets expected returns are $10\%, 10\%, 12\%, 4\%, 6\%, 4\%, 4\%, 1\%$
- Use 20 years monthly data (200103 - 202102) to estimate variance-covariance matrix
- Five assets with missing data (COM>USIG>EME>CORP>DME)

A Case Study With Missing Data

	# of Missing Observations	Missing Percentage
COM	69	29%
USIG	32	13%
EME	27	11%
CORP	18	8%
DME	7	3%
USLE	0	0%
TIPS	0	0%
CASH	0	0%

A Case Study With Missing Data (Cont.)

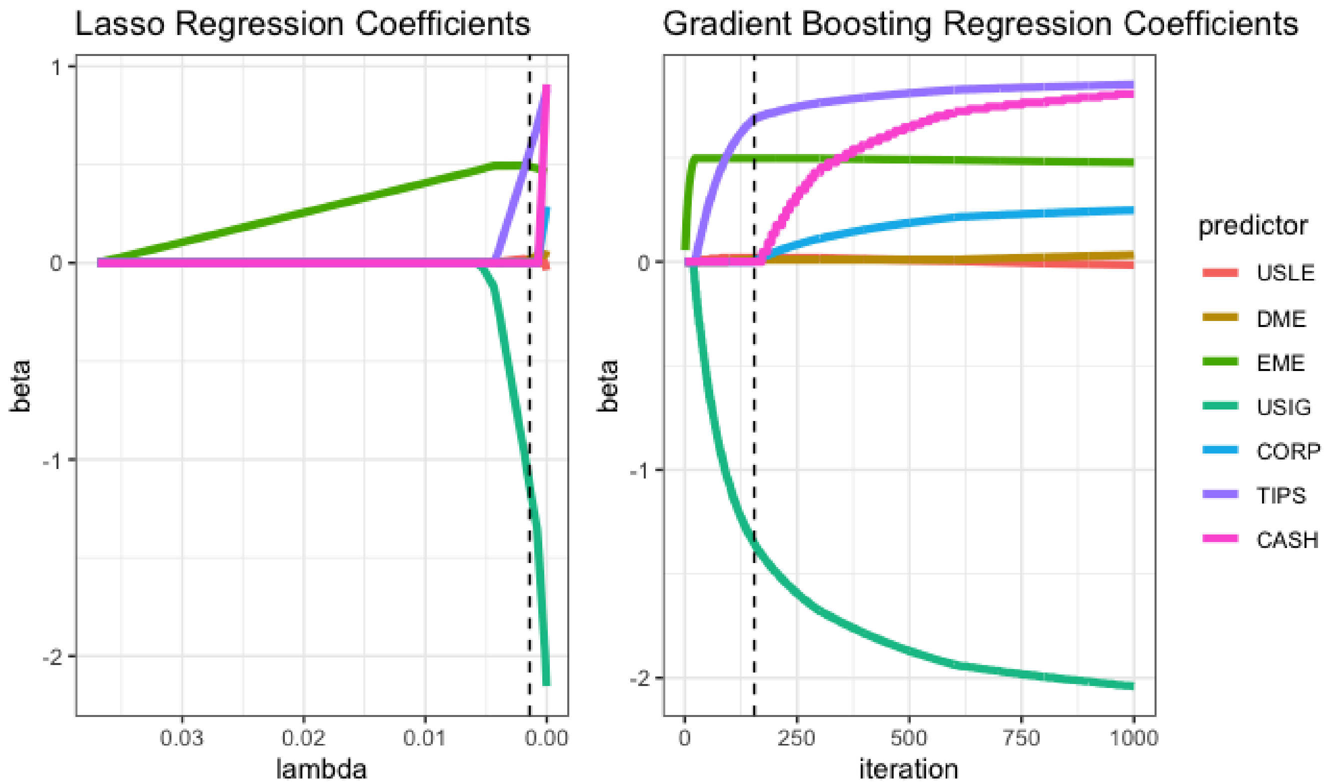
- We start with modeling COM returns
- Models
  - OLS
  - Lasso: use cross validation to select optimal λ
  - Gradient Boosting (GB): use AIC to select optimal number of iterations

A Case Study With Missing Data (Cont.)

