THE ELUSIVE FRONTIER

SIMON ELLERSGAARD & ROLF POULSEN

Abstract. A popular exercise in modern portfolio theory is the computation of an *efficient frontier* - the locus of portfolios which have the smallest possible variance for a given level of return. While no doubt a monumental result for the asset management industry, one does wonder just how robust these results are under parameter uncertainty (neither mean nor variance are *a priori* known quantities). In this note we demonstrate that the answer is "not very robust at all" principally due to the trouble with estimating the correct mean.

Remark 1. This note now form part of a chapter for an upcoming book *Finance 1 and Beyond* by Lando & Poulsen. Poulsen proposed the project. Ellersgaard did the coding and typed up this note.

Let $\mathbf{R} = (R_1, R_2, ..., R_N)^T$ be the **excess return** per unit time of N risky assets, i.e. let R_i be defined as the **rate of return** per unit time of the i^{th} asset *minus* the rate of return per unit time of the risk free asset for $i = 1, 2, ..., N$. The mean excess returns per unit time is represented by the vector $\boldsymbol{\mu} = E(\mathbf{R})$, and the covariance matrix per unit time is given by $\boldsymbol{\Sigma} = Var(\mathbf{R}) = E((\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^T)$. Assume the excess returns to be normally distributed, i.e. $\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Since the covariance matrix $\boldsymbol{\Sigma}$ by definition is positive definite, it follows that it admits a Cholesky decomposition:

$$\exists \boldsymbol{\sigma} \in \mathbb{R}^{N\times N} \text{ s.t. } \boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T,$$

where $\boldsymbol{\sigma}$ is a lower triangular matrix. In particular, one may readily check that $\mathbf{R}$ thence can be written as

$$\mathbf{R} = \boldsymbol{\mu} + \boldsymbol{\sigma}\mathbf{Z},$$

where $\mathbf{Z}$ is a random vector $\Omega \mapsto \mathbb{R}^n$ with distribution $\sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. More generally, the excess return vector over the time increment Δt is given by

$$\mathbf{R}_{\Delta t} = \boldsymbol{\mu}\Delta t + \boldsymbol{\sigma}\sqrt{\Delta t}\mathbf{Z},$$

in discrete analogy with geometric brownian motion.

Provided that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known to us, we can compute the mean-variance frontier and the capital market line using standard techniques. Nevertheless, it remains unclear whether these quantities can be reliably estimated, and indeed what bearing a negative answer to this query might have on our capital allocation. To test just how stable the mean-variance frontier is for empirical estimates of the mean and covariance, we design the following experiment:

(1) First, to get empirically plausible values for the estimators $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$ we use daily empirical data for five risky assets based on Kenneth French's "five industry portfolios" and the "Farma-French 3-Factors" available for free at `http://mba.tuck.dartmouth.edu/`

`pages/faculty/ken.french/data_library.html`. Unbiased sample estimates for the components μ_i and Σ_{ij} are given by

$$\bar{\mu}_i := \frac{1}{n\Delta t}\sum_{k=1}^{n}(R_{\Delta t})_{ik}, \tag{1}$$

and

$$\bar{\Sigma}_{ij} := \frac{1}{(n-1)\Delta t}\sum_{k=1}^{n}[(R_{\Delta t})_{ik} - \bar{\mu}_i\Delta t][(R_{\Delta t})_{jk} - \bar{\mu}_j\Delta t], \tag{2}$$

where $\{(R_{\Delta t})_{ij}\}_{j=1}^{n}$ is a time series of n consecutive Δt excess returns for asset i.

(2) For our present purposes we shall think of $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$ as the "true" parameters of the market.

(3) Using the equation $\mathbf{R}_{\Delta t} = \bar{\boldsymbol{\mu}}\Delta t + \bar{\boldsymbol{\sigma}}\sqrt{\Delta t}\mathbf{Z}$ where $\boldsymbol{\Sigma} = \bar{\boldsymbol{\sigma}}\bar{\boldsymbol{\sigma}}^T$ we now simulate m future evolutions of the five risky assets over a fixed temporal horizon.

