Homework 3 Write-up

Contributors: Ziheng Chen (zc2068), Cong Wang (cw3140), Chengyang Gu (cg3765)

Summary

We recommend the clipped estimator based on a comparison between four popular covariance cleaning estimates apart from the standard approach. The clipped estimator delivers one of the best portfolio variance reduction, reduces bias significantly, and is yet easily explainable. Additionally, the optimal shrinkage estimator adds to more complexity and reduces the variance further, but introduces more ad hoc bias. This result is consistent based on a rolling-based backtest using high-frequency TAQ data.

Problem 1

(a)

> A description of the methodology you are using to compare the covariance matrix estimators. Please motivate any assumptions you make.

Without specifying which covariance estimator is used, the purpose of this part is to explain how we evaluate this estimation.

We first divide the training and test sets. The training set consists of N/q observations, with N representing the number of stocks and q $(0 < q < 1)$ a pre-defined parameter. Each observation consists of mid-quote returns at any time t for all stocks. In our setting, after cleaning and removing stocks with too many outliers, we have $N = 476$. We choose $q = 0.5$ to retain consistency with BBP's paper. The test set is by default 3 trading days to follow the training set. In practice, we roll the training and test period window forward and perform multiple tests to exhaust all the data.

To evaluate the cleaned covariance matrix, we stick with BBP and use the Markowitz framework. It's important to keep in mind that we only care about the risk (or more precisely, volatility) as this objective which is second-order agrees with our optimization in this task.

Given the training set return dataset $D_t = \{r_{i,t}\}$, we estimate the predicted return $g = g(D_t)$ and the (cleaned) covariance matrix $\hat{\Sigma} = \hat{\Sigma}(D_t)$. The Markowitz portfolio can be constructed as

$$w := \frac{\hat{\Sigma}^{-1} g}{g^T \hat{\Sigma}^{-1} g} \in \mathbb{R}^N$$

With regard to verification, we first check that, by investing on w and compute the variance of P&L in the test set similar as in BBP (see eq. 13 in that paper), we equivalently estimate the covariance matrix Σ_{out} of the test set and compute $w^T \Sigma_{out} w$ as the variance. In other words, the variance of P&L is syntactically consistent with the Markowitz framework.

We provide a brief proof to show that both approaches are theoretically equivalent. Denote the random return vector as X. The covariance matrix of X is $cov(X) = E[(X - EX)^T (X - EX)]$. Stepwise P&L is Xw with $E[Xw] = \bar{X}w$. Thus, the variance of P&L is

$$
\begin{aligned}
var(Xw) &= E(w^T(X-\bar{X})^T(X-\bar{X})w) \\
&= w^T cov(X) w
\end{aligned}
$$

which is exactly the variance objective in the Markowitz framework.

In our experiment, however, we use a simplified variance estimator (using the notation in BBP)

$$
\hat{R}^2(t,w) = \frac{1}{T_{out}} \sum_{\tau=t+1}^{t+T_{out}} \left(\sum_{i=1}^{N} w_i r_{i,\tau} \right)^2
$$

The important assumption here is $E[r_{i,t}] = 0$ for all stocks i and time t. We think it's reasonable because we are working on cleaned high-frequency data, whose distribution is already asymptotically normal with mean close to zero. This contrasts with BBP in that they use daily return as observations. In that case, expected return may be worth considering.

We now explain how returns are predicted, i.e., the form of $g(D_t)$. Inspired by BBP 2016, we focus on three types of predictions:

1. Minimum variance portfolio: we simply let $g_i \equiv 1$. Applying the formula for w above, and recall the Markowitz framework, we know that w is the minimum variance portfolio.
2. Omniscient predictor: we use normalized realized returns as g. Note that in this case, g is not derived from the training set. We choose not to demean and de-variance the predictors because in our code, we directly operate on the covariance matrix. See more details below when we discuss how we estimate and clean covariance matrices.
3. Purely random predictor: we choose $g = \sqrt{N}v$, where v is sampled in the unit sphere, i.e., $||v||_2 = 1$. We follow eq. 16 in this page to generate v.

Finally, we simply compare the $\hat{R}^2$ across different covariance estimators and predictions to decide which estimator(s) delivers consistent improvements compared with the naive covariance matrix as the estimator.

(c)

> Provide an analysis of the covariance matrix estimators and their performance.
>
> Support your analysis with statistics/tables/figures as you see fit (you can get some ideas from the Axioma article).

Markowitz-based Volatility

Our covariance matrix estimators are based on 5-minute mid-quote returns using the TAQ quote high frequency data from June 27 to September 20, 2007. After removing 30 stocks with most missing records, our final data includes $N = 476$ stocks and $T_{total} = 5,005$ observations. As mentioned before, we follow the BBP and set $q = N/T = 0.5$ to estimate the covariance matrix estimators, while the rest of records are used as test dataset.

Regarding the choice of the predictor g, we follow the approach described the section “Optimal Shrinkage and portfolio optimization” and use #1. Minimum Variance #2. Omniscient case and #4. random long-short predictors throughout the experiment.

