ICEBERG RISK

An Adventure in Portfolio Theory
Final Instalment

Kent Osband
kosband@nyc.rr.com

June 2002

PRE-PUBLICATION REVIEW MANUSCRIPT
EXCLUSIVE TO READERS OF WILMOTT.COM
TO BE PUBLISHED BY TEXERE IN AUGUST 2002
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Preface</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part One: Standards of Deviation</strong></td>
<td></td>
</tr>
<tr>
<td>“Missing the Iceberg for the Ice Cubes”</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 1: Odd Odds of Odds</td>
<td>10</td>
</tr>
<tr>
<td>Chapter 2: New Angles on Pascal’s Triangle</td>
<td>22</td>
</tr>
<tr>
<td>Chapter 3: Amazing Tails of Risk</td>
<td>33</td>
</tr>
<tr>
<td>Chapter 4: More Amazing Tails</td>
<td>44</td>
</tr>
<tr>
<td>Chapter 5: The Creation of Correlation</td>
<td>57</td>
</tr>
<tr>
<td>Chapter 6: Defining Moments from De Finetti</td>
<td>69</td>
</tr>
<tr>
<td>Chapter 7: Big Risks in Value at Risk</td>
<td>80</td>
</tr>
<tr>
<td>Chapter 8: Good Approximations Behaving Badly</td>
<td>91</td>
</tr>
<tr>
<td>Chapter 9: The Abnormality of Normality</td>
<td>102</td>
</tr>
<tr>
<td>Chapter 10: Dependent Independence</td>
<td>116</td>
</tr>
<tr>
<td>Chapter 11: Just Desserts</td>
<td>129</td>
</tr>
<tr>
<td><strong>Part 2: Insights into Ignorance</strong></td>
<td></td>
</tr>
<tr>
<td>“More Dollars than Sense”</td>
<td>139</td>
</tr>
<tr>
<td>Chapter 12: Teaching Elephants to Dance</td>
<td>146</td>
</tr>
<tr>
<td>Chapter 13: Betting on Beta</td>
<td>160</td>
</tr>
<tr>
<td>Chapter 14: Relief through Beliefs</td>
<td>172</td>
</tr>
<tr>
<td>Chapter 15: Rating Risks without Regret</td>
<td>185</td>
</tr>
<tr>
<td>Chapter 16: Unusually Useful Utility</td>
<td>196</td>
</tr>
<tr>
<td>Chapter 17: Optimal Overlays</td>
<td>206</td>
</tr>
<tr>
<td>Chapter 18: Adjusted Advice</td>
<td>218</td>
</tr>
<tr>
<td>Chapter 19: Higher-Order High Jinks</td>
<td>229</td>
</tr>
<tr>
<td>Chapter 20: A Question of Rebalance</td>
<td>242</td>
</tr>
<tr>
<td>Chapter 21: Weighing the Options</td>
<td>255</td>
</tr>
<tr>
<td>Chapter 22: Fixing the Focus</td>
<td>268</td>
</tr>
<tr>
<td>Chapter 23: The Rooster Principle</td>
<td>282</td>
</tr>
</tbody>
</table>
17. OPTIMAL OVERLAYS

Conway didn’t tell anyone at work about his Saturday outing. While investment bankers are confirmed workaholics, asset managers consider it bad form to work weekends. And if they knew I was consorting with crazy people… Might as well wait to see what Devlin comes up with. For once Conway was glad to have a big inbox. He devoted himself to whittling it down, and the week sped by.

The next Saturday, when Conway arrived back at Club Med, Devlin and Regretta were waiting for him. “What took you so long?” asked Devlin.

“I took the first ferry I could. They don’t leave until 10 AM on weekends.”

“Never mind,” said Regretta. “That’s just Devlin’s way of saying he’s glad to see you.”


Devlin hid a blush. “Hey, cut the small talk. We’ve got work to do.” Flashing a wad of papers, he led them back to the lounge. “Are you ready for the answer, Conway?”

“I’m all ears.”

“Partition functions.”

“Partition which functions into what?”

“No, no. That’s their name. Partition functions. They’re weighted sums of exponentials.”

“Never heard of them.”

“I hadn’t either until I read Schrödinger on statistical thermodynamics. All we have to do is reinterpret the variables.”

“And what do you do with these partition functions?”

“Compare their values using different portfolio weights. The lowest value maximizes the overall certainty equivalent.”

“And that’s it?”

“As a first pass, yes.” Devlin beamed

Conway was dazed. Before he could respond, Regretta cut in. “Excuse me, Devlin. Now that you’ve told Conway the answer, would you mind telling me the question?”

“No problem. The question is: How do you combine CRR utility with conditional multivariate normality to quickly rank portfolios?”

“What makes you so keen on conditional multivariate normality?”
“Regretta, I’m surprised you should ask. I thought I explained that clearly at group therapy last month.”

“Gee, Devlin, I must have spaced out. I’m sure it was fascinating,” replied Regretta in mock seriousness.

“It was to me. It started back when I was working for Conway at Megabucks and we had to estimate value at risk on three binary options that were uncorrelated but not independent…”

Oh no, I’ve got to get us back on track. Conway cut in. “In a nutshell, Regretta, Devlin and I like conditional normality because it’s flexible enough to approximate nearly any kind of portfolio distribution without forcing you to estimate a lot of higher-order cross-moments.”

“Thanks,” said Regretta.

“Conditional multivariate normality,” corrected Devlin. “Let me explain the differences…”

“Later,” said Conway. “You and Regretta will have plenty of time after I leave.”

“Ooh, I can hardly wait,” said Regretta. “So, Devlin, how do partition functions help you get the answer?”

“They don’t help you get the answer. They are the answer. Since the expected utility in any given regime is exponential, the overall expected utility is just the weighted average of exponentials, with a minus sign stuck in front. And that’s what partition functions are, except for the sign.”

“Hold on, Devlin,” said Conway. We went through that last time. Expected CAR utility is exponential given normal distributions. Expected CRR utility isn’t. It isn’t even defined for normal distributions.”

“That depends on how you measure the portfolios.”

“You’ve been reading too much physics, Devlin. Whatever currency or time frame you choose, it’s still not going to allow you negative wealth.”

“We don’t need wealth to be negative. Unbounded returns will work nearly as well.”

“Fine, but returns have to be bounded too.”

“Not if you measure in logarithms.”

“What difference does that make?”

“Not much for changes of less than 10%. But the differences can be huge. Here are two charts I prepared to illustrate.”
“Very pretty, Devlin, but I am familiar with logarithms. What’s your point?”

“Think about it. If we focus on log returns rather than percentage returns, we don’t have to worry about infeasible values. Normality is allowed again, which makes it a lot easier to calculate expected utility.”

“I don’t see how. You still have to integrate over a negative power function times an exponential. I’m not familiar with any reduced form for that.”

“Oh, I see,” said Regretta. “While CRR utility is a power function of gross percentage returns, it’s an exponential function of log returns.”

“Bingo. So if we assume log returns are normal, what does the expression for CRR expected utility look like?”

“It looks like the expression for CAR expected utility, except that you’re integrating over log returns instead of absolute wealth. So it’s essentially a moment-generating function for the distribution of log returns.”

