Portfolios that Contain Risky Assets

Stochastic Models 4.

Kelly Objectives for Markowitz Portfolios

C. David Levermore
University of Maryland, College Park

Math 420: *Mathematical Modeling*
April 7, 2016 version

Portfolios that Contain Risky Assets
Part II: Stochastic Models

1. Independent, Identically-Distributed Models
2. Growth Rate Mean and Variance Estimators
3. Law of Large Numbers (Kelly) Objectives
4. Kelly Objectives for Markowitz Portfolios
5. Central Limit Theorem Objectives
6. Optimization of Mean-Variance Objectives
7. Utility Function Objectives

Stochastic Models 4. Kelly Objectives for Markowitz Portfolios

1. Introduction
2. Growth Rate Mean Estimators
3. Without Risk-Free Assets
4. One Risky-Free Rate Model

Stochastic Models 4. Kelly Objectives for Markowitz Portfolios

Introduction. We now want to apply the Kelly criterion to classes of Markowitz portfolios. Given a daily return history $\{\mathbf{r}(d)\}_{d=1}^{D}$ on N risky assets, a daily return μ_{si} on a safe investment, and a daily return μ_{cl} on a credit line, the Markowitz portfolio with allocation $\mathbf{f}$ in risky assets has the daily return history $\{r(d,\mathbf{f})\}_{d=1}^{D}$ where

$$r(d,\mathbf{f}) = \mu_{\mathsf{rf}}\Big(1 - \mathbf{1}^{\mathsf{T}}\mathbf{f}\Big) + \mathbf{r}(d)^{\mathsf{T}}\mathbf{f}\,, \tag{1}$$

with

$$\mu_{\mathsf{rf}} = \begin{cases} \mu_{\mathsf{si}} & \text{if } \mathbf{1}^{\mathsf{T}}\mathbf{f} \leq 1\,, \\ \mu_{\mathsf{cl}} & \text{if } \mathbf{1}^{\mathsf{T}}\mathbf{f} > 1\,. \end{cases} \tag{2}$$

The *one risk-free rate model* for risk-free assets assumes $0 < \mu_{\mathsf{si}} = \mu_{\mathsf{cl}}$. The *two risk-free rate model* for risk-free assets assumes $0 < \mu_{\mathsf{si}} < \mu_{\mathsf{cl}}$. Portfolios *without risk-free assets* are modeled by adding the constraint $\mathbf{1}^{\mathsf{T}}\mathbf{f} = 1$.

We will consider only classes of *solvent Markowitz portfolios*. This means that we require $\mathbf{f} \in \Omega^+$ where

$$\Omega^+ = \left\{ \mathbf{f} \in \mathbb{R}^N \,:\, 1 + r(d, \mathbf{f}) > 0 \;\; \forall d \right\}. \tag{3}$$

It can be shown that *$r(d, \mathbf{f})$ is a concave function of $\mathbf{f}$ over $\mathbb{R}^N$ for every d*. This means that for every d and every $\mathbf{f}_0$, $\mathbf{f}_1 \in \mathbb{R}^N$ we can show that

$$r(d, \mathbf{f}_s) \geq (1 - s)\, r(d, \mathbf{f}_0) + s\, r(d, \mathbf{f}_1) \quad \text{for every } s \in [0, 1]\,,$$

where $\mathbf{f}_s = (1 - s)\,\mathbf{f}_0 + s\,\mathbf{f}_1$. This concavity implies that for every $\mathbf{f}_0$, $\mathbf{f}_1 \in \Omega^+$ and every $s \in [0, 1]$ we have

$$\begin{aligned} 1 + r(d, \mathbf{f}_s) &\geq 1 + (1 - s)\, r(d, \mathbf{f}_0) + s\, r(d, \mathbf{f}_1) \\ &= (1 - s)\Big(1 + r(d, \mathbf{f}_0)\Big) + s\Big(1 + r(d, \mathbf{f}_1)\Big) \geq 0\,, \end{aligned}$$

whereby $\mathbf{f}_s \in \Omega^+$. Therefore *$\Omega^+$ is a convex set*.

The solvent Markowitz portfolio with allocation $\mathbf{f}$ has the growth rate history $\{x(d,\mathbf{f})\}_{d=1}^{D}$ where

$$x(d,\mathbf{f}) = \log\Big(1 + r(d,\mathbf{f})\Big)\,. \tag{4}$$

Notice that the growth rate history is only defined for solvent portfolios.

Because $r(d,\mathbf{f})$ is a concave function over $\mathbf{f} \in \mathbb{R}^N$ for every d and because $\log(1+r)$ is an increasing, strictly concave function of r over $r \in (-1,\infty)$, we can show that *$x(d,\mathbf{f})$ is a concave function of $\mathbf{f}$ over Ω^+ for every d*. Indeed, for every $\mathbf{f}_0$, $\mathbf{f}_1 \in \Omega^+$ and every $s \in [0,1]$ we have

$$\begin{aligned}
x(d,\mathbf{f}_s) &= \log\Big(1 + r(d,\mathbf{f}_s)\Big) \\
&\geq \log\Big(1 + (1-s)\,r(d,\mathbf{f}_0) + s\,r(d,\mathbf{f}_1)\Big) \\
&\geq (1-s)\,\log\Big(1 + r(d,\mathbf{f}_0)\Big) + s\,\log\Big(1 + r(d,\mathbf{f}_1)\Big) \\
&= (1-s)\,x(d,\mathbf{f}_0) + s\,x(d,\mathbf{f}_1)\,.
\end{aligned}$$

Growth Rate Mean Estimators. If we use an IID model for the class of solvent Markowitz portfolios then the Kelly criterion says that for maximal long-term growth we should pick $\mathbf{f} \in \Omega^+$ to maximize the growth rate mean $\gamma(\mathbf{f})$ of the underlying probability distribution for growth rates. Because we do not know $\gamma(\mathbf{f})$, the best we can do is to maximize an estimator $\widehat{\gamma}(\mathbf{f})$ for $\gamma(\mathbf{f})$. Here we will explore different expressions for $\widehat{\gamma}(\mathbf{f})$.