In terms of evaluating the performance of covariance matrix, we set default out-of-sample set as 3 days or $3 \times 77 = 231$ records and compute number of out-of-sample using formula (11) in BBP

$$n = \lfloor \frac{T_{tot} - T - 1}{T_{out}} \rfloor$$

We use $T = 952$ records as training data to estimate three covariance matrix and roll out-of-sample (length=231) to compute the variance estimator using formula (13) in BBP paper until the data is exhausted. Alternatively, as mentioned before, we also roll both training and test data to evaluate the covariance matrix and they conclude to similar result.

Table 1 (Roll out-of-sample data)

Out-of-sample Portfolio Volatility (%)	Minimum Variance	Omniscient	Random Uniform
Empirical (uncentered)	9.6484452	3.5768163	1.3450456
Clipped	9.2276980	3.1231274	1.2471107
OptShrinkage	9.1939608	3.0602467	1.2275518
Empirical (centered) (benchmark)	9.6527823	3.5774962	1.3461573
EWMA $(\alpha = 0.9)$	83.1517845	101.8867633	205.1721253

The optimal weights from Markovitz formula (14) naturally sums to 1 for Minimum Variance predictor but not always sum to 1 for the other two predictors. We compensate for that by allowing for a risk-free asset, which we assume to have **zero return and zero volatility** in this high-frequency setting. Another way to view it is that, it adds to extra penalty if to impose the constraint $||w||_1 = 1$ for other return predictors other than minimum variance portfolio.

In table 1, Empirical (centered) covariance matrix in the fourth row is our benchmark which is also the uncleaned version of covariance matrix estimator using in-sample training data. Empirical (uncentered) computes the maximum likelihood covariance estimator in a similar way but assuming zero expectation additionally. The clipped covariance estimation “clips” the eigenvalues of an empirical covariance matrix by using PCA to choose an appropriate K eigenvalues in order to provide a cleaned estimator. OptShrinkage follows the method in BBP2016 which is an ad hoc method proposed to correct for a systematic bias for small eigenvalues.

As we observe each column in table 1, the cleaned versions of is effective in the sense of shrinking the benchmark covariance matrix by removing noise in the data. For each predictor case, all three covariance matrix (Empirical (uncentered), Clipped, and OptShrinkage) generated lower portfolio risk compared to benchmark. And OptShrinkage (pyRMT) is lowest across all the predictor cases. This conclusion is consistent with the result in the BBP2016risk paper.

As we observe each row in table 1, within each covariance matrix estimator, the random uniform predictor always gives lowest portfolio risk. It’s probably due to the randomness in generating the uniform estimate of future stock return. The minimum variance predictor corresponds to the highest portfolio risk while the omniscient predictor sits in the middle.

Table 2 (Roll both training and out-of-sample data)

Out-of-sample Portfolio Volatility (%)	Minimum Variance	Omniscient	Random Uniform
Empirical (uncentered)	8.3055379	5.9205378	1.2690937
Clipped	7.1644516	5.7079075	0.9044781
OptShrinkage	7.2162787	5.7323686	0.9361581
Empirical (centered) (benchmark)	8.3055367	5.9226240	1.2703541
EWMA $(\alpha = 0.9)$	546.8765754	236.2035113	177.5162272

In table 2, we rolled both training and out-of-sample data. It comes up with the same conclusion as in table 1 that the cleaned versions of estimator (Clipped and OptShrinkage) is effective in the sense of shrinking the benchmark covariance matrix by removing noise in the data. Across different predictors, cleaned versions also did a good job in shrinking the portfolio risk. However, we did notice some difference in this rolling method. Both Minimum Variance and Random Uniform predictors cases have lower portfolio risk and Omniscient case has higher risk compared to table 1. In general, it does have a consistent result as BBP 2016 risk paper.

Visualization of P&L

We include a brief investigation of P&L under different estimators. In this part, we only consider the three estimators in BBP.