“Which works out to what? Here’s a pen and paper if you need it.”

Regretta took the pen and paper and worked through the integration. “It’s an exponential. That makes the overall expected utility a weighted average of these exponentials, with the weights given by the probabilities of the various regimes. There’s your partition function.”

“You got it, except that the exponentials are typically preceded by a minus sign. Now let me ask you a harder question. What does each exponent in each exponential tell you?”

Regretta went back to her paper and experimented rearranging terms. “It’s a multiple of the certainty equivalent in that regime, the guaranteed return that will yield the same expected utility.”
“Guaranteed log return,” corrected Devlin. “And what multiple of the certainty equivalent is it?”

“The power in the power function, which equals one minus the CRR.”

“That’s right. What does that make the formula for certainty equivalent in each regime?”

“The conditional mean log return less ½(CRR-1) times the conditional variance.”

“So what is the economic interpretation of ½(CRR-1)?”

“The price of variance in terms of an equivalent guaranteed log loss. The more risk averse you are, the higher the price.”

“Good job, Regretta. Perfect. Do you understand now, Conway?”

“Not completely. When we first tried to calculate expected CRR utility under normality, the answer was negative infinity. How did remeasuring things in log terms get rid of it?”

“I didn’t just remeasure. I redefined the probability density so that log returns are normal instead of percentage returns.”

“How do you know that’s better?”

“With most daily returns you can’t tell the difference. For huge crashes, though, lognormality won’t let you wind up with less than nothing while normality will.”

“I grant you lognormality is better in that respect. But who’s to say lognormality doesn’t screw up big somewhere else?”

“There’s also a theoretical reason for favoring long-term lognormality. It’s the difference between multiplication and addition.”

_I wish Devlin’s explanations needed less explanation._ “Meaning?”

“The cumulative gross return equals the product of the gross return over each subperiod, so the logs are additive. If the subperiod returns are independent and identically distributed, the long-term log is bound by the Central Limit Theorem to approach normality, making the cumulative gross return lognormal.”

“I see. So what are the implications for horizons of a few quarter to a few years?”

“The only noticeable difference is that lognormality adds skewness. Here’s an example I charted where log returns are normally distributed with mean 10% and standard deviation 20%. The dark heavy line marks their density. The solid area represents the corresponding lognormal density for percentage returns.”
Conway pondered the chart. “You know, it’s really not such a big deal. Strange we didn’t think of this before.”

“That’s because we were thinking too much like standard portfolio theorists. If we had been options theorists we would have thought of it right away. They nearly always assume lognormality.”

“Why don’t portfolio theorists assume it too?”

“Because portfolios of lognormal assets are a lot harder to deal with than portfolios of multivariate normal assets.”

“Really? In what way?”

“Portfolios of lognormal assets aren’t lognormal. They aren’t normal either.”

“Not even if they’re multivariate lognormal?”

“Not even then. They’re sort of part-normal, part-lognormal, and how much they lean toward one or the other depends on how many assets you have and how big the weights are.”

“It doesn’t sound very tractable.”

“It isn’t. So when options theorists deal with portfolios—currency baskets, say—they typically just gloss over the differences between logs and percentages. Specifically, they assume the portfolio is lognormal, with a log mean that’s the weighted average of the asset log means and a variance that’s a quadratic form in the covariance matrix.”
“How good an approximation is that?”

“It’s a first-order Taylor’s approximation: great if the changes are small enough, awful if they’re huge. To take an extreme example, suppose an asset comprising 1% of your portfolio completely bites the dust, while the other 99% holds its value. What’s the log portfolio return?”

“Minus 1%.”

“Of course. But that’s not what the standard approximation would tell us. The log return on the failing asset would be minus infinity, so that the average of the logs would have to be minus infinity too, making the whole portfolio look worthless.”

“Ahhh, so that’s why portfolio theory tries to avoid log returns.”

“Even while the other half of finance theory can’t live without it. It’s schizophrenic, isn’t it?”

Conway nodded. And here I had to come to a madhouse to find out. “So what do you propose to do?”

“For now I’m simply going to ignore the discrepancy too.”

“Did I hear that right? You, Devlin, going with a standard practice you know is wrong? It’s hard enough getting you to follow standard practice when it’s right.

Devlin smiled. “Give me time. I can’t see any way around it yet. Besides, it’s not as if I’m confusing something that single-regime models keep straight.”

“Speaking of keeping things straight,” said Regretta, “I’m having some trouble myself. Do you mind quickly reviewing your multi-regime alternative?”

“No problem. Estimate the conditional mean and variance using the log-linear approximation for portfolios. Subtract a multiple of the variance from the mean to form the certainty equivalent for every regime. Convert to conditional expected utilities and calculate total expected utility as their probability-weighted average. Then pick the portfolio weights to maximize the total expected utility.”

“Each step sounds easy. Still...”

“It is easy. Like baking a layer cake. Here’s a recipe I printed up to help you both remember. I put tildes on most of the variables to remind you that we’re dealing with log returns rather than the more commonly used percentage returns.”
Quick Recipe for Portfolio Optimization

1) Divide world into various conditionally lognormal regimes, assigning probability $p_k$ to each regime $k$.

2) For each regime, estimate the vector mean $\bar{M}_k$ and the covariance matrix $\bar{\Sigma}_k$ of excess log returns.

3) Given portfolio weights $\omega$, estimate the conditional log mean $\bar{m}_k$ of the portfolio as $\omega'\bar{M}_k$ and the conditional variance $\bar{\sigma}_k$ as $\omega'\bar{\Sigma}_k\omega$.

4) Given constant relative risk aversion $c>1$, estimate the conditional log certainty equivalent $\bar{C}E_k$ as $\bar{m}_k - \frac{1}{2}(c-1)\bar{\sigma}_k$.

5) Calculate the conditional expected utility $\bar{EU}_k$ as $-\exp\left(\frac{1}{1-c}\bar{C}E_k\right)$.

6) Calculate the aggregate expected utility $\bar{EU}$ as the probability-weighted average $\sum p_k\bar{EU}_k$.

7) Calculate the aggregate log certainty equivalent $\bar{C}E$ as logarithm of $|\bar{EU}|$ divided by $1-c$.

8) Choose $\omega$ to maximize $\bar{EU}$ or $\bar{C}E$.

Conway read over the recipe. “Thanks, Devlin. But all this talk about cakes and recipes makes me hungry. How about you?”

“Famished,” said Devlin. “Let’s go to lunch.”

“I’m not that hungry but I’ll join you,” said Regretta. And they headed off to the dining room.

While they’re eating, let’s chew over these partition functions. Given the tools we’ve developed already, they won’t be hard to digest. The biggest challenge is to keep the distinction between logarithms and percentages only partially straight. Think of it as an exercise in studied carelessness. We’ll tidy up later.
Terminology

For $P_t$, the price of an asset or portfolio at time $t$ net of the risk-free rate (that is, the nominal price divided by a cumulative risk-free index), let $x_t$ denote the excess percentage return $\frac{P_t - P_{t-1}}{P_{t-1}}$ and $\tilde{x}_t$, the logarithmic excess return $\ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(1 + x_t)$. I will usually shorten “logarithm” to “log” and drop the time subscript as understood. I will denote the vector counterpart of $x$ by $\tilde{X} = \ln(1 + X)$ and indeed use $\sim$ (tilde) to identify log counterparts more generally. However, do not assume that the log moments equal the logs of one plus the corresponding percentage moments. Due to nonlinearity this will hardly ever be the case.