Given and allocation $\mathbf{f}$ and weights $\{w(d)\}_{d=1}^{D}$ such that

$$w(d) > 0 \quad \forall d\,, \qquad \sum_{d=1}^{D} w(d) = 1\,, \tag{5}$$

the growth rate history $\{x(d, \mathbf{f})\}_{d=1}^{D}$ yields the estimator

$$\widehat{\gamma}(\mathbf{f}) = \sum_{d=1}^{D} w(d)\, x(d, \mathbf{f}) = \sum_{d=1}^{D} w(d)\, \log\Big(1 + r(d, \mathbf{f})\Big)\,. \tag{6}$$

This is clearly defined for every $\mathbf{f} \in \Omega^+$.

Here we collect some facts about $\widehat{\gamma}(\mathbf{f})$ considered as a function over Ω^{+}.

Fact 1. $\widehat{\gamma}(\mathbf{0}) = \log(1 + \mu_{\text{si}})$.

Fact 2. $\widehat{\gamma}(\mathbf{f})$ *is concave over* Ω^{+}.

Fact 3. *For every* $\mathbf{f} \in \Omega^{+}$ *we have the upper bound*

$$\widehat{\gamma}(\mathbf{f}) \leq \log\Big(1 + \widehat{\mu}(\mathbf{f})\Big), \tag{7}$$

where $\widehat{\mu}(\mathbf{f})$ is the estimator of the return mean given by

$$\begin{aligned} \widehat{\mu}(\mathbf{f}) = \sum_{d=1}^{D} w(d)\, r(d,\mathbf{f}) &= \mu_{\text{rf}}\Big(1 - \mathbf{1}^{\mathsf{T}}\mathbf{f}\Big) + \sum_{d=1}^{D} w(d)\, \mathbf{r}(d)^{\mathsf{T}}\mathbf{f} \\ &= \mu_{\text{rf}}\Big(1 - \mathbf{1}^{\mathsf{T}}\mathbf{f}\Big) + \mathbf{m}^{\mathsf{T}}\mathbf{f}\,. \end{aligned} \tag{8}$$

Remark. Fact 1 shows that bound (7) of fact 3 is an equality when $\mathbf{f} = \mathbf{0}$.

Proof. Definitions (1) and (2) of $r(d, \mathbf{f})$ and μ_{rf} respectively show that

$$r(d, \mathbf{0}) = \mu_{\mathsf{rf}}\Big(1 - \mathbf{1}^{\mathsf{T}}\mathbf{0}\Big) + \mathbf{r}(d)^{\mathsf{T}}\mathbf{0} = \mu_{\mathsf{rf}} = \mu_{\mathsf{si}}\,.$$

Then definition (6) of $\widehat{\gamma}(\mathbf{f})$ yields

$$\begin{aligned}\widehat{\gamma}(\mathbf{0}) &= \sum_{d=1}^{D} w(d)\,\log\Big(1 + r(d, \mathbf{0})\Big)\\ &= \sum_{d=1}^{D} w(d)\,\log(1 + \mu_{\mathsf{si}}) = \log(1 + \mu_{\mathsf{si}})\,.\end{aligned}$$

Therefore we have proved **Fact 1**. $\square$

Proof. Because $x(d, \mathbf{f})$ is a concave function of $\mathbf{f}$ over Ω^{+} for every d, and because definition (6) shows that $\widehat{\gamma}(\mathbf{f})$ is a linear combination of these concave functions with positive coefficients, it follows that $\widehat{\gamma}(\mathbf{f})$ is concave over Ω^{+}. This proves **Fact 2**. $\square$

The proof of **Fact 3** uses the *Jensen inequality*. This inequality states that if the function $g(z)$ is convex (concave) over an interval $[a, b]$, the points $\{z(d)\}_{d=1}^{D}$ all lie within $[a, b]$, and the nonnegative weights $\{w(d)\}_{d=1}^{D}$ sum to one, then

$$g(\bar{z}) \leq \overline{g(z)} \qquad \left(\overline{g(z)} \leq g(\bar{z})\right), \tag{9}$$

where

$$\bar{z} = \sum_{d=1}^{D} z(d)\, w(d)\,, \qquad \overline{g(z)} = \sum_{d=1}^{D} g(z(d))\, w(d)\,.$$

For example, if we take $g(z) = z^p$ for some $p > 1$, so that $g(z)$ is convex over $[0, \infty)$, and we take $z(d) = w(d)$ for every d then because the points $\{w(d)\}_{d=1}^{D}$ all lie within $[0, 1]$, the Jensen inequality yields

$$\bar{w}^p = \left(\sum_{d=1}^{D} w(d)^2\right)^p \leq \sum_{d=1}^{D} w(d)^{p+1} = \overline{w^p}\,.$$

The Jensen inequality can be proved for the case when $g(z)$ is convex and differentiable over $[a,b]$ by starting from the inequality

$$g(z) \geq g(\bar{z}) + g'(\bar{z})(z - \bar{z}) \qquad \text{for every } z \in [a,b]\,.$$

This inequality simply says that the tangent line to the graph of g at $\bar{z}$ lies below the graph of g over $[a,b]$. By setting $z = z(d)$ in the above inequality, multiplying both sides by $w(d)$, and summing over d we obtain

$$\begin{aligned}
\sum_{d=1}^{D} g(z(d))\, w(d) &\geq \sum_{d=1}^{D} \Big(g(\bar{z}) + g'(\bar{z})(z(d) - \bar{z})\Big)\, w(d) \\
&= g(\bar{z}) \sum_{d=1}^{D} w(d) + g'(\bar{z}) \left(\sum_{d=1}^{D} \Big(z(d) - \bar{z}\Big)\, w(d) \right) .
\end{aligned}$$

The Jensen inequality then follows from the definitions of $\bar{z}$ and $\overline{g(z)}$.