Figure 1: Profit and Loss of minimum variance portfolio

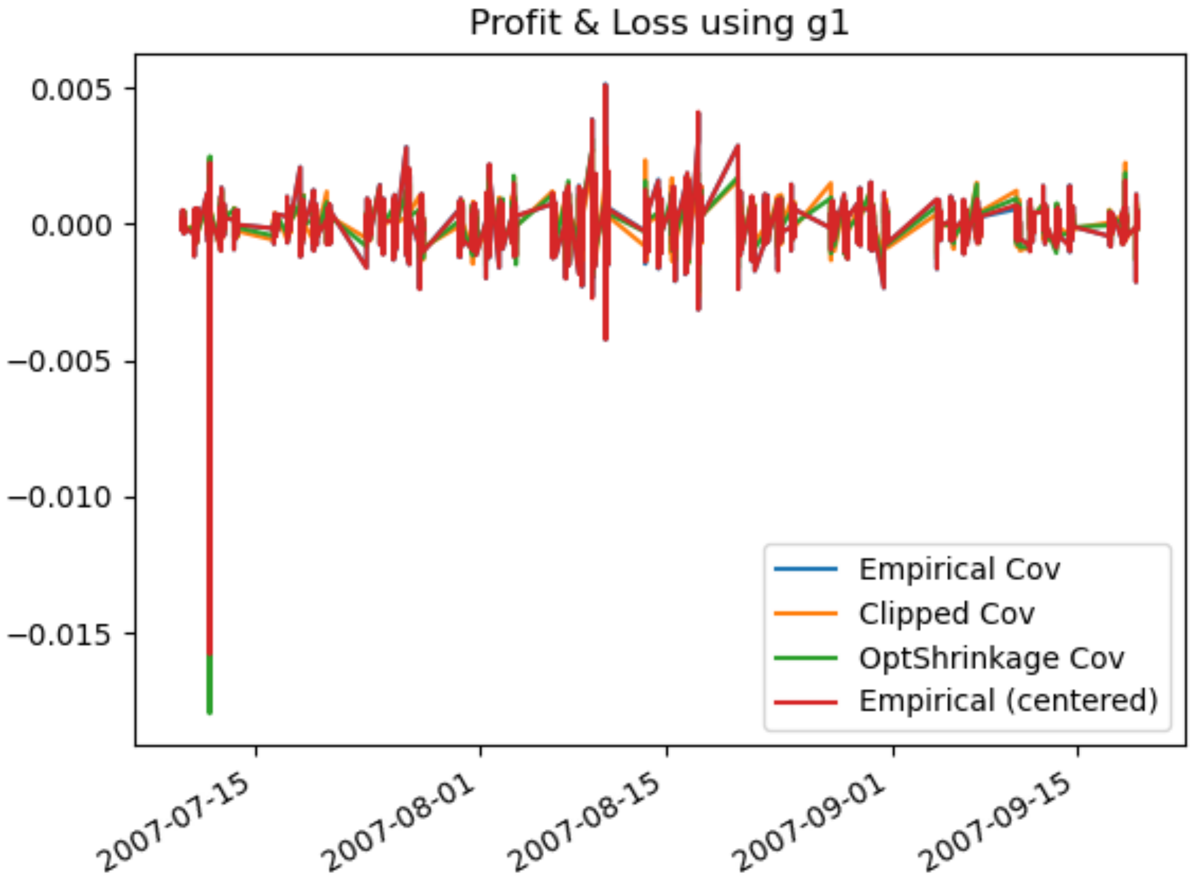
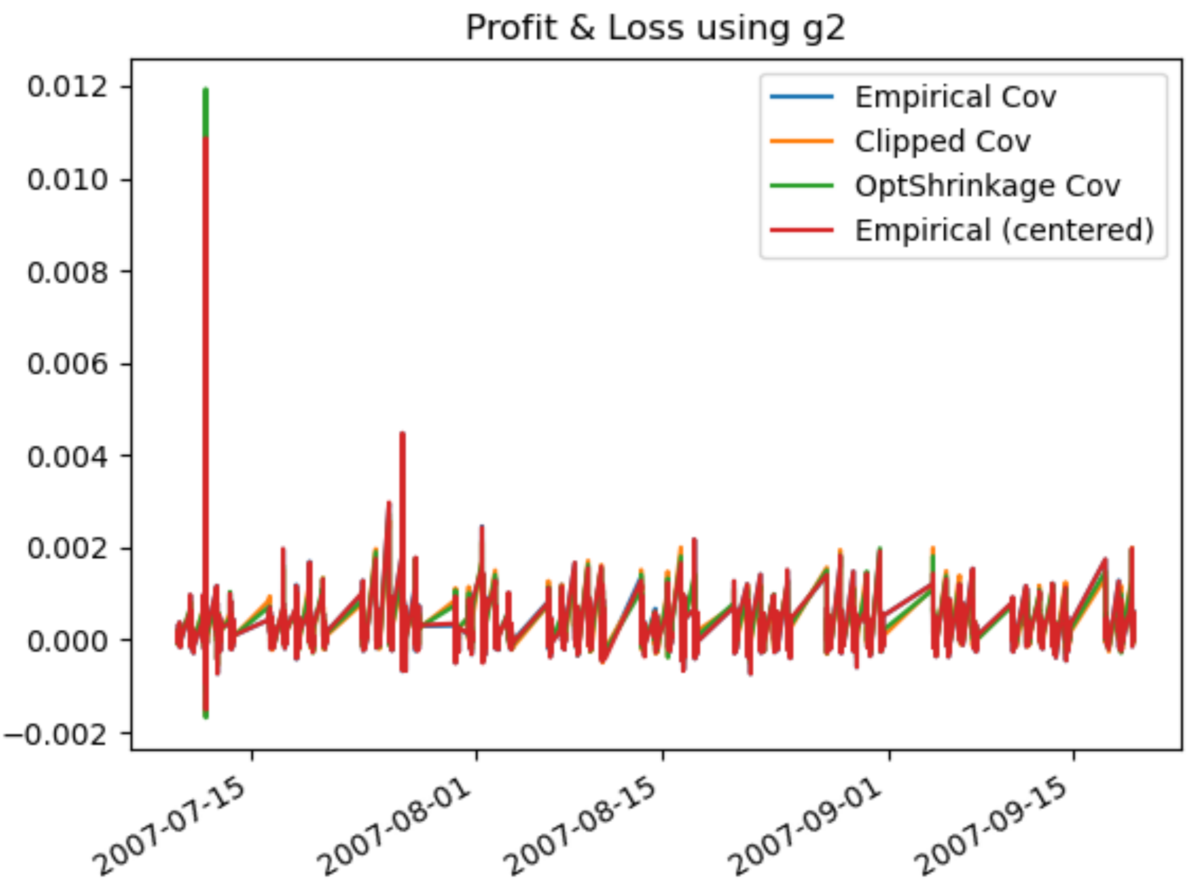