As a shorthand I will also sometimes describe the percentage mean $\omega'M$ and variance $\omega'\Sigma\omega$ of a portfolio by the letters $m$ and $v$, without explicitly including the risky asset weights $\omega$ on which $m$ and $v$ depend. To be more precise, $m$, $v$, and their log counterparts will serve as placeholders for various approximations. Similarly, $\omega^*$ will identify an approximately optimal portfolio mix. Most of both this appendix and the next are devoted to working out and refining the approximations.

Differences between Log and Percentage Change

A Taylor series expansion shows that:

$$\tilde{x} \equiv \ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots.$$  

When $x$ is close to zero, the higher-order terms in the Taylor expansion will be small compared with the first, so that $\tilde{x}$ will approximately equal $x$. The gap $x - \tilde{x}$ is only 0.0002, or 2 basis points (bps) at $x=2\%$, 12 bps at 5%, and 47 bps at 10%. As $x$ gets larger the gap widens rapidly: 600 bps at $x=40\%$ and 9000 bps at $x=200\%$. The widening is even more dramatic when $x$ is negative: 1100 bps at $x=-40\%$ and infinite at $x=-100\%$.

Two Notions of Average Change

If the average log changes have been zero, the ending price will equal the starting price, because $\ln\left(\frac{P_t}{P_0}\right) = \ln\left(\frac{P_t}{P_{t-1}}\right) + \ln\left(\frac{P_{t-1}}{P_{t-2}}\right) + \cdots + \ln\left(\frac{P_1}{P_0}\right) = \sum_i \tilde{x}_i = 0$. Since percentage changes always exceed log changes except at zero, the average percentage change will be positive even though the price has returned to its original value. In an extreme case, average percentage changes can be positive even when the asset is now worthless. This makes percentage change a misleading measure in finance.

However, log changes can be misleading too. Consider a portfolio composed of many equally weighted assets. The portfolio will be stable if and only if the average percentage change on each asset is zero. Now if the average percentage change is zero, the average log change will be zero unless every price is still. So logarithms will underestimate the changes in a portfolio. In an
extreme case, if one of the assets becomes worthless, the average log change will be $-\infty$ even though the portfolio as a whole might have appreciated.

**Average Log Change and Relative Risk Aversion**

Average changes can be related to relative risk aversion as follows. An investor is risk neutral with a CRR coefficient $c$ of 0 if she is indifferent to any bet with a zero expected percentage return. Her $c$ equals 1 if she is indifferent to any zero expected return to her log wealth—for example, even odds of halving her wealth or doubling it. Conversely, an investor whose $c$ exceeds 1, as is considered typical, would rather hold cash than expect the average logarithm of her wealth to drift downward. As noted earlier, $c<1$ would switch the sign of the partition function.

**Lognormality**

A normal distribution is unbounded. That is, any value is feasible, including extremely negative ones. That can’t be true of prices. So analysts typically assume logarithmic normality. A gross percentage return $1+x$ is said to be lognormally distributed if its log is normally distributed. That is, the cumulative distribution $F$ is given by:

$$F(1 + x) = \int_{-\infty}^{\ln(1 + x)} \frac{1}{\sigma \sqrt{2\pi}} \cdot \exp \left( -\frac{(\tilde{x} - \tilde{\mu})^2}{2\tilde{\sigma}^2} \right) d\tilde{x}$$

where $\tilde{\mu}$ and $\tilde{\sigma}^2$ denote the mean and variance respectively of $\tilde{x} = \ln(1 + x)$. Differentiating both sides shows that

$$f(1 + x) = \frac{1}{\sigma (1 + x) \sqrt{2\pi}} \cdot \exp \left( -\frac{(\ln(1 + x) - \tilde{\mu})^2}{2\tilde{\sigma}^2} \right)$$

**Moments of a Lognormal Distribution**

The $n^{th}$ moment of a lognormal distribution can be calculated as:

$$E[(1 + x)^n] = \int_0^{\infty} \frac{(1 + x)^n}{\sigma \sqrt{2\pi}} \cdot \exp \left( -\frac{(\ln(1 + x) - \tilde{\mu})^2}{2\tilde{\sigma}^2} \right) dx = \int_{-\infty}^{\infty} \frac{\exp(n\tilde{x})}{\tilde{\sigma} \sqrt{2\pi}} \cdot \exp \left( -\frac{\tilde{x}^2}{2\tilde{\sigma}^2} \right) d\tilde{x}$$

where $\mathcal{M}(\cdot)$ denotes the moment-generating function for a normal distribution. It follows that:

- $E[x] = \exp\left(\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}\right) - 1$. A Taylor series expansion indicates that the mean percentage change exceeds the mean log change by $\frac{1}{2}\left(\tilde{\mu}^2 + \tilde{\sigma}^2\right)$ plus higher-order terms.
• \( \text{Var}[x] = \text{Var}[1+x] = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2) = \exp(2\mu + \sigma^2) - 1 \), which reduces to \( \sigma^2 + 2\mu\sigma^2 \) plus higher-order terms.

• Skewness \( E \left[ \left( \frac{x - E[x]}{\sqrt{\text{Var}[x]}} \right)^3 \right] \) works out to—I will spare readers the gory details—

\[
\left( \exp(\sigma^2) + 2 \right) \sqrt{\exp(\sigma^2) - 1} \), which reduces to \( 3\sigma \) plus higher-order terms. So the lognormal density is indeed positively skewed, but only modestly if \( \sigma \) is small.

• Kurtosis \( E \left[ \left( \frac{x - E[x]}{\sqrt{\text{Var}[x]}} \right)^4 \right] \) works out to—again I shall be gracious—

\[
2 \exp(3\sigma^2) + 3 \exp(2\sigma^2) - 6 \), which reduces to \( 16\sigma^2 \) plus higher-order terms.

**Dependence on Time Periods**

Finance theory generally assumes that the underlying laws of motion are stable and that price movements at different, non-overlapping times are uncorrelated. In that case, mean returns and variances will both scale linearly over time, while standard deviations will scale with the square root of time. To verify this, decompose a given change into the sum of uncorrelated component changes and apply the formulas for moments of sums.

At short time periods \( \sigma \) will dominate \( \mu \) even though they both are shrinking toward zero. In that case, lognormal and normal densities will essentially match subject to a mean shift of \( \frac{1}{2}\sigma^2 \). Conversely, over long horizons lognormal densities will be far more upward skewed and fat-tailed than normal. In practice, asset allocation decisions tend to focus on intermediate time periods, where one can reasonably convert from one form to the other but a mean shift of \( \frac{1}{2}\sigma^2 \) may not suffice.