Remark. There is an integral version of the Jensen inequality that we do not give here because we do not need it.

Proof. Let $\mathbf{f} \in \Omega^+$. Then the points $\{r(d,\mathbf{f})\}_{d=1}^D$ all lie within an interval $[a,b] \subset (-1,\infty)$. Because $\log(1+r)$ is a concave function of r over $(-1,\infty)$, the Jensen inequality (9) and definition (8) of $\widehat{\mu}(\mathbf{f})$ yield

$$\widehat{\gamma}(\mathbf{f}) = \sum_{d=1}^{D} w(d) \log\Big(1 + r(d,\mathbf{f})\Big) \\ \leq \log\left(1 + \sum_{d=1}^{D} w(d)\, r(d,\mathbf{f})\right) = \log\Big(1 + \widehat{\mu}(\mathbf{f})\Big)\,.$$

This establishes the upper bound (7), whereby **Fact 3** is proved. $\square$

Remark. Under very mild assumptions on the return history $\{\mathbf{r}(d)\}_{d=1}^D$ that are always satisfied in practice we can strenghten Fact 2 to *$\widehat{\gamma}(\mathbf{f})$ is strictly concave over Ω^+* and strengthen bound (7) of Fact 3 to the strict inequality

$$\widehat{\gamma}(\mathbf{f}) < \log\Big(1 + \widehat{\mu}(\mathbf{f})\Big) \quad \text{when } \mathbf{f} \neq \mathbf{0}\,. \tag{10}$$

Without Risk-Free Assets. We now consider solvent Markowitz portfolios without risk-free assets. The associated allocations $\mathbf{f}$ belong to

$$\Omega = \left\{ \mathbf{f} \in \Omega^{+} \,:\, \mathbf{1}^{\mathsf{T}}\mathbf{f} = 1 \right\}. \tag{11}$$

On this set the growth rate mean estimator (6) reduces to

$$\widehat{\gamma}(\mathbf{f}) = \sum_{d=1}^{D} w(d)\, \log\!\left(1 + \mathbf{r}(d)^{\mathsf{T}}\mathbf{f}\right). \tag{12}$$

This is an infinitely differentiable function over Ω^{+} with

$$\begin{aligned} \nabla_{\mathbf{f}}\, \widehat{\gamma}(\mathbf{f}) &= \sum_{d=1}^{D} w(d)\, \frac{\mathbf{r}(d)}{1 + \mathbf{r}(d)^{\mathsf{T}}\mathbf{f}}\,, \\ \nabla_{\mathbf{f}}^{2} \widehat{\gamma}(\mathbf{f}) &= -\sum_{d=1}^{D} w(d)\, \frac{\mathbf{r}(d)\, \mathbf{r}(d)^{\mathsf{T}}}{(1 + \mathbf{r}(d)^{\mathsf{T}}\mathbf{f})^{2}}\,. \end{aligned} \tag{13}$$

The Hessian matrix $\nabla_{\mathbf{f}}^2 \,\widehat{\gamma}(\mathbf{f})$ has the following properties.

Fact 4. $\nabla_{\mathbf{f}}^2 \,\widehat{\gamma}(\mathbf{f})$ is nonpositive definite for every $\mathbf{f} \in \Omega$.

Fact 5. $\nabla_{\mathbf{f}}^2 \,\widehat{\gamma}(\mathbf{f})$ is negative definite for every $\mathbf{f} \in \Omega$ if and only if the vectors $\{\mathbf{r}(d)\}_{d=1}^{D}$ span $\mathbb{R}^N$.

Remark. Fact 4 implies that $\widehat{\gamma}(\mathbf{f})$ is concave over Ω, which was already proven in Fact 2. Fact 5 implies that $\widehat{\gamma}(\mathbf{f})$ is strictly concave over Ω when the vectors $\{\mathbf{r}(d)\}_{d=1}^{D}$ span $\mathbb{R}^N$, which is always the case in practice.

Proof. Let $\mathbf{f} \in \Omega$. Then for every $\mathbf{y} \in \mathbb{R}^N$ we have

$$\mathbf{y}^{\mathsf{T}} \nabla_{\mathbf{f}}^2 \,\widehat{\gamma}(\mathbf{f}) \mathbf{y} = - \sum_{d=1}^{D} w(d) \, \frac{(\mathbf{r}(d)^{\mathsf{T}} \mathbf{y})^2}{(1 + \mathbf{r}(d)^{\mathsf{T}} \mathbf{f})^2} \leq 0 \,.$$

Therefore $\nabla_{\mathbf{f}}^2 \,\widehat{\gamma}(\mathbf{f})$ is nonpositive definite, which proves Fact 4. $\square$

Proof. Let $\mathbf{f} \in \Omega$. Then by the calculation in the previous proof we see that for every $\mathbf{y} \in \mathbb{R}^N$

$$\mathbf{y}^\top \nabla^2_{\mathbf{f}}\, \widehat{\gamma}(\mathbf{f})\mathbf{y} = 0 \quad \Longleftrightarrow \quad \mathbf{r}(d)^\top \mathbf{y} = 0 \ \forall d\,.$$

First, suppose that $\nabla^2_{\mathbf{f}}\, \widehat{\gamma}(\mathbf{f})$ is not negative definite. Then there exists an $\mathbf{y} \in \mathbb{R}^N$ such that $\mathbf{y}^\top \nabla^2_{\mathbf{f}}\, \widehat{\gamma}(\mathbf{f})\mathbf{y} = 0$ and $\mathbf{y} \neq \mathbf{0}$. The vectors $\{\mathbf{r}(d)\}_{d=1}^D$ must then lie in the hyperplane orthogonal (normal) to $\mathbf{y}$. Therefore the vectors $\{\mathbf{r}(d)\}_{d=1}^D$ do not span $\mathbb{R}^N$.