(4) For each future evolution in the simulated data, we re-compute sample estimates for the mean and the covariance matrix. Label these by $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\Sigma}}_i$ where $i = 1, 2, ..., m$. Obviously the *expected values* of the random variables $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ will be the "true" parameters $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$.

(5) Using the mean-variance equation

$$\hat{\sigma}_P^2(\mu_P) = \frac{a - 2b\mu_P + c\mu_P^2}{d}$$

or, equivalently,

$$\mu_p(\hat{\sigma}_P^2) = \frac{b}{c} \pm \sqrt{\frac{1}{c}\left[d\hat{\sigma}_P^2 + \frac{b^2}{c} - a\right]},$$

where $a = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $b = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1} = \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $c = \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$, and $d = ac - b^2$ we plot the mean-variance frontier $(\hat{\sigma}_P^2, \mu_P)$ for the scenarios:

(a) $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ (the simulated sample estimates).
(b) $(\hat{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$ (simulated sample mean, "true" covariance).
(c) $(\bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ ("true" mean, simulated covariance).

(6) We conjecture that the frontiers will be all over the place when we use the simulated sample mean, $\hat{\boldsymbol{\mu}}$. The simple explanation for this is that the real parameter $\bar{\boldsymbol{\mu}}$ is difficult to estimate reliably (as we have seen elsewhere, for log-returns the estimator is a telescoping sum, meaning that only the first and last data point in the stock price process end up determining the expected return). A similar problem does not pertain to the covariance.

The true mean ($\bar{\boldsymbol{\mu}}$) and covariance matrix ($\bar{\boldsymbol{\Sigma}}$) for French's five industry portfolios are exhibited in tables 1 and 2. The estimators are based on daily data points collected over the 20 year horizon July 1995 to July 2015. The associated mean-variance frontier, market portfolio, and capital market line (CML) are exhibited in the lefthand part of figure 1. Notice that both the frontier and the CML are parabolic functions in (variance,mean)-space - had we plotted the corresponding curves in (standard deviation,mean)-space they would respectively transform to a hyperbola and a straight line. Furthermore, notice the trending inverse relationship between the expected return of the portfolios and their variances (marked by x in the figure): one would

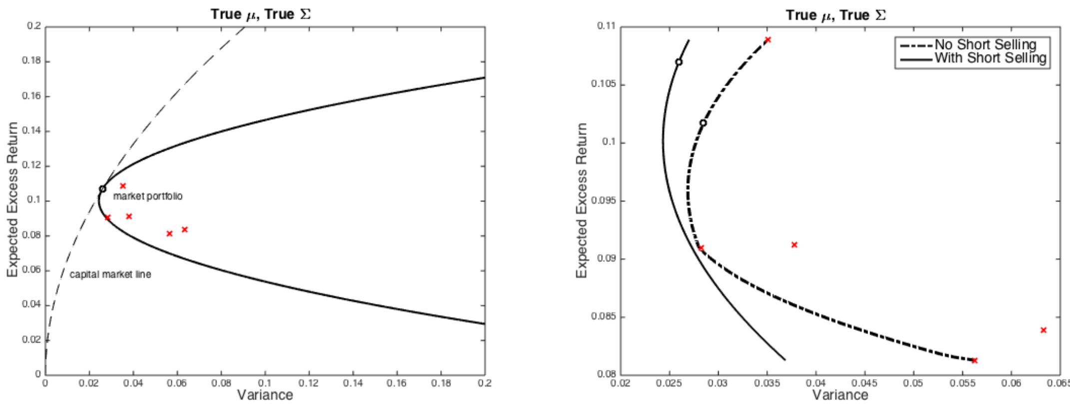

FIGURE 1. **Left:** The mean-variance frontier computed from five industrial indices, marked by x in the diagram. The white dot represents the associated market portfolio, while the dotted line connecting the origin and the MP represents the capital market line. **Right:** The mean-variance frontier with and without short-selling restrictions.

probably guess that taking on more risk (volatility) would be compensated by the promises of a higher expected return, but clearly this is not the case in this concrete empirical example!