**Multivariate Lognormality**

A vector \( 1+X \) is said to be multivariate lognormal if its log \( L \equiv \ln(1+X) \) is multivariate normal. I denote the latter’s mean and variance by \( \bar{M} \) and \( \bar{\Sigma} \) respectively. Multivariate normality of the logs is of course equivalent to saying that every linear combination \( \omega'X \) is normal with mean \( \omega'\bar{M} \) and variance \( \omega'\bar{\Sigma}\omega \).

**Portfolios of Lognormal Assets**

Let’s consider a one-shot portfolio game. An allocation is made among risky assets with percentage weights \( \omega \), and the portfolio is held undisturbed for one period. The net excess percentage returns on the portfolio then takes the simple form \( \omega'X \). Unfortunately, the corresponding log return is not \( \omega'\bar{X} \). Rather, it equals \( \ln(1 + \omega'X) = \ln \left( 1 + \omega' \left( \exp(\bar{X}) - 1 \right) \right) \).
That makes portfolios of lognormal assets messier to deal with than portfolios of normal assets, at least in one-shot games.

Moreover, portfolios of lognormal assets are hardly ever lognormal, and don’t fit any other standard distribution type either. However, this shortcoming is not peculiar to lognormality. Any portfolio that can be neatly characterized by a few summary parameters, regardless of the numbers of assets or their weights, has to be 100% multivariate normal. To make any other approach tractable we have to look for approximations.

**First Approximations for Lognormality**

To deal with portfolios of lognormal assets, finance typically encourages some selective amnesia. One approach treats $X$ as multivariate normal, ignoring the ludicrous implications for long-term asset prices or likelihoods of negative wealth. The other approach treats log portfolio returns as if they were $\omega' \tilde{X}$ instead of $\ln(1+\omega'X)$. Using the first approach, the portfolio percentage mean return and variance equal $\omega' \tilde{M}$ and $\omega' \tilde{\Sigma} \omega$ respectively. Using the second approach, the log portfolio mean and variance are estimated as $\omega' \tilde{M}$ and $\omega' \tilde{\Sigma} \omega$ respectively.

We can semi-justify either approach by taking a very short time period and ignoring possible jumps. Applying successive first-order Taylor series approximations reduces the logarithmic return $\ln(1+\omega'X)$ to the net percentage return $\omega'X = \omega' \left( \exp \left( \tilde{X} \right) - 1 \right)$ and in turn to the average log return $\omega' \tilde{X}$. I say “semi-justify” because the time horizons and risks we’re interested in warn us not to ignore second-order terms and jumps. But Devlin’s approach already addresses much of that by allowing for multiple regimes. So I will follow Devlin’s lead and accept the approximations as a starting point.

**Expected CRR Utility under Lognormality**

CRR utility can be written as $\text{sign}(1-c) \cdot (1+x)^{1-c} = \text{sign}(1-c) \cdot \exp \left( (1-c) \tilde{x} \right)$. Given lognormality, expected CRR utility given by:

$$EU = \text{sign}(1-c) \cdot \int_{-\infty}^{\infty} \frac{\exp \left( (1-c) \tilde{x} \right)}{\sigma \sqrt{2\pi}} \cdot \exp \left( -\frac{\left( \tilde{x} - \tilde{\mu} \right)^2}{2\tilde{\sigma}^2} \right) d\tilde{x}$$

$$= \text{sign}(1-c) \cdot \exp \left( (1-c) \tilde{\mu} + \frac{1}{2}(1-c)^2 \tilde{\sigma}^2 \right) \cdot \exp \left( (1-c) \omega' \tilde{M} + \frac{1}{2}(1-c)^2 \omega' \tilde{\Sigma} \omega \right)$$

The second step is yet another application of moment-generating functions, while the third step follows from the log-linear approximation.

**Certainty Equivalence under Lognormality**

The log certainty equivalent $\tilde{CE}$ of a risky portfolio is the guaranteed log return that would yield the same expected utility $EU$. It is readily checked that $\tilde{CE} = \frac{1}{1-c} \ln \left( \left| EU \right| \right)$. Applying the
approximations for lognormality discussed above, it follows that for a portfolio $\omega'X$ with log mean $\bar{m}$ and variance $\bar{v}$:

$$\bar{CE} \equiv \bar{m} - \frac{1}{c-1}(c-1)\bar{v} \equiv \omega'\bar{M} - \frac{1}{2}(c-1)\omega'\bar{\Sigma}\omega.$$ 

The factor $\frac{1}{2}(c-1)$ measures the cost of variance in terms of equivalent certain loss. At first glance this suggests that someone with $c<1$ likes risk. For example, when $c=0$, connoting risk neutrality, $\bar{CE}$ equals $\bar{m} + \frac{1}{2}\bar{v}$, so the investor might willingly accept a bet with negative $\bar{m}$. However, the apparent mistake vanishes when we recall that $\bar{m}$ denotes the mean log return rather than the log mean return. The log mean return is indeed $\bar{CE}$ itself, as it should be for risk neutrality.

**Expected CRR Utility under Conditional Lognormality**

For each possible risk regime $k$, calculate a conditional certainty equivalent $\bar{CE}_k$ and convert to a conditional expected utility $EU_k$. Then calculate the overall expected utility $EU$ as the expectation of $EU_k$ given perceived probabilities $p_k$ that regime $k$ occurs:

$$EU = \sum_k p_k EU_k = \text{sign}(1-c) \cdot \sum_k p_k \exp\left( (1-c) \cdot \bar{CE}_k \right)$$

$$\approx \text{sign}(1-c) \cdot \sum_k p_k \exp\left( (1-c) \left( \omega'\bar{M}_k - \frac{c-1}{2} \omega'\bar{\Sigma}_k\omega \right) \right).$$

Devlin’s recipe presents only the formula for the typical case $c>1$.

**Analogy to Thermodynamics**

In his lectures on statistical thermodynamics, Schrödinger defined the partition function

$$Z = \sum_\ell \exp\left( - \frac{\varepsilon_\ell}{kT} \right),$$

where $k$ is the Boltzman constant, $T$ is the temperature, and $\varepsilon_\ell$ is the energy in state $\ell$. He noted that $-kT \ln(Z)$ measures free energy. So if we interpret $1-c$ as $1/kT$, $EU$ as $-Z$, $\bar{CE}_k$ as the conditional energy and $\bar{CE}$ as the aggregate free energy, the equation is analogous. Higher risk aversion corresponds to lower temperature and more muted energies.

To generate the probability measure $p$ over the various states, simply allow each state to occur multiple times with $p_k$ measuring its relative frequency. Thermodynamics uses this procedure too. In fact, each conditional expected utility $EU_k$ is itself the reduced form of a partition function. Variance represents random collisions that dissipate energy and hence reduce $\bar{CE}_k$, with a given collision rate relatively more costly when the risk aversion is high or temperature low.
18. ADJUSTED ADVICE

“That was delicious,” said Conway after lunch. “I always wondered what truffles tasted like.”

“Not me,” said Devlin. “Mom taught me to throw away any food with fungus growing in it. But I did like the double fudge chocolate cake. At least the bit that Regretta left me.”

“Don’t exaggerate, Devlin. I just took a little nibble,” said Regretta, guiltily.

“If you call that a little nibble, you’re a capybara.” Devlin chuckled. Regretta’s face turned red.