Conversely, suppose that the vectors $\{\mathbf{r}(d)\}_{d=1}^D$ do not span $\mathbb{R}^N$. Then there must be a nonzero vector $\mathbf{y}$ that is orthogonal to their span. This means that $\mathbf{y}$ satisfies $\mathbf{r}(d)^\top \mathbf{y} = 0$ for every d, whereby $\mathbf{y}^\top \nabla^2_{\mathbf{f}}\, \widehat{\gamma}(\mathbf{f})\mathbf{y} = 0$. Therefore $\nabla^2_{\mathbf{f}}\, \widehat{\gamma}(\mathbf{f})$ is not negative definite.

Both directions of the characterization in Fact 5 are now proven. $\square$

Henceforth we will assume that the covariance matrix $\mathbf{V}$ is positive definite. Recall that this is equivalent to assuming that the set $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ spans $\mathbb{R}^N$. Because this condition implies that the set $\{\mathbf{r}(d)\}_{d=1}^{D}$ spans $\mathbb{R}^N$, by Fact 5 it implies that $\nabla_{\mathbf{f}}^2 \widehat{\gamma}(\mathbf{f})$ is negative definite for every $\mathbf{f} \in \Omega$. Therefore the estimator $\widehat{\gamma}(\mathbf{f})$ is a strictly concave function over Ω.

Remark. Because $\widehat{\gamma}(\mathbf{f})$ is a strictly concave function over Ω, if it has a maximum then it has a unique maximizer. Indeed, suppose that $\widehat{\gamma}(\mathbf{f})$ has maximum $\widehat{\gamma}_{\mathsf{mx}}$ over Ω, and that $\mathbf{f}_0$ and $\mathbf{f}_1 \in \Omega$ are maximizers of $\widehat{\gamma}(\mathbf{f})$ with $\mathbf{f}_0 \neq \mathbf{f}_1$. For every $s \in (0,1)$ define $\mathbf{f}_s = (1-s)\,\mathbf{f}_0 + s\,\mathbf{f}_1$. Then for every $s \in (0,1)$ we have $\mathbf{f}_s \in \Omega$ and, by the strict concavity of $\widehat{\gamma}(\mathbf{f})$ over Ω,

$$\begin{aligned}
\widehat{\gamma}(\mathbf{f}_s) &> (1-s)\,\widehat{\gamma}(\mathbf{f}_0) + s\,\widehat{\gamma}(\mathbf{f}_1) \\
&= (1-s)\,\widehat{\gamma}_{\mathsf{mx}} + s\,\widehat{\gamma}_{\mathsf{mx}} = \widehat{\gamma}_{\mathsf{mx}}\,.
\end{aligned}$$

But this contradicts the fact that $\widehat{\gamma}_{\mathsf{mx}}$ is the maximum of $\widehat{\gamma}(\mathbf{f})$ over Ω. Therefore at most one maximizer can exist.

Recall that Ω^+ is the intersection of the half spaces

$$1 + \mathbf{r}(d)^\mathsf{T}\mathbf{f} > 0\,, \qquad \text{for } d = 1, \cdots, D\,,$$

and that Ω is the intersection of Ω^+ with the hyperplane $\mathbf{1}^\mathsf{T}\mathbf{f} = 1$.

The set Ω^+ is the intersection of the half-spaces $1 + \mathbf{r}(d)^\mathsf{T}\mathbf{f} > 0$. The set Ω is the intersection of Ω^+ with the hyperplane $\mathbf{1}^\mathsf{T}\mathbf{f} = 1$. For many return histories $\{\mathbf{r}(d)\}_{d=1}^{D}$ the set Ω is bounded. In such cases we will have $1 + \mathbf{r}(d)^\mathsf{T}\mathbf{f} \searrow 0$ for at least one d as $\mathbf{f}$ approaches the boundary of Ω. But then we will have $\log(1 + \mathbf{r}(d)^\mathsf{T}\mathbf{f}) \to -\infty$ for at least one d as $\mathbf{f}$ approaches the boundary of Ω. Therefore we will have $\widehat{\gamma}(\mathbf{f}) \to -\infty$ as $\mathbf{f}$ approaches the boundary of Ω. Therefore $\widehat{\gamma}(\mathbf{f})$ has a maximizer in Ω.

The maximizer of $\widehat{\gamma}(\mathbf{f})$ over Ω can be found numerically by methods that are typically covered in graduate courses. Rather than seek the maximizer of $\widehat{\gamma}(\mathbf{f})$ over Ω, we will replace the estimator $\widehat{\gamma}(\mathbf{f})$ with a new estimator for which finding the maximizer is easier. The hope is that the maximizer of $\widehat{\gamma}(\mathbf{f})$ and the maximizer of the new estimator will be close.

This strategy rests upon the fact that $\widehat{\gamma}(\mathbf{f})$ is itself an approximation. The uncertainties associated with it translate into uncertainities about its maximizer. The hope is that the difference between the maximizers of $\widehat{\gamma}(\mathbf{f})$ and of the new estimator will be within these uncertainties.