Figure 2: Profit and Loss of omniscient predictor-based portfolio



From these figures, we realize that all the covariance estimators tend to be consistent with regard to capturing the volatility. Yet it's unclear whose performance stands out. This is reasonable because covariance cleaning procedures at best provides marginal improvement (as we have seen in the previous discussion). Mathematically, the covariance serves as the second-order estimation and thus should not have a large impact on the realized profit and loss curve. This also asks for a finer method to evaluate the estimators, for examle the bias statistics.

Bias Statistics

Next we compute the Bias Statistics discussed in the Axioma paper in order to evaluate the bias of different estimators.

Table 3 Bias Statistics (Roll out-of-sample data only)

Bias Statistics	Minimum Variance	Omniscient	Random Uniform
Empirical (uncentered)	0.7090	1.3351	1.3671
Clipped	**0.5982**	**1.5308**	1.1510
OptShrinkage	0.6663	**1.5353**	1.2931
Empirical (centered)	0.7089	1.3342	1.3677
EWMA	0.9257	0.4431	1.0315

95% confidence interval: $[0.6570, 1.3430]$. This interval is computed using: $\lfloor 1-\sqrt{2/T}, 1+\sqrt{2/T} \rfloor$, where $T = (T_{total} - N/q)/T_{out} = (5005 - 476/0.5)/273 = 14$.

Denote $r_t, t = 1, 2, \ldots T$, where t represents 3 days return converted from 5-minute return, $\sigma_t, t = 1, 2, \ldots T$ the standard deviation across 3 days out-of-sample. Then we can compute $Z_t = r_t/\sigma_t$. and thus the Bias statistics $\sqrt{Var(Z_t)}$

As we see in the table 3, the majority of bias statistics falls within the 95% confidence interval indicating that the null hypothesis that the model is unbiased cannot be rejected and thus concludes that our covariance matrix is accurately estimated. However, we did notice that some Bias statistics falls outside the interval. For example Clipped with Minimum Variance predictor has 0.5982 which is towards under-predicting risk over time. Omniscient case tends to be slightly over-predicting risk over time for Clipped and OptShrinkage covariance matrix.

(d)

> Provide a summary of your findings from (c) as a set of recommendations of which covariance matrix estimator to use. Pretend you are working at a startup hedge fund trading U.S. equities and you are are writing this as a memo to the CIO.

We recommend the clipped estimator based on a comparison between four popular covariance cleaning estimates apart from the standard approach. The clipped estimator delivers one of the best portfolio variance reduction, reduces bias significantly, and is yet easily explainable. Additionally, the optimal shrinkage estimator adds to more complexity and reduces the variance further, but introduces more ad hoc bias. This result is consistent based on a rolling-based backtest using high-frequency TAQ data.

We have compared several covariance estimators under Markowitz optimal portfolio framework. We solve following the mean-variance optimization problem by Lagrangian multiplier method.

$$\begin{cases} min_{w \in R^N} \frac{1}{2} w^* C w \\ s.t. w^* g \geq G \end{cases}$$

The solution is

$$w_c = \frac{C^{-1} g}{g^* C^{-1} g}$$

where C is the covariance matrix of assets in our portfolio and g is the expected return of the asset in the portfolio. Forming expectations of future returns is primarily from financial analysts. Thus, correctly estimating the covariance may be arguably the second most important thing.

1. We do recommend that we use cleaned versions of covariance estimators to alleviate the issue above. The clipped covariance matrix tends to do a better job than Empirical estimator as we can see in table 1 and table 2. It generates relatively lower portfolio risk compared to the Empirical estimator. Intuitively, the clipped covariance matrix uses the PCA technique to choose an appropriate K eigenvalue to construct a cleaned covariance matrix estimator.
2. The optimal shrinkage (OptShrinkage) covariance matrix also does a good job compared to Empirical estimator. It produces lower portfolio risk compared to Empirical estimator. OptShrinkage follows the method in BBP2016 which is an ad hoc method proposed to correct for a systematic bias for small eigenvalues.
3. We don't recommend using sample covariance (e.g., Empirical covariance estimator) in the solution

formula (5) at all because estimation error leads to poor performance in Mean-Variance optimization. Also, we often face curse of dimensionality when the number of assets in the portfolio is larger than the period T we want to forecast. Intuitively, we shouldn't use past covariance of asset to forecast that for the future time period.

4. When we use Bias Statistics to evaluate the performance of Clipped and optimal shrinkage covariance estimators, we see that the Bias Statistics tend to fall into the 95% confidence interval indicating that the null hypothesis that the model is unbiased cannot be rejected and thus concludes that our covariance matrix is accurately estimated.

(e)

> Extra credit: Implement the covariance matrix estimator described in "Exponential Weighting and Random-Matrix-Theory-Based Filtering of Financial Covariance Matrices for Portfolio Optimization" by Szilard Pafka, Marc Potters and Imre Kondor. Then evaluate and compare its performance with your other 3 estimators from above. What do you find?