Poor Devlin, so artlessly gauche. I better help him out.

“Devlin, this may come as a surprise to you, but not every woman likes to be compared to a giant water rodent. Is it possible you meant a hungry sleek mink instead?”

“All I said was…,” Devlin began to reply, until he saw Conway motion him to zip it. Devlin turned to Regretta and eyed her slowly up and down. “Yes, I suppose I did. Sorry, Regretta.”

“Never mind,” said Regretta, “but thank you both. Let’s get back to work, shall we?”

“I agree,” said Conway. “Come on, Devlin, it’s time to show us the solution.”

“What do you mean? I already gave you the whole recipe.”

“Not quite. It’s missing the last line.”

“Oh, I get it. You want an explicit formula for the optimal portfolio mix.”

“Of course.”

“And you figure I gave you the recipe only to whet your appetite for the simpler answer.”

“That thought had crossed my mind.”

“Then brace yourself. There’s no explicit closed-form formula. Not with multiple risky regimes.”

“You mean we have to solve it numerically case-by-case?”

“I’m afraid so. But it’s not as bad as it sounds. With only a few assets and scenarios, I’ve found that a spreadsheet solver optimizes in a few seconds at most. And exponential polynomials seem sufficiently well-behaved that I suspect even complicated problems can be solved reasonably quickly, though I haven’t tested this with any rigor.”

“Can you draw out any qualitative implications of your model without crunching a lot of numbers?” asked Regretta,

“A few. For example, if the CRR exceeds one, rational gamblers won’t like downward skews or fat tails. They’ll prefer thin-tailed distributions that slant upwards.”

“That’s reassuring. How did you verify that?”
“By taking a fourth-order Taylor series expansion in the certainty equivalents and examining the
signs on skewness and kurtosis.”

“Interesting. Do the higher moments make a lot of difference or just a little?”

“That depends. Remember your chart on the maximal CRR tolerance for risk? I could drive up
the Sharpe ratios sky-high on some of those boom-or-bust bets without the investor taking
them. Conversely, if I dilute the risk enough, any investor will stomach it given a positive
expected return.”

“So much for ranking portfolios by their Sharpe ratios.”

“I wouldn’t say that. They do rank most portfolios correctly, provided you can lever or dilute risk
as needed. In fact, the optimization itself can be interpreted as a modified Sharpe ratio
maximization. The optimal risky portfolio weights equal the inverse of a weighted average
covariance matrix times the corresponding weighted average vector of excess returns, divided
by the CRR less one. The riskfree asset picks up the slack.”

“What are the weights? The probabilities of each regime?”

“No, they’re risk-adjusted probabilities. They’re proportional to the actual probabilities times the
marginal expected utility in that regime. The worse the regime, the higher the expected marginal
utility. So the adjustments make you focus more on improving performance in the worst regimes
than the pure probabilities alone suggest.”

“Doesn’t this give you an explicit formula for the portfolio weights? I thought you said there
wasn’t one.”

“There isn’t. When you modify the portfolio, you modify the risk adjustments. So you would need
to start with an initial guess and iterate. Or rather, have a computer do it quickly.”

Conway spoke up. “Alright, Devlin, I think I understand how your model works and its theoretical
merits. What are its main practical benefits?”

“Confidence, mostly. In most of the examples I’ve played with, the portfolio recommendations
don’t improve much over a single-regime approximation.”

“Oh, really? Don’t they help you deal with uncertainty?”

“Not if the uncertainty is just white noise around the true means. Then the single-regime method
is just as good, provided the unconditional variance incorporates both the risk and the
uncertainty the way you suggested.”

“So what makes the two approaches noticeably differ?”

“Icebergs. To mitigate disasters my model will emphasize keeping low-earning lifeboats
around—T-bills and the like—even at the expense of lower Sharpe ratios. Do you recall the
example you gave in your first visit with small odds of a common 30% crash?”

“And uncorrelated otherwise? Yes, I remember.”
“Well, I took the same parameters as yours—mean returns of 5% and 3% respectively barring crash, volatility of 10% barring crash—and optimized subject to a crash risk of 4.5% and a CRR of 4. Treating the asset pair as bivariate normal, as standard theory requires, advises investing about five times as much in the first asset than the second, with virtually no T-bill holdings. The partition function method boosts the cash holdings to 30%. Moreover, it takes roughly equal bites out of both assets, resulting in the second asset nearly disappearing from the portfolio.” Devlin laid out a chart.

Common Disaster Risks, Different Advice

Unconditional Mean/Variance

CRR Utility Maximization

“That’s a sharp contrast. Only how am I supposed to persuade someone who doesn’t know one model from the other that yours is better?”

Devlin shrugged his shoulders. “Beats me. You know how finance people are. They’re so geared to figuring out the consensus view that they have trouble thinking for themselves. The first question you’ll be asked when you walk in the door is who else uses the system you’re recommending. The second question is why not. You’ll have two strikes against you before you sit down.”

Regretta chimed in. “On top of that, you’re not likely to be judged fairly, comparing the portfolio outcomes with the information you had ex ante. Most often a crisis won’t occur, in which case your recommendations will appear to have fallen short. And if a crisis does occur, people will be inclined to fault you for not hedging even more.”

“I realize that. But isn’t there anything I can say to make the new approach more palatable?” asked Conway. Regretta and Devlin gave him a blank look. “Please, help me out.”

After a pause Devlin spoke. “I can’t identify with the bozos. Maybe that’s why I’m cooped up and they’re running free. But I can tell you what intuitively appeals to me. First, the allocations don’t tend to be as extreme as what standard theory recommends. You don’t need as many ad hoc bounds to generate an interior solution or weight the market consensus as highly. You just need to acknowledge the iceberg risks that most investors and fund managers have in the back of their minds anyway.”

Osband, ICEBERG RISK
“I thought one of the features of iceberg risk is that people don’t know exactly what it is. If the captain of the Titanic had even thought about icebergs he probably would have taken more precautions.”

“I agree. You can never fully expect the unexpected. That limits how much you can forestall. Still, the very exercise of looking for possible disasters tends to expand your awareness. And by tweaking the probability forecasts within the model you can better appreciate what thresholds matter and how. That’s the second great appeal of the model. It helps embed risk management into asset allocation, instead of imposing it as an afterthought.”

“Can you give us an example?”

“Sure. I have to warn you: it’s very stark and stylized. But it does spotlight iceberg risk. Moreover, it’s simple enough to calculate without normal or log-normal approximation.”

“Good. I’m all for simple.”

“Imagine a risky asset has only two possible outcomes: The first outcome falls $s$ standard deviations below the mean...”

“Hey, I remember that one. It’s the distribution you used to show that an $s$-standard deviation tail can have probability $\frac{1}{s^2 + 1}$. The second outcome falls $\frac{1}{s}$ standard deviations above the mean.”

“Exactly. Now suppose an investor’s portfolio consists of this asset and risk-free T-bills. Her only choice is what fraction to hold in each. How according to standard theory should $s$ affect her choice?”

“It shouldn’t. The mean, the standard deviation, and Regretta’s risk aversion will decide everything. Tail risk per se won’t matter.”