The new growth rate estimator will be expressed in terms of the return mean vector $\mathbf{m}$ and return covariance matrix $\mathbf{V}$, so that we can use the efficient frontiers developed earlier.

One way to approximate $\widehat{\gamma}(\mathbf{f})$ is to use the *quadratic Taylor approximation* of $\log(1+r)$ for small r. That approximation is

$$\log(1+r) \approx r - \tfrac{1}{2}r^2\,. \tag{14}$$

When this approximation is used in (12) we obtain the *quadratic growth rate mean estimator*

$$\begin{aligned}
\widehat{\gamma}_{\mathsf{q}}(\mathbf{f}) &= \sum_{d=1}^{D} w(d)\left(\mathbf{r}(d)^{\mathsf{T}}\mathbf{f} - \tfrac{1}{2}(\mathbf{r}(d)^{\mathsf{T}}\mathbf{f})^2\right) \\
&= \left(\sum_{d=1}^{D} w(d)\,\mathbf{r}(d)\right)^{\mathsf{T}}\mathbf{f} - \tfrac{1}{2}\mathbf{f}^{\mathsf{T}}\left(\sum_{d=1}^{D} w(d)\,\mathbf{r}(d)\,\mathbf{r}(d)^{\mathsf{T}}\right)\mathbf{f} \\
&= \mathbf{m}^{\mathsf{T}}\mathbf{f} - \tfrac{1}{2}\mathbf{f}^{\mathsf{T}}\left(\mathbf{m}\mathbf{m}^{\mathsf{T}} + \mathbf{V}\right)\mathbf{f}\,.
\end{aligned} \tag{15}$$

We obtained this estimator twice earlier using the moment and cumulant generating functions.

The following table shows that the quadratic Taylor approximation (14) to $\log(1+r)$ is pretty good when $|r| < .25$ and is not too bad when $.25 < |r| < .5$. It is breaks down completely when $1 \le |r|$.

r	$\log(1+r)$	$r - \frac{1}{2}r^2$	$r - \frac{1}{2}r^2 + \frac{1}{3}r^3$
$-.5$	$-.693$	$-.625$	$-.667$
$-.4$	$-.511$	$-.480$	$-.502$
$-.3$	$-.357$	$-.345$	$-.354$
$-.2$	$-.223$	$-.220$	$-.223$
$-.1$	$-.105$	$-.105$	$-.105$
$.0$	$.000$	$.000$	$.000$
$.1$	$.095$	$.095$	$.095$
$.2$	$.182$	$.180$	$.183$
$.3$	$.262$	$.255$	$.264$
$.4$	$.336$	$.320$	$.341$
$.5$	$.405$	$.375$	$.417$

We can also approximate $\widehat{\gamma}(\mathbf{f})$ by the *second-order Taylor approximation* of $\log(1+r)$ for $r = \mathbf{r}(d)^{\mathsf{T}}\mathbf{f}$ near $\widehat{\mu}(\mathbf{f}) = \mathbf{m}^{\mathsf{T}}\mathbf{f}$. That approximation is

$$\log(1+r) \approx \log\Big(1+\mathbf{m}^{\mathsf{T}}\mathbf{f}\Big) + \frac{(\mathbf{r}(d)-\mathbf{m})^{\mathsf{T}}\mathbf{f}}{1+\mathbf{m}^{\mathsf{T}}\mathbf{f}} - \frac{1}{2}\frac{\Big((\mathbf{r}(d)-\mathbf{m})^{\mathsf{T}}\mathbf{f}\Big)^2}{(1+\mathbf{m}^{\mathsf{T}}\mathbf{f})^2}\,.$$

When this approximation is used in (12) we obtain the estimator

$$\widehat{\gamma}_{\mathsf{s}}(\mathbf{f}) = \log\Big(1+\mathbf{m}^{\mathsf{T}}\mathbf{f}\Big) - \frac{1}{2}\frac{\mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}}{(1+\mathbf{m}^{\mathsf{T}}\mathbf{f})^2}\,. \tag{16}$$

We obtained this estimator earlier using the cumulant generating function.

Remark. The estimator (16) satisfies the upper bound (7) from Fact 3. However, it is not concave and does not have a maximum. This makes it a poor candidate for a new growth rate mean estimator.

A slight modification of (16) yields a much better candidate for a new growth rate mean estimator — namely, the mean-centered estimator

$$\widehat{\gamma}_{\mathsf{m}}(\mathbf{f}) = \log\Big(1 + \mathbf{m}^{\mathsf{T}}\mathbf{f}\Big) - \frac{1}{2}\,\frac{\mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}}{1 + 2\mathbf{m}^{\mathsf{T}}\mathbf{f}}\,, \tag{17}$$

which is defined on the half-space $H = \Big\{\mathbf{f} \in \mathbb{R}^N \,:\, 0 < 1 + 2\mathbf{m}^{\mathsf{T}}\mathbf{f}\Big\}$.

The estimator $\widehat{\gamma}_{\mathsf{m}}(\mathbf{f})$ clearly satisfies the upper bound (7) from Fact 3 for every $\mathbf{f} \in H$. Moreover, we have the following.

Fact 6. $\widehat{\gamma}_{\mathsf{m}}(\mathbf{f})$ is strictly concave over the half-space H.

Proof. This will follow upon showing that $\widehat{\gamma}_{\mathsf{m}}(\mathbf{f})$ is the sum of two functions, the first of which is concave over H and the second of which is strictly concave over H.