Out of interest, the righthand side of figure 1 also exhibits the mean-variance frontier in the event we enforce the "no short-selling" restriction $\boldsymbol{\pi} \geq 0$.[1] Given the linearity of portfolio returns, such a frontier can only be drawn between the data point with the lowest expected excess return and the data point with the highest expected excess return. Unsurprisingly, a ban on short-selling entails a dampening on the market portfolio from $(\hat{\sigma}_P^2, \mu_P) = (0.0259, 0.1070)$ to $(\hat{\sigma}_P^2, \mu_P) = (0.0285, 0.1017)$, and therefore more conservative Sharpe ratios for rational investors (from 0.6640 to 0.6026).

Asset 1	Asset 2	Asset 3	Asset 4	Asset 5
0.0909	0.0912	0.0839	0.1089	0.0813

TABLE 1. "True" mean excess return for the five industrial indices (French 1995-2015).

Asset 1	Asset 2	Asset 3	Asset 4	Asset 5
0.0282	0.0266	0.0311	0.0236	0.0336
0.0266	0.0378	0.0343	0.0252	0.0373
0.0311	0.0343	0.0633	0.0290	0.0448
0.0236	0.0252	0.0290	0.0351	0.0305
0.0336	0.0373	0.0448	0.0305	0.0563

TABLE 2. "True" covariance matrix for the five industrial indices (French 1995-2015).

[1] This problem must be solved numerically; e.g. using MATLAB's quadprog function. For consistency the depicted true mean-variance frontier has also been computed numerically.