“Right. And that’s not very intuitively appealing, is it?”

“We’ve gone over that already. Nearly everyone feels that tail risk matters. Sometimes it matters a lot.”

“Then you’ll like what the partition functions have to say. I worked through some examples with an excess mean return of 5%, a standard deviation of 10%, and a CRR of 3. When $s=1$, so that the tails are short and symmetric, the recommended portfolio weight on the risky asset is nearly 2, meaning that you should borrow to double up. At $s=3$—that is, a 10% chance of a 25% loss with an 8.3% gain otherwise—you don’t lever at all. At $s=10$, that is a 1% chance of losing 95% and 99% chance of gaining 6%, you shouldn’t put more than half of your portfolio at risk.”

“Double up or cut in half: That’s a huge variation. But all we all know that two portfolios can look quite different and still yield comparable risk-adjusted returns. How much difference does strategy choice make here?”

“Well, we know it makes a huge difference to apply advice founded on limited downside to the case of $s=10$. If you lever up a bet that already risks you losing nearly everything, then your expected utility will be minus infinity.”
“In theory, yes. In real life bankruptcy considerations would kick in so that’s not such a clean example.”

“Fair enough. But even when \( s=2 \), wrongly assuming \( s=1 \) can cut your risk-adjusted return nearly in half. And if \( s=4 \), the error can cost you over 15 percentage points in risk-adjusted return. gap and the gap widens exponentially as \( s \) increases. Here’s a chart I prepared to illustrate.”

\[
\text{COSTS OF IGNORING TAIL RISK (}\mu=5\%, \sigma=10\%, \text{ CRR}=3)\]

Conway admired the chart. “Superb. Even your bozos will be impressed. But if I may offer a friendly editing suggestion, why not say ‘standard advice’ instead of ‘advice using \( s=1 \)’? It’s shorter and clearer.”

“It’s shorter but not clearer. Standard advice assumes normality, which we know is technically incompatible with CRR utility. I could convert normality to a log normal approximation but that’s not standard either, and there’s more than one way to do the conversion.”

Regretta was intrigued. “Multiple conversion techniques? What do you mean? Don’t you just want to match the means and variances?”

“Depends what you mean by match. Suppose I apply the approximation favored in options theory, where the log and percentage returns have identical variances and means that differ by half the variance. Suppose I also apply the approximation of my recipe, which treats log returns just like percentages when it comes to calculating portfolio means and variances. The optimal portfolio share for the risky asset then works out to 225%, which exceeds the 199% recommended for \( s=1 \) and will sacrifice even more risk-adjusted return. However, for more accuracy, you might want to drop either approximation or both.”

“What difference does it make?”
“If you drop the first approximation but not the second, the optimal share is 245%. If you drop the second approximation but not the first, it’s 167%. If you drop both approximations, the answer is 211%. In any case the recommended advice leaves you way over-exposed to iceberg risk for any $s$ over 2.5.”

“That’s still a big divergence in the recommendations. What do you think causes it?”

“Two things. To begin with, you’re approximating risk-adjusted returns rather than the optimal portfolio mix. A second-order change in the returns means a first-order change in their slope, and a first-order change in their slope can easily mean a first-order change in the location of the maximum. On top of that, the top risk-adjusted returns tend to be characterized less by sharp peaks than by ridges, so that two substantially different portfolio mixes may both yield nearly optimal results for a given problem.”

“But I take it some near-solutions may be much less robust to misspecification than others.”

“Exactly. The near-solution for normality with a 245% risky asset share is much more vulnerable to icebergs than the near-solution with a 167% share. Lack of robustness is one of the main weaknesses of standard mean-variance optimization.”

“But you can’t prevent misspecification even in your model. ‘Garbage in, garbage out’ is a fundamental law. You can’t repeal it.”

“No, I can’t. But my model does serve to amplify even modest warnings of risk. For example, suppose you’re certain that one of the two-point distributions we’ve been looking at applies and that $s$ equals either 1 or 4, but you’re completely uncertain which. A naïve approach would average the best portfolio for $s=1$ with the best portfolio for $s=4$, yielding a risky asset share of 142%. But my approach would tell you to model your uncertainty as two different, equally likely regimes, one in which $s=1$ and one in which $s=4$. What do you think it will advise?”

“Less than 142%, I presume.”

“A lot less. Expected utility is maximized when the share is just 101%, as if you were certain $s$ were 2.9. Your risk-adjusted return will be 299 basis points, nearly half again as high as the naïve approach would yield.”

Regretta nodded in approval. “Neat. Really neat. I think you’ve won a convert.”

“Make that two,” said Conway. “How about building me a portfolio optimizer that implements your recipe? I have a budget for consultants I can pay you out of.”

“Not so fast. I have to improve the recipe first.”

*You just hate being agreed with, don’t you?* “I thought you were happy with the recipe.”

“Not happy enough. Those approximations to portfolio moments bother me. I’m worried they’re too crude.”

“Is that all? Don’t make better the enemy of good.”
“It’s not always good. For example, for a CRR of less than one my recipe goes berserk and tells you to opt for unbounded risk.”

“Aahhh, so that’s why you restricted it. But according to you and Regretta it’s not a very important case anyway.”

“It’s important that my recipe fails without my knowing why.”

_Such a perfectionist._ “Look, Devlin, why not build the simple version first and tweak it later?”

“Build the simple version yourself if you want. Maybe Regretta will help you. I’d rather think more about the problem.”

“Now that’s a good idea. How about it, Regretta?”

Regretta glanced at Devlin before replying to Conway. “Well, I wouldn’t want to undercut Devlin…”

“I don’t mind,” said Devlin. “Really.”

“OK, then, I’ll do it. But can I ask you one question, Conway?”

“Sure, go ahead.”

“How much are you willing to pay?”…

While Conway and Regretta negotiate terms, let’s review Devlin’s calculations and try to understand his concerns.

_An Approximate Certainty Equivalent_

While the formula synthesizing the conditional certainty equivalents into the aggregate is straightforward, it doesn’t provide much intuition for their interaction. Here’s one way to gain some insight. First, using fourth-order Taylor series expansions for the various

\[
\exp \left( (1-c)\bar{CE}_i \right),
\]

estimate the partition function as:

\[
|EU| \equiv 1 + (1-c)E\left[\bar{CE}_i\right] + \frac{(1-c)^2}{2} E\left[\bar{CE}_i^2\right] + \frac{(1-c)^3}{6} E\left[\bar{CE}_i^3\right] + \frac{(1-c)^4}{24} E\left[\bar{CE}_i^4\right]
\]

which we can express by the shorthand \(1+Q\). Next, use a fourth-order Taylor series expansion

\[
\ln(1+Q) \equiv Q - \frac{Q^2}{2} + \frac{Q^3}{3} - \frac{Q^4}{4}
\]

to approximate \(\frac{1}{1-c} \ln \left(|EU|\right)\). Ignoring terms of fifth order and higher, some tedious algebra yields:

\[
\bar{CE} \equiv \text{Mean} - \frac{c-1}{2} \text{Var} + \frac{(c-1)^2}{6} \text{Var}^{\frac{3}{2}} \text{Skew} - \frac{(c-1)^3}{24} \text{Var}^2 \text{Kurt}
\]
where \(\text{Mean}, \text{Var}, \text{Skew},\) and \(\text{Kurt}\) denote respectively the mean, variance, skewness, and kurtosis of the \(\tilde{CE}_i\). 