The function $\log\left(1 + \mathbf{m}^\mathsf{T}\mathbf{f}\right)$ is infinitely differentiable over H with

$$\nabla_{\mathbf{f}} \log\left(1 + \mathbf{m}^\mathsf{T}\mathbf{f}\right) = \frac{\mathbf{m}}{1 + \mathbf{m}^\mathsf{T}\mathbf{f}}\,,$$
$$\nabla_{\mathbf{f}}^2 \log\left(1 + \mathbf{m}^\mathsf{T}\mathbf{f}\right) = -\frac{\mathbf{m}\,\mathbf{m}^\mathsf{T}}{(1 + \mathbf{m}^\mathsf{T}\mathbf{f})^2}\,.$$

Because its Hessian is nonpositive definite, the function $\log\left(1 + \mathbf{m}^\mathsf{T}\mathbf{f}\right)$ is concave over H.

The harder part of the proof is to show that

$$-\frac{1}{2}\,\frac{\mathbf{f}^\mathsf{T}\mathbf{V}\mathbf{f}}{1 + 2\mathbf{m}^\mathsf{T}\mathbf{f}} \quad \text{is strictly concave over } H\,. \tag{18}$$

This follows from the next two facts. Our proof of **Fact 6** will be completed after those facts are established.

Fact 7. Let $\mathrm{b}, \mathbf{x} \in \mathbb{R}^N$ such that $1 + \mathbf{b}^\top \mathbf{x} > 0$. Then $\mathbf{I} + \mathbf{x}\,\mathbf{b}^\top$ is invertible with

$$\left(\mathbf{I} + \mathbf{x}\,\mathbf{b}^\top\right)^{-1} = \mathbf{I} - \frac{\mathbf{x}\,\mathbf{b}^\top}{1 + \mathbf{b}^\top \mathbf{x}}\,. \tag{19}$$

Proof. Just check that

$$\begin{aligned}
\left(\mathbf{I} + \mathbf{x}\,\mathbf{b}^\top\right)\left(\mathbf{I} - \frac{\mathbf{x}\,\mathbf{b}^\top}{1 + \mathbf{b}^\top \mathbf{x}}\right) &= \left(\mathbf{I} + \mathbf{x}\,\mathbf{b}^\top\right) - \frac{\left(\mathbf{I} + \mathbf{x}\,\mathbf{b}^\top\right)\mathbf{x}\,\mathbf{b}^\top}{1 + \mathbf{b}^\top \mathbf{x}} \\
&= \mathbf{I} + \mathbf{x}\,\mathbf{b}^\top - \frac{\mathbf{x}\,\mathbf{b}^\top + \mathbf{x}\,\mathbf{b}^\top \mathbf{x}\,\mathbf{b}^\top}{1 + \mathbf{b}^\top \mathbf{x}} \\
&= \mathbf{I} + \mathbf{x}\,\mathbf{b}^\top - \frac{1 + \mathbf{b}^\top \mathbf{x}}{1 + \mathbf{b}^\top \mathbf{x}}\,\mathbf{x}\,\mathbf{b}^\top = \mathbf{I}\,.
\end{aligned}$$

The assertions of **Fact 7** then follow. $\square$

Fact 8. Let $\mathbf{A} \in \mathbb{R}^{N\times N}$ be symmetric and positive definite. Let $\mathbf{b} \in \mathbb{R}^N$. Let X be the half-space given by

$$X = \left\{ \mathbf{x} \in \mathbb{R}^N \; : \; 1 + \mathbf{b}^\mathsf{T}\mathbf{x} > 0 \right\}.$$

Then

$$\phi(\mathbf{x}) = \frac{1}{2}\,\frac{\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}}{1 + \mathbf{b}^\mathsf{T}\mathbf{x}} \quad \text{is strictly convex over } X\,.$$

Proof. The function $\phi(\mathbf{x})$ is infinitely differentiable over X with

$$\nabla_{\mathbf{x}}\,\phi(\mathbf{x}) = \frac{\mathbf{A}\mathbf{x}}{1 + \mathbf{b}^\mathsf{T}\mathbf{x}} - \frac{1}{2}\,\frac{\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}\,\mathbf{b}}{(1 + \mathbf{b}^\mathsf{T}\mathbf{x})^2}\,,$$

$$\nabla_{\mathbf{x}}^2\phi(\mathbf{x}) = \frac{\mathbf{A}}{1 + \mathbf{b}^\mathsf{T}\mathbf{x}} - \frac{\mathbf{A}\mathbf{x}\,\mathbf{b}^\mathsf{T} + \mathbf{b}\,\mathbf{x}^\mathsf{T}\mathbf{A}}{(1 + \mathbf{b}^\mathsf{T}\mathbf{x})^2} + \frac{\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}\,\mathbf{b}\,\mathbf{b}^\mathsf{T}}{(1 + \mathbf{b}^\mathsf{T}\mathbf{x})^3}\,.$$

Then using (19) of **Fact 7** the Hessian can be expressed as

$$\begin{aligned}\nabla^2_{\mathbf{x}}\phi(\mathbf{x}) &= \left(\mathbf{I} - \frac{\mathbf{b}\,\mathbf{x}^\top}{1+\mathbf{b}^\top\mathbf{x}}\right)\frac{\mathbf{A}}{1+\mathbf{b}^\top\mathbf{x}}\left(\mathbf{I} - \frac{\mathbf{x}\,\mathbf{b}^\top}{1+\mathbf{b}^\top\mathbf{x}}\right) \\ &= \left(\mathbf{I}+\mathbf{x}\,\mathbf{b}^\top\right)^{-\top}\frac{\mathbf{A}}{1+\mathbf{b}^\top\mathbf{x}}\left(\mathbf{I}+\mathbf{x}\,\mathbf{b}^\top\right)^{-1}.\end{aligned}$$

Because $\mathbf{A}$ is positive definite and $1+\mathbf{b}^\top\mathbf{x} > 0$ for every $\mathbf{x} \in X$, this shows that $\nabla^2_{\mathbf{x}}\phi(\mathbf{x})$ is positive definite for every $\mathbf{x} \in X$. Therefore $\phi(\mathbf{x})$ is strictly convex over X, thereby proving **Fact 8**. $\square$.