Aside from the three estimators, we introduce the exponenial weighted moving average (EWMA) covariance estimator defined as

$$C_{ij} = \frac{1-\alpha}{1-\alpha^T} \sum_{\tau=1}^{T} \alpha^{T-\tau} r_{i,\tau} r_{j,\tau},\ \alpha < 1$$

The coefficient $(1-\alpha)/(1-\alpha^T)$ simply serves as the normalizer. The implied assumption is (again) $E[r_{i,t}] = 0$ for all stock and time. With $\alpha \to 1$, this normalizer degenerates to $1/T$, and this estimator simply becomes the uncentered empirical estimator. As an intuition, a higher α corresponds to an increased significance of past observations.

We have shown the results using the EWMA estimator in the previous tables. We realize that EWMA works best in bias reduction but increases the variance significantly. We attribute this to the natrue of high frequency trading where variance is usually accumulated. The EWMA puts too much weight on more recent observations and thus prone to this variance accumulation.

Problem 2

We start with some preparations. The **efficient frontier** is the set of optimal portfolios that offer the highest expected return for a defined level of risk, or the lowest risk for a given level of expected return. We show that the efficient frontier of risky assets is a hyperbola, in the (σ, μ) plane space. We then solve for the efficient frontier when a risk-free asset is available and discuss this in three different situations.

Let's derive the efficient frontier of risky assets. This is to

$$\min_w \frac{1}{2} w' \Sigma w$$

s.t.

$$w' \mu = \mu_0$$

$$w' e = 1$$

The Lagrangian is in the form of

$$L(w, \lambda, \gamma) = \frac{1}{2} w' \Sigma w + \lambda(1 - w'e) + \gamma(\mu_0 - w'\mu)$$

We set each of $\frac{\partial L}{\partial w}, \frac{\partial L}{\partial \lambda}, \frac{\partial L}{\partial \gamma}$ to 0 to get

$$\Sigma w^* - \lambda^* e - \gamma^* \mu = 0$$

$$\mu' w^* = 0$$

$$e' w^* = 1$$

Solving these equations, we can get

$$w^* = \lambda^* \Sigma^{-1} e + \gamma^* \Sigma^{-1} \mu$$

$$\lambda^* = \frac{1}{\Delta}(C - \mu_0 B)$$

$$\gamma^* = \frac{1}{\Delta}(\mu_0 A - B)$$

Where

$$A = e' \Sigma^{-1} e$$

$$B = e' \Sigma^{-1} \mu$$

$$C = \mu' \Sigma^{-1} \mu$$

Next, we find out the shape of the efficient frontier in the (σ, μ) square. When μ_0 varies, we need to find out the shape of the smallest σ.

$$(\sigma^*)^2 = (w^*)' \Sigma w^* = (w^*)' \Sigma(\lambda^* \Sigma^{-1} e + \gamma^* \Sigma^{-1} \mu) = \lambda^* (w^*)' e + \gamma^* (w^*)' \mu$$

According that

$$(w^*)' \mu = \mu_0$$

$$(w^*)' e = 1$$

We get

$$(\sigma^*)^2 = \lambda^* + \gamma^* \mu_0 = \frac{1}{\Delta}(C - \mu_0 B) + \frac{1}{\Delta}(\mu_0 A - B)\mu_0 = \frac{A\mu_0^2 - 2B\mu_0 + C}{\Delta}$$

This is a hyperbola symmetric about $\mu = \frac{A}{B}$, and an important fact is that this is just the expected return of the GMV portfolio.

(a)

We now work on the analytical solutions of different situations. Although these three situations have different efficient frontier shapes, the logic to derive the optimal portfolio is the same, we try to minimize $\frac{1}{2} w' \Sigma w$, when a μ_0 is given.

Any covariance matrix is symmetric and positive semi-positive. So, this is a quadratic optimization problem.

$$\min_{w_R} \frac{1}{2} w'_R \Sigma w_R$$

subject to

$$\mu_0 = w'_R \mu + (1 - w'_R \iota) R_f$$

The optimal portfolio w_R can be solved by Lagrange method. We build a Lagrangian

$$L = \frac{1}{2} w'_R \Sigma w_R - \lambda(w'_R \mu + R_f - R_f w'_R \iota - \mu_0)$$

We have partial derivatives with respect to w and λ

$$\frac{\partial L}{\partial w_R} = \Sigma w_R - \lambda \mu + R_f \lambda \iota$$

$$\frac{\partial L}{\partial \lambda} = w'_R \mu + R_f - R_f w'_R \iota - \mu_0$$

Setting these two partial derivatives to be 0, the solution of w_R and λ are the answer we want,

$$w^*_R = \lambda^* \Sigma^{-1} (\mu - R_f \iota)$$

$$(\mu - R_F \iota)' w^*_R = \mu_0 - R_f$$

Combine these two equations,

$$\lambda^* (\mu - R_F \iota)' \Sigma^{-1} (\mu - R_f \iota) = \mu_0 - R_f$$

$$\lambda^* = \frac{\mu_0 - R_f}{(\mu - R_F \iota)' \Sigma^{-1} (\mu - R_f \iota)}$$

and then we can get the result

$$w^*_R = \lambda^* \Sigma^{-1} (\mu - R_f \iota)$$

When the line is tangent, 1-w_R=0, and we can get a tangent portfolio, in this case, $w'_{tang} \iota$=1, hence

$$w'_{tang} = \frac{\Sigma^{-1} (\mu - R_f \iota)}{\iota' \Sigma^{-1} (\mu - R_f \iota)}$$

Using w_{tang}, we can easily get μ_{tang} and σ_{tang}, and then we can define efficient frontier in the three situations.