Negative skewness (a leftward slant in the distribution) penalizes the aggregate \(\tilde{CE}\) over and above variance, as does positive kurtosis (fat tails). Disaster risk—a small chance of a huge loss—nearly always subtracts from the skewness and adds to the kurtosis. Hence it reduces risk-adjusted returns more than mean-variance analysis suggests. The more risk averse the investor is, the more the higher-order moments matter.

**Optimization Given a Single Regime**

With only a single regime, maximizing expected utility amounts to maximizing a single \(\tilde{CE}\) expression. For a perfectly lognormal asset, as we recall, this would equal the mean log return less \(\frac{1}{2}(c-1)\) times the variance of log returns. For a portfolio of lognormal assets, Devlin’s recipe approximates \(\tilde{CE}\) as \(\omega'\tilde{M} - \frac{1}{2}(c-1)\omega'\tilde{\Sigma}\omega\). Maximizing with respect to \(\omega\) yields optimal risky shares \(\omega^*\) of \(\frac{1}{c-1}\tilde{\Sigma}^{-1}\tilde{M}\).

The second derivative matrix, or Hessian, of \(\tilde{CE}\) is just \((1-c)\tilde{\Sigma}\), which is negative or positive definite depending on the sign of \(c-1\). Therefore, \(\omega^*\) does indeed maximize risk-adjusted returns in the typical case \(c>1\). However, when \(c<1\), \(\omega^*\) yields a minimum rather than a maximum, and investors are advised to seek unbounded risks. We’ll be looking for ways to patch that.

**Relation to Standard Sharpe Ratio Maximization**

Recall that multiples of \(\tilde{M}\) maximize the Sharpe ratio. So Devlin’s model appears to incorporate standard Sharpe ratio maximization as a special case. Unfortunately, along the way we shifted the definition of means and covariances to apply to log returns instead of percentage returns. That makes the correspondence only approximate.

Some of the discrepancy reflects errors in Devlin’s method. For example, while his method treats log portfolio returns as linear in the portfolio weights, a second-order Taylor’s expansion reveals a quadratic element as well:

\[
E\left[\ln(1+\omega'X)\right] \equiv E[\omega'X] - \frac{1}{2} E\left[(\omega'X)^2\right] = \omega'M - \frac{1}{2}\omega'MM'\omega - \frac{1}{2}\omega'\Sigma\omega.
\]

If we substitute this quadratic form for \(\omega'M\) in Devlin’s approximation to \(\tilde{CE}\) while assuming that \(\tilde{\Sigma} \equiv \Sigma\), the optimal mix is revised to

\[
\omega^* = \frac{1}{c+M'S^2M}M = \frac{1}{c+SS^2}M = \frac{1}{c+S^2}\Sigma^{-1}M.
\]

where \(S^*\) denotes the maximal Sharpe ratio and \(S\) the actual Sharpe ratio. The first equality can be verified by pre-multiplying both sides by \(c\Sigma + MM'\). The second equality follows from the
equation for the maximum Sharpe ratio, while the last equality follows from $\omega^*$ maximizing the Sharpe ratio.

In short, the revised optimal mix maximizes the Sharpe ratio without unbounded risk, even for $c < 1$. That’s an important step toward reconciling different measures. Bear in mind, however, that the revision applies a second-order approximation only to the portfolio mean and not the variance.

**Lognormal Approximations via Moment-Matching**

One way to fit a lognormal approximation to returns is to match the first two moments. Recalling the formulas for lognormal moments from the previous chapter, we must choose $\mu$ and $\sigma^2$ to satisfy:

$$\exp(\mu + \frac{1}{2} \sigma^2) - 1 = \mu ; \quad \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1) = \sigma^2$$

where $m$ and $v$ denote the target percentage mean and variance. This has solution:

$$\sigma^2 = \ln \left( 1 + \frac{\sigma^2}{(1+\mu)^2} \right) \equiv \frac{\sigma^2}{(1+\mu)^2}$$

$$\mu = \ln(1+\mu) - \frac{1}{2} \sigma^2 \equiv \mu - \frac{1}{2} \left( \mu^2 + \sigma^2 \right).$$

The same relation applies between portfolio statistics $m$, $v$, $\tilde{m}$ and $\tilde{v}$. We can thus relate the Sharpe ratio $\tilde{S} \equiv \frac{\tilde{m}}{\sqrt{\tilde{v}}}$ based on log returns to the standard Sharpe ratio $S \equiv \frac{m}{\sqrt{v}}$ as:

$$\tilde{S} = m - \frac{1}{2} \left( m^2 + v \right) \equiv (1+m)(S - \frac{1}{2} mS - \frac{1}{2} \sqrt{v}) \equiv S + \frac{1}{2} mS - \frac{1}{2} \sqrt{v} = S + \frac{1}{2} \sqrt{v} \left( S^2 - 1 \right).$$

Hence, the two Sharpe ratios give nearly the same readings around the value of one. Above one the log-based Sharpe rewards more for variance than the standard Sharpe. Below one it penalizes more for variance.

**A Few Examples**

To illustrate the impact of the various approximations mentioned above, let us consider Devlin’s example in which the risky asset has mean $\mu = 5\%$ and standard deviation $\sigma = 10\%$, for a Sharpe of exactly 0.5. For a lognormal density to match that mean and standard deviation, it must set $\{\tilde{\mu} = 4.428\%, \tilde{\sigma} = 9.502\%\}$ for a log-based Sharpe of 0.466. In comparison, the standard lognormal approximation used in finance would set $\{\tilde{\mu} = 4.5\%, \tilde{\sigma} = 10\%\}$ for a log-based Sharpe of 0.450. The approximations $\sigma(1-\mu)$ and $\mu-\frac{1}{2}(\mu^2 + \sigma^2)$ given above are much sounder: they set $\{\tilde{\mu} = 4.424\%, \tilde{\sigma} = 9.500\%\}$ for a log-based Sharpe of 0.466.
Now let’s consider a portfolio that doubles up on the risky asset. Applying the approximations in Devlin’s recipe to fit a lognormal density to the portfolio, we will calculate \( \bar{\mu} = 9\% \), \( \bar{\sigma} = 20\% \) using the standard benchmark estimate versus \( \bar{\mu} = 8.855\% \), \( \bar{\sigma} = 19.005\% \) using the better benchmarks. However, an even better procedure would be to apply the conversion formulas directly to the doubled-up portfolio, that is, to \( m = 10\% \) and \( \nu = 20\% \). This generates \( \bar{\mu} = 7.905\% \), \( \bar{\sigma} = 18.039\% \) for a log-based Sharpe of 0.438.

Devlin’s approximation fares better when the risky asset is diluted. For example, if the risky asset comprises only half the portfolio, Devlin’s recipe would set \( \bar{\mu} = 2.214\% \), \( \bar{\sigma} = 4.750\% \) using the best lognormal approximation, whereas a more direct calculation on the portfolio would set \( \bar{\mu} = 2.350\% \), \( \bar{\sigma} = 4.875\% \). Still, the discrepancies remain a bit too large for comfort and help explain Devlin’s dissatisfaction.