By setting $\mathbf{A} = \mathbf{V}$ and $\mathbf{b} = 2\mathbf{m}$ in **Fact 8** and using the fact that the negative of a strictly convex function is strictly concave, we establish (18), thereby completing the proof of **Fact 6**. $\square$

Next, we identify a class of solvent Markowitz portfolios whose allocations lie within H.

Fact 9. $\Omega_{\frac{1}{2}} = \left\{ \mathbf{f} \in \Omega \,:\, \frac{1}{2} \leq 1 + \mathbf{r}(d)^{\mathsf{T}}\mathbf{f} \;\; \forall d \right\} \subset H$.

Proof. Because $\Omega_{\frac{1}{2}} = \left\{ \mathbf{f} \in \Omega \,:\, 0 \leq 1 + 2\mathbf{r}(d)^{\mathsf{T}}\mathbf{f} \;\; \forall d \right\}$, it is clear that $0 \leq 1 + 2\mathbf{m}^{\mathsf{T}}\mathbf{f}$ for every $\mathbf{f} \in \Omega_{\frac{1}{2}}$ with equality if only if $0 = 1 + 2\mathbf{r}(d)^{\mathsf{T}}\mathbf{f}$ for every d. But this implies that $(\mathbf{r}(d) - \mathbf{m})^{\mathsf{T}}\mathbf{f} = 0$ for every d, which implies that $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ does not span $\mathbb{R}^N$, which contradicts the assumption that $\mathbf{V}$ is positive definite. Therefore $0 < 1 + 2\mathbf{m}^{\mathsf{T}}\mathbf{f}$ for every $\mathbf{f} \in \Omega_{\frac{1}{2}}$. $\square$

Remark. This class excludes portfolios that would have dropped 50% in value during a single trading day over the history considered. This seems like a reasonable constraint for any long-term investor.

Finally, we introduce an estimator that is a hybrid of the quadratic estimator (15) and the mean-centered estimator (17). Specifically, by using the first term from the mean-centered estimator (17) and the volatility term from the quadratic estimator (15) we obtain

$$\widehat{\gamma}_{\mathrm{h}}(\mathbf{f}) = \log\Big(1 + \mathbf{m}^{\mathsf{T}}\mathbf{f}\Big) - \frac{1}{2}\,\mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}\,. \tag{20}$$

This estimator is defined over the half-space where

$$1 + \mathbf{m}^{\mathsf{T}}\mathbf{f} > 0\,.$$

This contains the half-space H over which the mean-centered estimator (17) was defined. It also contains Ω, the set of allocations for solvent Markowitz portfolios.

Remark. This hybrid estimator satisfies the upper bound (7) and is strictly concave over the half-space on which it is defined. It is easier to use than the mean-centered estimator (17), but relatively underestimates the risk when $\mathbf{m}^{\mathsf{T}}\mathbf{f} < 0$.

One Risk-Free Rate Model. We now consider solvent Markowitz portfolios with risk-free assets. We use the one risk-free rate model with daily return μ_{rf}. The associated allocations $\mathbf{f}$ belong to Ω^{+}. The growth rate mean estimator (6) is

$$\widehat{\gamma}(\mathbf{f}) = \sum_{d=1}^{D} w(d) \, \log\Big(1 + r(d, \mathbf{f})\Big) \,, \tag{21}$$

where from (1) we have

$$r(d, \mathbf{f}) = \mu_{\mathsf{rf}}\Big(1 - \mathbf{1}^{\mathsf{T}}\mathbf{f}\Big) + \mathbf{r}(d)^{\mathsf{T}}\mathbf{f} \,. \tag{22}$$

This can be recast as

$$\widehat{\gamma}(\mathbf{f}) = \sum_{d=1}^{D} w(d) \, \log\Big(1 + \mu_{\mathsf{rf}} + \tilde{\boldsymbol{r}}(d)^{\mathsf{T}}\mathbf{f}\Big) \,, \tag{23}$$

where the so-called *excess returns* $\tilde{\boldsymbol{r}}(d)$ are given by

$$\tilde{\boldsymbol{r}}(d) = \mathbf{r}(d) - \mu_{\mathsf{rf}}\mathbf{1} \,. \tag{24}$$

The entries of $\tilde{\boldsymbol{r}}(d)$ are called excess returns because they are the returns in excess of the return of the risk-free asset. The excess return mean $\tilde{\boldsymbol{m}}$ is given by

$$\begin{aligned}\tilde{\boldsymbol{m}} &= \sum_{d=1}^{D} w(d)\,\tilde{\boldsymbol{r}}(d) = \sum_{d=1}^{D} w(d)\left(\mathbf{r}(d) - \mu_{\mathsf{rf}}\mathbf{1}\right)\\ &= \sum_{d=1}^{D} w(d)\,\mathbf{r}(d) - \mu_{\mathsf{rf}}\mathbf{1}\sum_{d=1}^{D} w(d) = \mathbf{m} - \mu_{\mathsf{rf}}\mathbf{1}\,.\end{aligned} \tag{25}$$