Both Lasso and GB ended up with similar model coefficients and selected the same set of predictors:

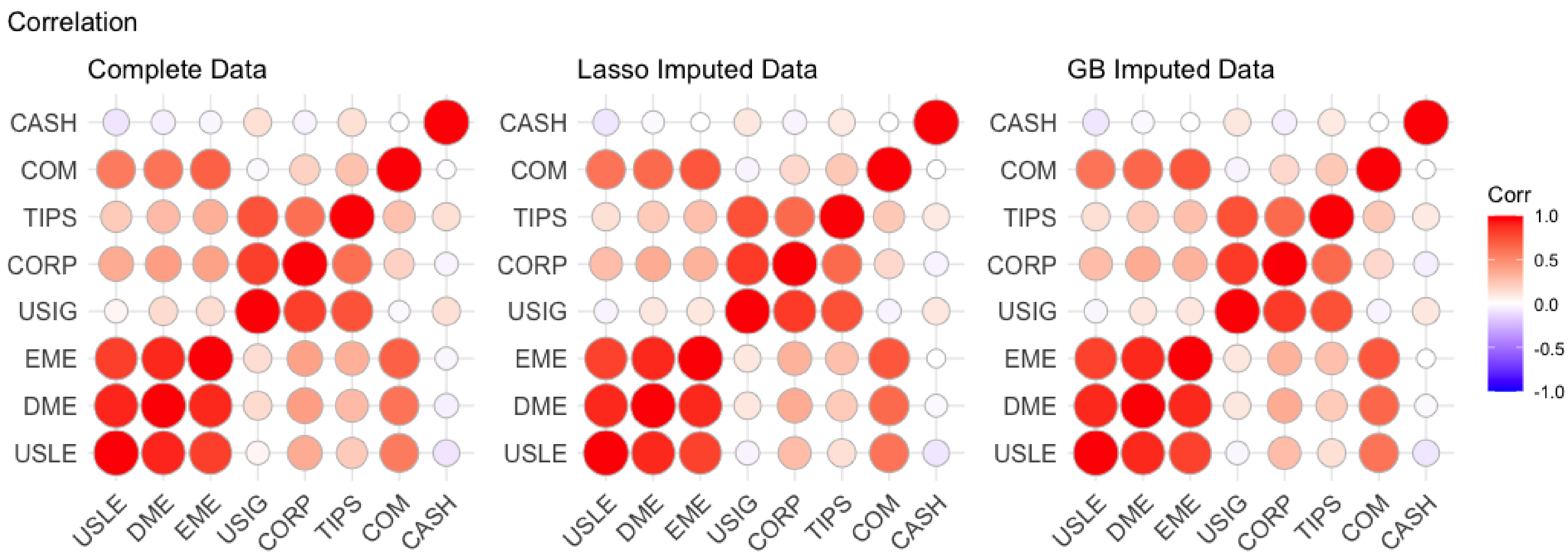
predictor	OLS	Lasso	GB
(Intercept)	-0.0037	-0.0047	-0.0044
USLE	-0.0495	0.0126	0.0195
DME	0.0709	0.0167	0.0095
EME	0.4595	0.4888	0.4968
USIG	-2.2064	-1.1314	-1.3575
CORP	0.3048	0.0000	0.0000
TIPS	0.8943	0.5655	0.6889
CASH	0.9522	0.0000	0.0000

A Case Study With Missing Data (Cont.)



A Case Study With Missing Data (Cont.)

Now let's use *ImputeR* to impute all the missing data, and compare the estimated correlation matrix:



A Case Study With Missing Data (Cont.)