Situation 1. $E(R_{GMV}) < R_f$

In this case, we have a tangent line, and it goes to the right-up corner. Hence, the efficient frontier is

$$\mu_e = \sigma_e \frac{\mu_{tang} - R_f}{\sigma_{tang}} + R_f \ for \ \mu_e >= R_f$$

Situation 2. $E(R_{GMV}) = R_f = \frac{B}{A}$

In this case, we do not have a tangent portfolio, but we still can get the efficient frontier by using the property of hyperbola. For a hyperbola in the form of $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$, its asymptotic lines are $y = \frac{b}{a}x$ and $y = -\frac{b}{a}x$. Remember that our hyperbola is (by rearranging terms)

$$\frac{\sigma^2}{\frac{1}{A}} - \frac{(\mu - \frac{B}{A})^2}{\frac{\Delta}{A^2}} = 1$$

Hence, the efficient frontier are

$$\mu_e = \sqrt{\frac{\Delta}{A}}\sigma_e + R_f \ for \ \sigma > 0 \ and \ \mu_e >= R_f$$

$$\mu_e = -\sqrt{\frac{\Delta}{A}}\sigma_e + R_f \ for \ \sigma < 0 \ and \ \mu_e >= R_f$$

When we take negative σ values, it means that we short that portfolio.

Situation 3. $E(R_{GMV}) < R_f$

In this case, we have a tangent line, and it goes to the left-up corner. Hence, the efficient frontier is

$$\mu_e = \sigma_e \frac{\mu_{tang} - R_f}{\sigma_{tang}} + R_f \ for \ \mu_e >= R_f$$

Although this looks exactly the same as the situation 1, pay attention that the $\frac{\mu_{tang} - R_f}{\sigma_{tang}}$ part is smaller than zero so this line goes to left and σ take negative values.

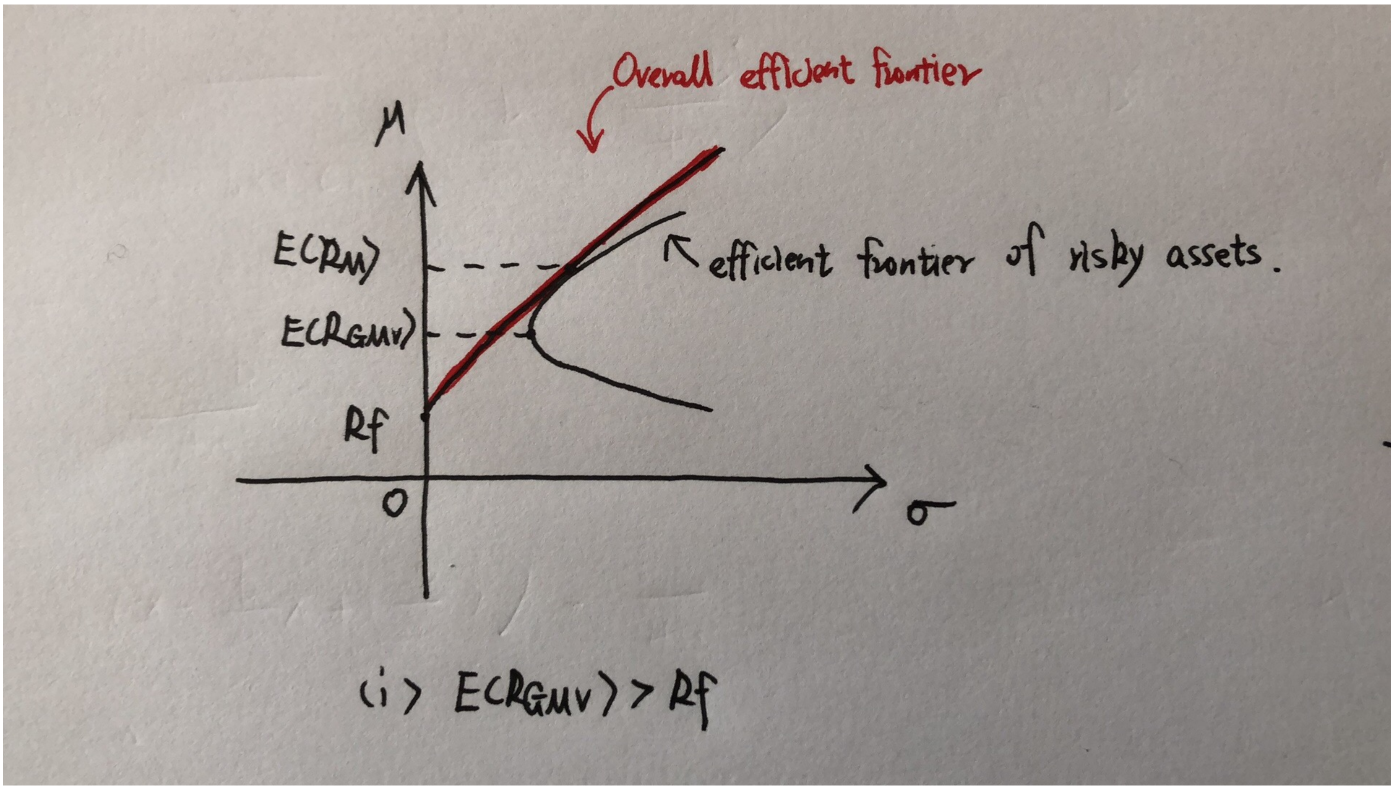
(b)