Because $\tilde{\boldsymbol{r}}(d) - \tilde{\boldsymbol{m}} = \mathbf{r}(d) - \mathbf{m}$ for every d, the excess return covariance matrix $\tilde{\boldsymbol{V}}$ equals the return covariance matrix $\mathbf{V}$. In other words,

$$\begin{aligned}\tilde{\boldsymbol{V}} &= \sum_{d=1}^{D} w(d)\left(\tilde{\boldsymbol{r}}(d) - \tilde{\boldsymbol{m}}\right)\left(\tilde{\boldsymbol{r}}(d) - \tilde{\boldsymbol{m}}\right)^{\mathsf{T}}\\ &= \sum_{d=1}^{D} w(d)\left(\mathbf{r}(d) - \mathbf{m}\right)\left(\mathbf{r}(d) - \mathbf{m}\right)^{\mathsf{T}} = \mathbf{V}\,.\end{aligned}$$

We now want to develop estimators for $\widehat{\gamma}(\mathbf{f})$ given by (23) in terms of $\mathbf{m}$ and $\mathbf{V}$, or what is equivalent, in terms of $\tilde{\boldsymbol{m}}$ and $\mathbf{V}$. The key to doing this quickly is the following observation.

The law of logarithms yields

$$\log\Bigl(1 + \mu_{\mathsf{rf}} + \tilde{\boldsymbol{r}}(d)^{\mathsf{T}}\mathbf{f}\Bigr) = \log\Bigl(1 + \mu_{\mathsf{rf}}\Bigr) + \log\left(1 + \frac{\tilde{\boldsymbol{r}}(d)^{\mathsf{T}}\mathbf{f}}{1 + \mu_{\mathsf{rf}}}\right) .$$

Therefore $\widehat{\gamma}(\mathbf{f})$ given by (23) may be expressed as

$$\widehat{\gamma}(\mathbf{f}) = \log\Bigl(1 + \mu_{\mathsf{rf}}\Bigr) + \sum_{d=1}^{D} w(d)\, \log\left(1 + \frac{\tilde{\boldsymbol{r}}(d)^{\mathsf{T}}\mathbf{f}}{1 + \mu_{\mathsf{rf}}}\right) . \qquad (26)$$

Notice that the sum above has the same form as the growth rate estimator (12) with

$$\frac{\tilde{\boldsymbol{r}}(d)}{1 + \mu_{\mathsf{rf}}} \quad \text{replacing} \quad \mathbf{r}(d) .$$

Therefore we simply can use the same estimators as before with

$$\frac{\tilde{\boldsymbol{m}}}{1+\mu_{\mathsf{rf}}} \quad \text{and} \quad \frac{\mathbf{V}}{(1+\mu_{\mathsf{rf}})^2} \quad \text{replacing} \quad \mathbf{m} \quad \text{and} \quad \mathbf{V}\,. \tag{27}$$

For example, if we apply recipe (27) to the quadratic estimator (15) then we obtain the quadratic estimator

$$\begin{aligned} \widehat{\gamma}_{\mathsf{q}}(\mathbf{f}) &= \log\Big(1+\mu_{\mathsf{rf}}\Big) + \frac{\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f}}{1+\mu_{\mathsf{rf}}} - \frac{1}{2}\,\frac{(\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f})^2 + \mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}}{(1+\mu_{\mathsf{rf}})^2} \\ &= \log\Big(1+\mu_{\mathsf{rf}}\Big) + \frac{\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f}}{1+\mu_{\mathsf{rf}}} - \frac{1}{2}\,\frac{\mathbf{f}^{\mathsf{T}}\Big(\tilde{\boldsymbol{m}}\,\tilde{\boldsymbol{m}}^{\mathsf{T}} + \mathbf{V}\Big)\mathbf{f}}{(1+\mu_{\mathsf{rf}})^2}\,. \end{aligned} \tag{28}$$

This estimator is defined over $\mathbb{R}^N$.

If we apply recipe (27) to the mean-centered estimator (17) then we obtain the mean-centered estimator

$$\begin{aligned}\widehat{\gamma}_{\mathrm{m}}(\mathbf{f}) &= \log\!\left(1+\mu_{\mathrm{rf}}\right) + \log\!\left(1+\frac{\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f}}{1+\mu_{\mathrm{rf}}}\right) - \frac{1}{2}\,\frac{\dfrac{\mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}}{(1+\mu_{\mathrm{rf}})^2}}{1+2\,\dfrac{\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f}}{1+\mu_{\mathrm{rf}}}} \qquad (29)\\ &= \log\!\left(1+\mu_{\mathrm{rf}}+\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f}\right) - \frac{1}{2}\,\frac{\dfrac{\mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}}{1+\mu_{\mathrm{rf}}}}{1+\mu_{\mathrm{rf}}+2\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f}}\,.\end{aligned}$$

This estimator is defined over the half-space where

$$1+\mu_{\mathrm{rf}}+2\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f} > 0\,.$$

If we apply recipe (27) to the hybrid estimator (20) then we obtain the hybrid estimator

$$
\begin{aligned}
\widehat{\gamma}_{\mathsf{h}}(\mathbf{f}) &= \log\big(1+\mu_{\mathsf{rf}}\big) + \log\left(1+\frac{\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f}}{1+\mu_{\mathsf{rf}}}\right) - \frac{1}{2}\,\frac{\mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}}{(1+\mu_{\mathsf{rf}})^2} \\
&= \log\big(1+\mu_{\mathsf{rf}}+\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f}\big) - \frac{1}{2}\,\frac{\mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}}{(1+\mu_{\mathsf{rf}})^2}\,.
\end{aligned}
\tag{30}
$$

This estimator is defined over the half-space where

$$
1+\mu_{\mathsf{rf}}+\tilde{\boldsymbol{m}}^{\mathsf{T}}\mathbf{f} > 0\,.
$$

This contains the half-space over which the mean-centered estimator (29) was defined. It also contains Ω, the set of allocations for solvent Markowitz portfolios.