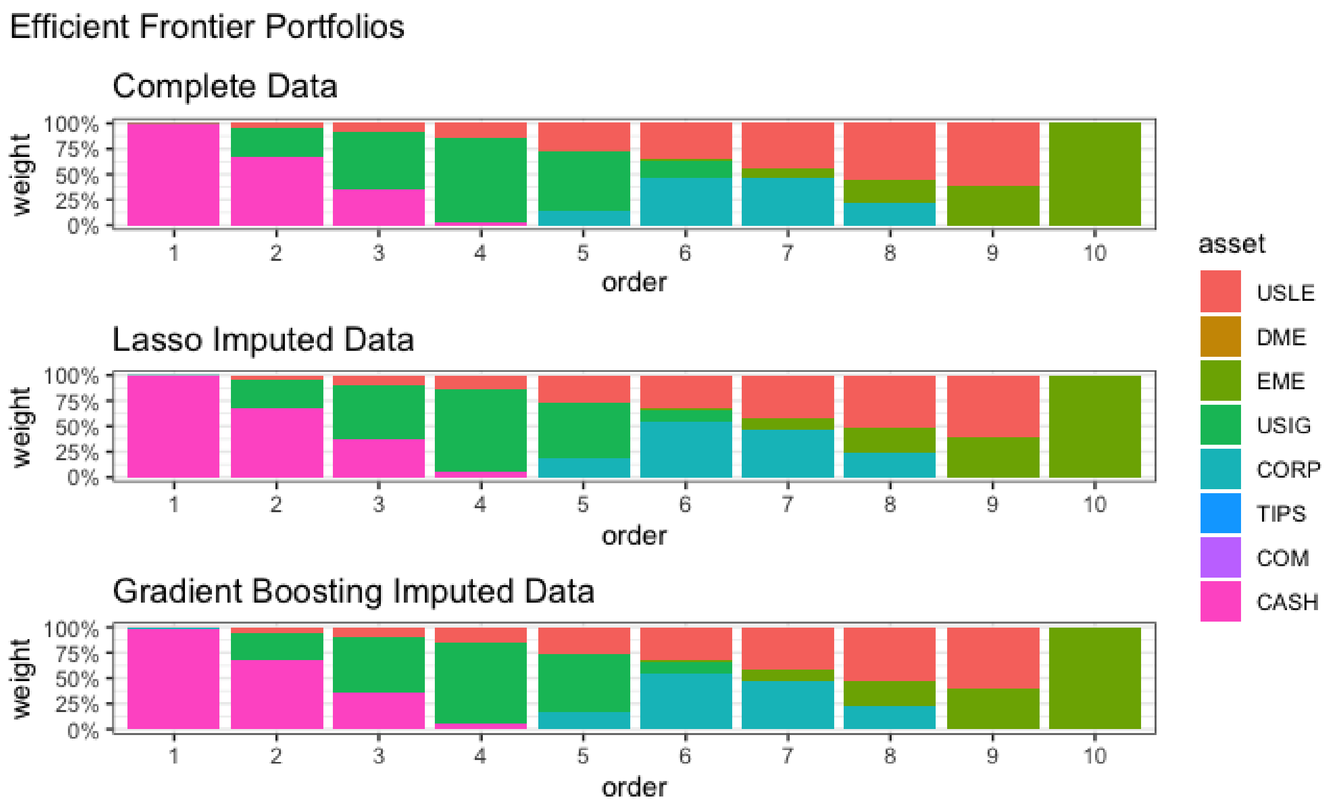
Volatility estimates:

	Complete	Lasso	GB
USLE	0.16	0.15	0.15
DME	0.19	0.18	0.18
EME	0.23	0.22	0.22
USIG	0.04	0.04	0.04
CORP	0.08	0.07	0.07
TIPS	0.07	0.07	0.07
COM	0.19	0.17	0.17
CASH	0.01	0.01	0.01

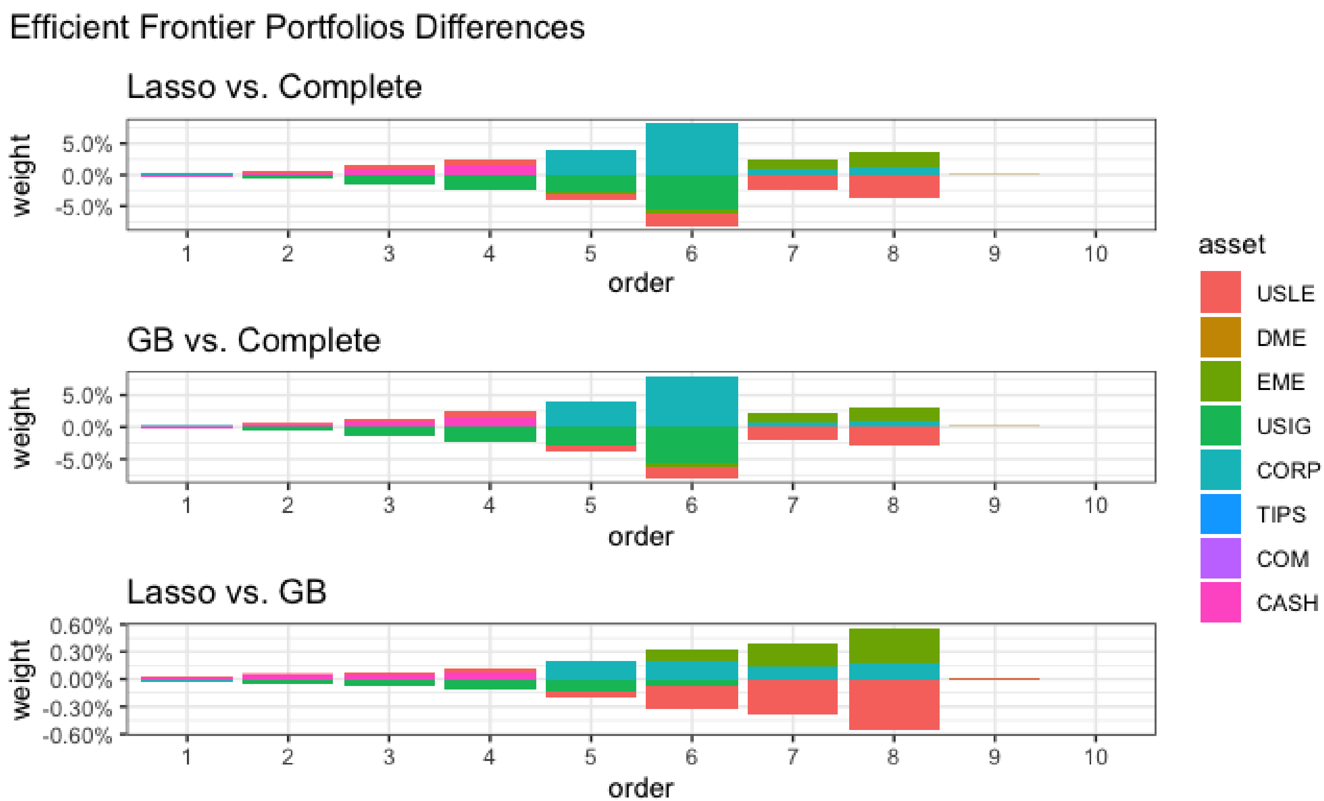
Both Lasso and GB produced similar volatility estimates and are slightly different from volatility estimates by only using the complete data.