We then draw the graphs according to each situation and investigate when such portfolios will be optimal:

1. We can long the risky assets only, long the risky-free asset only or long a mixture of risky and risk-free assets.
2. We can short the risky-free asset and invest the proceeds into the risky assets.
3. We can short the risky assets and invest the proceeds into the risk-free asset.

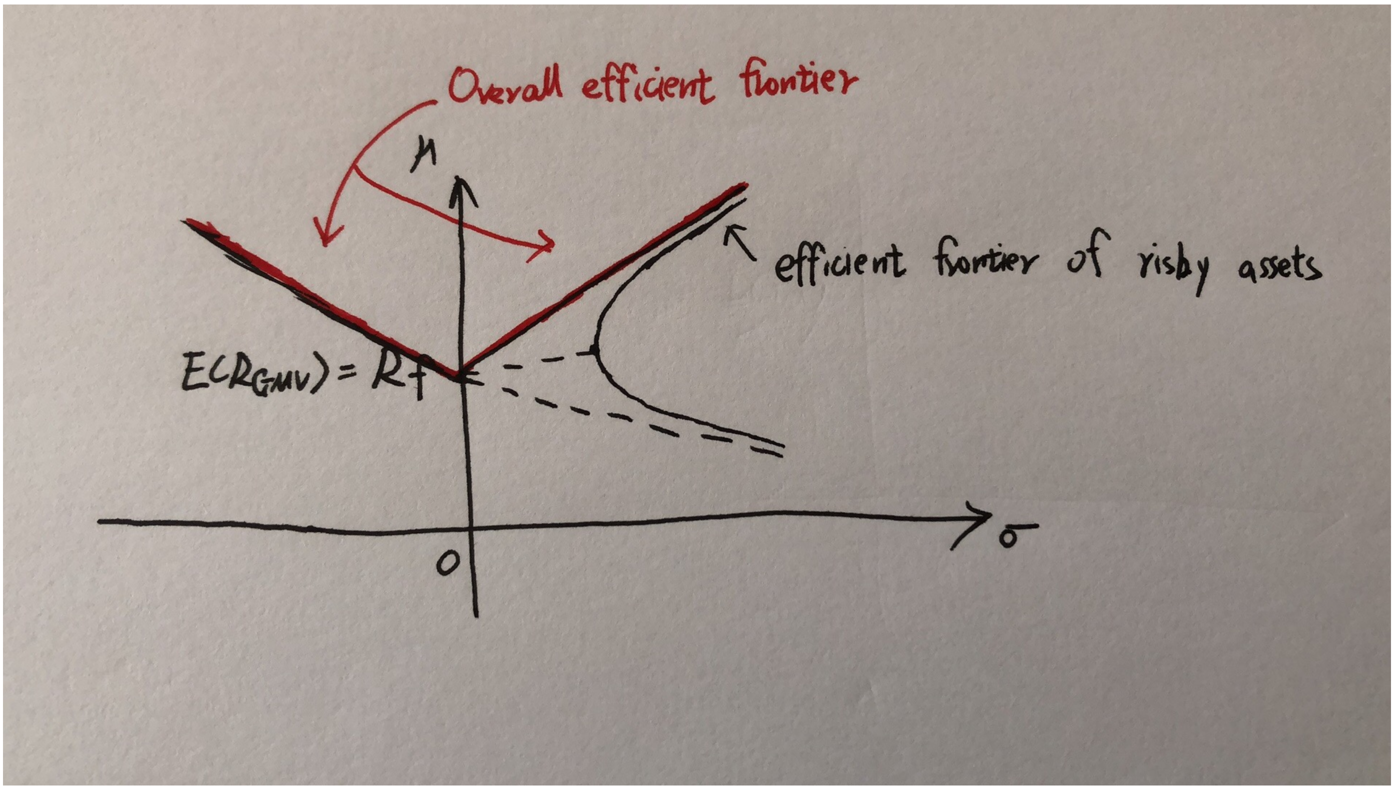
Recall that our purpose is still to minimize overall σ, when μ is given. So, let's consider what kind of action we should take when R_f varies