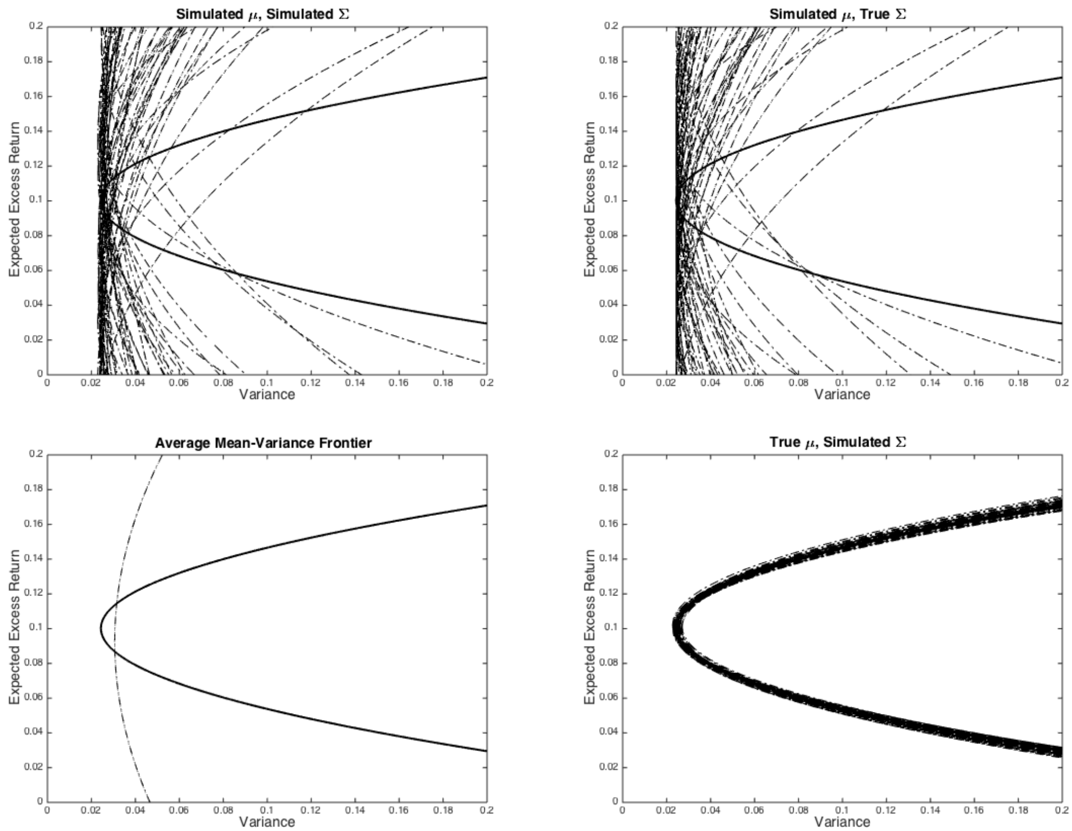

FIGURE 2. **Top left:** The mean-variance frontiers for $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ (the simulated sample estimates). **Top right:** The mean variance-frontiers for $(\hat{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$ (simulated sample mean, "true" covariance). **Bottom left:** The average mean-variance frontier for the simulated sample estimates. Specifically, the dotted line represents the horizontal average of the dotted lines in the top left figure. **Bottom right:** The mean variance-frontiers for $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$ ("true" mean, simulated covariance).

Next we simulate 50 evolutions of the five industrial indices five years into the future. The mean-variance frontiers generated by the associated estimators $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\Sigma}}_i$ for $i = 1, 2, ..., 50$ are exhibited in figure 2. Immediately we notice that frontiers which utilise the sample estimator $\hat{\boldsymbol{\mu}}$ are *highly* scattered with respect to the "true" mean-variance frontier, irrespective of whether we use the "true" or the sample covariance. On the other hand, if we use the "true" drift $\bar{\boldsymbol{\mu}}$ the associated frontiers collapse to something resembling the "true" mean-variance frontier. The implication is clear: mis-specifications of the expected mean return of the risky assets will invariably have a devastating impact on the way rational investors think they should invest vis-a-vis how they ought to invest given full information about the governing dynamics. This form of model mis-specification can significantly curb their welfare gains. The problem, of course, is that the "true" drift $\bar{\boldsymbol{\mu}}$ is notoriously unreliable to estimate (recall that for log returns only the first and final data points in the time series go into the estimation). This point is highlighted in the lefthand part of figure 3 where we plot the sample parameter $\hat{\boldsymbol{\mu}}$ across the different simulations.

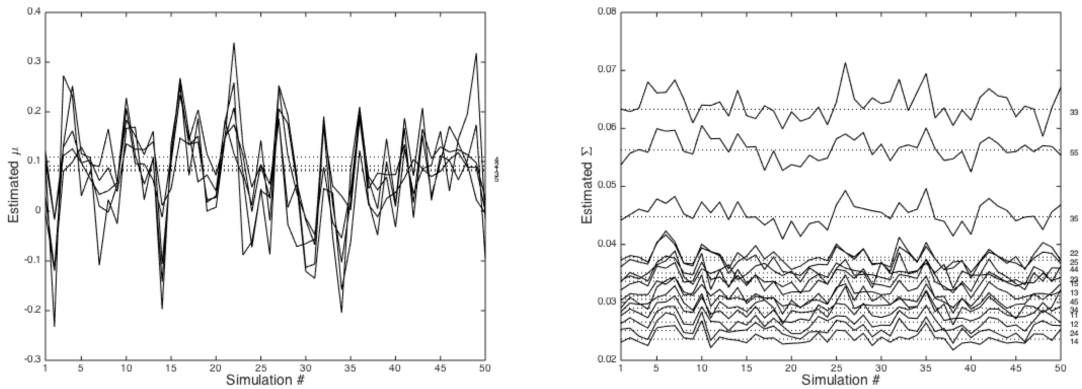

FIGURE 3. **Left:** The stability of the estimator $\hat{\boldsymbol{\mu}}$ across different simulations. The dotted lines represent the "true" means. Clearly, the estimators are highly erratic. **Right:** The stability of the estimator $\hat{\boldsymbol{\Sigma}}$ across different simulations. The dotted lines represent the "true" covariances. Clearly, the estimators are stable.

Clearly, the simulated drift estimators oscillate wildly around their "true" counterparts. Given the quadratic nature of the covariance estimator, an analogous problem does not prevail here: there is no significant information/welfare loss affecting rational investors in deploying $\hat{\boldsymbol{\Sigma}}$ over $\bar{\boldsymbol{\Sigma}}$.