A Case Study With Missing Data (Cont.)

Finally, we generate MVO efficient frontier portfolios:



A Case Study With Missing Data (Cont.)



Some Other Things

Resampled Efficient Frontier

- What about the uncertainties of μ and Σ?
- Michaud and Michaud (2008) proposed resampled efficient frontier©

1. Simulate n batches of return data from μ and Σ
2. For $i = 1, \ldots, n$, get $\hat{\mu}_i$ and $\hat{\Sigma}_i$ based on the i-th batch of simulated data
3. For $i = 1, \ldots, n$, get efficient portfolio weights $\mathbf{w}_i$ based on $\hat{\mu}_i$ and $\hat{\Sigma}_i$
4. Calculate resampled efficient portfolio weights as: $\frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i$

Other Common Optimization Constraints/Penalty

- Linear
  - Group weights: $c \leq \mathbf{A}\mathbf{w} \leq d$
  - Risk factor beta: $c \leq \mathbf{F}\mathbf{w} \leq d$
  - Max turnover: $||\mathbf{w}_t - \mathbf{w}_{t-1}||_1 \leq c$
  - Max leverage: $||\mathbf{w}||_1 \leq c$
- Integer
  - Cardinality: $\sum_i I(w_i \neq 0) \leq c$
  - Min position size: $I(|w_i| < c)I(w_i \neq 0) = 0$
- Quardratic
  - Concentration: $||\mathbf{w}||_2$

References

Brinson, Gary P., L. Randolph Hood, and Gilbert L. Beebower. 1986. "Determinants of Portfolio Performance." *Financial Analysts Journal* 42 (4): 39–44. http://www.jstor.org/stable/4478947.

Feng, L., S. Moritz, G. Nowak, AH Welsh, and TJ O'Neill. 2018. "ImputeR: A General Multivariate Imputation Framework." http://github.com/SteffenMoritz/imputeR.

Geltner, David. 1993. "Estimating Market Values from Appraised Values Without Assuming an Efficient Market." *The Journal of Real Estate Research* 8 (3): 325–45. http://www.jstor.org/stable/44095449.

Getmansky, Mila, Andrew W. Lo, and Igor Makarov. 2004. "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns." *Journal of Financial Economics* 74 (3): 529–609. https://doi.org/https://doi.org/10.1016/j.jfineco.2004.04.001.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2nd ed. Springer. http://www-stat.stanford.edu/~tibs/ElemStatLearn/.

Hofner, Benjamin, Andreas Mayr, Nikolay Robinzonov, and Matthias Schmid. n.d. "Model-Based Boosting in R: A Hands-on Tutorial Using the R Package Mboost." *Computational Statistics* 29 (1): 3–35. https://doi.org/10.1007/s00180-012-0382-5.

Michaud, Richard O., and Robert O. Michaud. 2008. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88. https://doi.org/https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.