Situation 1. $E(R_{GMV}) < R_f$



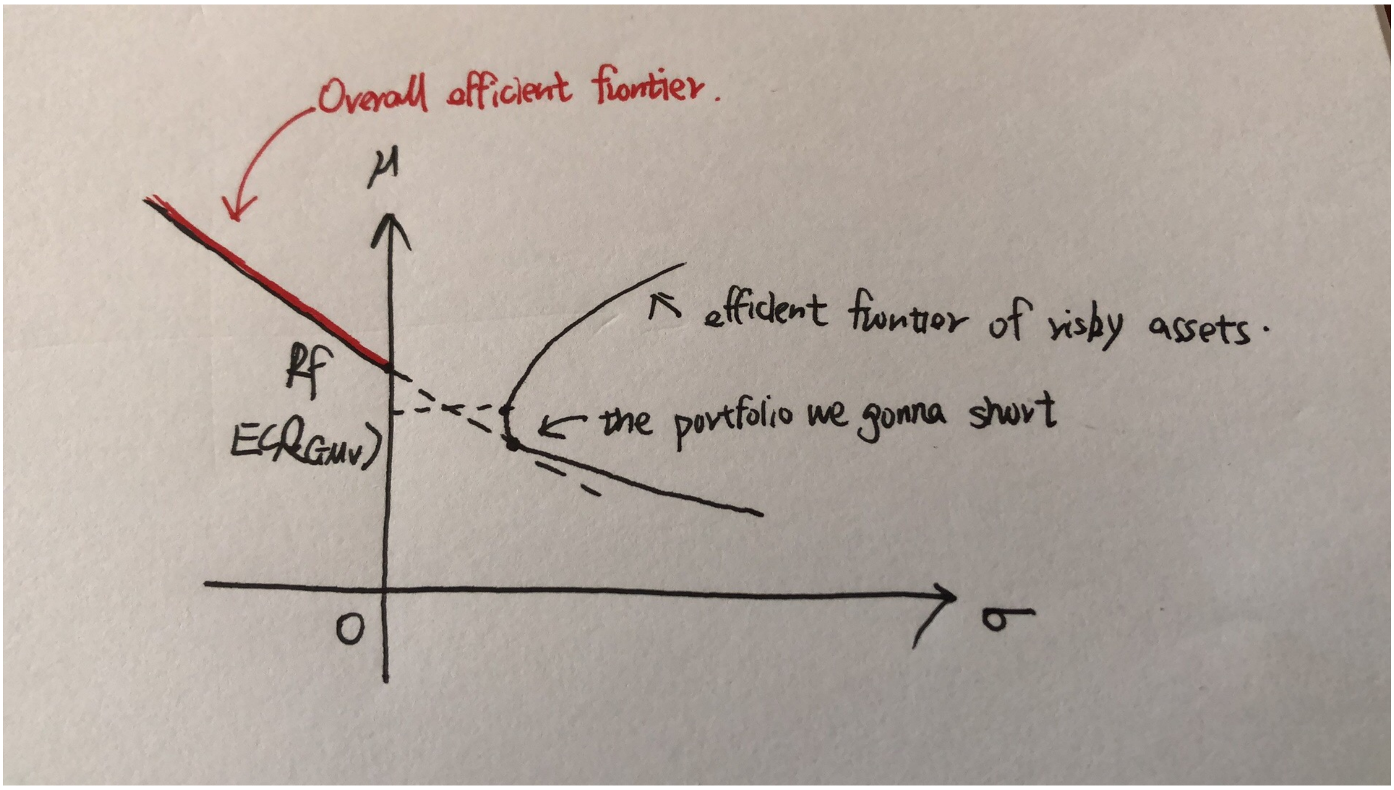
This happens when $E(R_{GMV}) > R_f$. This is what we always see in textbooks, and in this situation, the overall efficient frontier is just the tangent line of the hyperbola starting from the point (0, R_f). In the part left to the tangent point, we invest in both risk-free asset and risky assets. In the part right to the tangent point, we need to short the risk-free asset and invest the proceeds in the risky assets.

Situation 2. $E(R_{GMV}) = R_f$



This happens when $E(R_{GMV}) = R_f$. According to the property of a hyperbola, if you start from R_f, and manage to get a tangent line, you cannot make it, but you can get two asymptotic lines instead. So, these two asymptotic line become overall efficient frontier. We need to talk about the part in the second quadrant. This means that you short the risky assets, and you invest the proceeds into the risk-free asset. The risk asset you short is in the direction of the dotted line in the first quadrant. If we short an asset, we need to find out the asset with the lowest expected return at a given variation, so we exploit the lower asymptotic line. As the efficient frontier is in the first quadrant, it is just like what we always meet, investing in both risk-free asset and risky assets. One small difference here is that you may not fully reach this position, because this is an asymptotic line, but you can get one which is very close. The reason why there are two efficient frontiers is that they have the same expected return when the σ^2 is given.

Situation 3. $E(R_{GMV}) < R_f$



This happens when $E(R_{GMV}) < R_f$. We would like to short a portfolio with the lowest expected return possible and invest the proceeds into the risk-free market. This guarantees that we can have the highest expected return possible whenever we are given a σ^2.

(c)

We expect to observe situation 2 and 3 above in real life, which could occur when stock prices are falling, and the economy is in a recession, then the tangency portfolio will have a negative Sharpe slope. In this case, efficient portfolios involve shorting the tangency portfolio and investing the proceeds in T-Bills.