**Optimization with Multiple Regimes**

With multiple possible regimes, the first-order conditions for optimizing Devlin’s approximate formula \( \bar{EU} \) for expected utility become:

\[
\frac{d\bar{EU}}{d\omega} = \sum_k p_k EU_k \left( (1-c) \bar{\mu}_k - (1-c)^2 \bar{\sigma}_k \omega \right) = 0
\]

This implies

\[
(c-1)\omega^* = \left( \sum_k p_k EU_k \bar{\Sigma}_k \right)^{-1} \sum_k p_k EU_k \bar{M}_k
\]

for the optimal portfolio \( \omega^* \). The Hessian is \((c-1)^2 p_k EU_k \bar{\Sigma}_k\), and since the \( EU_k \) all carry the sign of 1-c, this will negative semi-definite and thereby guarantee that \( \omega^* \) maximizes \( \bar{EU} \) if \( c > 1 \). If \( c < 1 \), Devlin’s method will fail as miserably as it does in the single-regime case: \( \omega^* \) will minimize \( \bar{EU} \) and investors will be advised to take unbounded risks.

**Dependence of Optimal Mix on Risk Aversion**

In classic finance theory, the optimal mix is proportional to the inverse of a covariance matrix times a vector of means, while risk aversion has no impact apart from leverage. At first glance, the formula above for \( \omega^* \) preserves the same properties. On closer examination, risk aversion enters the expressions for \( EU_k \) in a way that’s impossible to disentangle. There is no general closed-form solution and the risky bundle’s composition varies with investor risk aversion. Optimization has to be done iteratively to readjust utility weights.

**A Sharpe Ratio Interpretation of the Optimal Mix**

There is a sense however in which a kind of Sharpe ratio maximization still applies. Form a new vector \( \lambda \) with elements \( \lambda_k = \frac{p_k EU_k}{\sum_k p_k EU_k} \). Since the \( \lambda_k \) sum to one, they can be interpreted as defining risk-adjusted probabilities for the various regimes. Provided \( c > 1 \), a “disaster regime” in
which expected utility is highly negative will carry more weight than its simple probability of occurrence suggests. The more investors dislike risk, the more they will weight the worst regime relative to the best regime.

Next form a weighted average excess returns vector \( \bar{\mathbf{M}} = \sum_k \lambda_k \bar{\mathbf{M}}_k \) and an adjusted covariance matrix \( \bar{\mathbf{\Sigma}} = \sum_k \lambda_k \bar{\mathbf{\Sigma}}_k \) using the \( \lambda_k \) as weights. In economic terms \( \bar{\mathbf{M}} \) and \( \bar{\mathbf{\Sigma}} \) represent risk-adjusted means and risk-adjusted covariances respectively. The optimal portfolio then takes a very simple form:

\[
\omega^* = \frac{1}{c-1} \bar{\mathbf{\Sigma}}^{-1} \bar{\mathbf{M}}
\]

with an implied risk-free share of \( 1-\omega^* \mathbf{1} \). The optimal portfolio maximizes a risk-adjusted Sharpe ratio, with the same multiplier as in the single-regime case.

**Incorporating Normally Distributed Uncertainty**

Suppose you are confident that log returns are multivariate normal, with true covariance matrix \( \bar{\mathbf{\Sigma}} \), and that while you are uncertain about the mean log return, your beliefs about the mean can be described as multivariate normally distributed with mean \( \bar{\mathbf{M}} \) and covariance matrix \( \bar{\mathbf{\Sigma}} \). Operationally, this is identical to saying that your beliefs about log returns are multivariate normal with mean \( \bar{\mathbf{M}} \) and covariance matrix \( \bar{\mathbf{\Sigma}} + \bar{\mathbf{\Sigma}} \).

One way to verify this result is to model expected utility in two steps. First calculate expected utility conditional on mean \( \bar{\mathbf{X}} \) and portfolio share \( \omega \) as

\[
\exp\left((1-c)\left(\omega'\bar{\mathbf{X}} - \frac{1}{2}(1-c)\omega'\bar{\mathbf{\Sigma}}\omega\right)\right).
\]

Then integrate the latter over your beliefs about \( \bar{\mathbf{X}} \):

\[
EU = \int_{\mathbb{R}^n} \exp\left((1-c)\left(\omega'\bar{\mathbf{X}} - \frac{1}{2}(1-c)\omega'\bar{\mathbf{\Sigma}}\omega\right)\right) \cdot \left(2\pi \right)^{\frac{n}{2}} \cdot \exp\left(-\frac{1}{2}(\bar{\mathbf{X}} - \bar{\mathbf{M}})' \bar{\mathbf{\Sigma}}^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{M}})\right) d\bar{\mathbf{X}}
\]

\[
= \exp\left(-\frac{1}{2}(1-c)^2\omega'\bar{\mathbf{\Sigma}}\omega\right) \cdot \mathcal{M}_n\left((1-c)\omega\right)
\]

\[
= \exp\left(-\frac{1}{2}(1-c)^2\omega'\bar{\mathbf{\Sigma}}\omega\right) \cdot \exp\left((1-c)\omega'\bar{\mathbf{M}} - \frac{1}{2}(1-c)^2\omega'\bar{\mathbf{\Sigma}}\omega\right)
\]

\[
= \exp\left((1-c)\left(\omega'\bar{\mathbf{M}} - \frac{1}{2}(1-c)\omega'\left(\bar{\mathbf{\Sigma}} + \bar{\mathbf{\Sigma}}\right)\omega\right)\right)
\]

This is just the expected utility given a portfolio \( \omega \) of risky assets having mean \( \bar{\mathbf{M}} \) and covariance matrix \( \bar{\mathbf{\Sigma}} + \bar{\mathbf{\Sigma}} \).

**Moment-Generating Function for Multivariate Normal Density**

In the calculations above, \( \mathbb{R}^n \) denotes the \( n \)-dimensional space of real numbers, while \( \mathcal{M}_n \) is a moment-generating function for an \( n \)-variate normal distribution. Let us verify that \( \mathcal{M}_n(\omega) \) is indeed

\[
\exp\left(\omega'\bar{\mathbf{M}} - \frac{1}{2}\omega'\bar{\mathbf{\Sigma}}\omega\right).
\]

One method is to rearrange the exponents in the integral:
\[
M_a(\omega) = \int \exp(\omega' \tilde{X}) \cdot \left(2^n \pi^n \Xi\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\tilde{X} - \tilde{M})' \tilde{\Xi}^{-1} (\tilde{X} - \tilde{M})\right) d\tilde{X} \\
= \exp(\omega' \tilde{M} - \frac{1}{2} \omega' \tilde{\Xi} \omega) \cdot \int \left(2^n \pi^n \tilde{\Xi}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\tilde{X} - \tilde{M} - \tilde{\Xi} \omega)' \tilde{\Xi}^{-1} (\tilde{X} - \tilde{M} - \tilde{\Xi} \omega)\right) d\tilde{X} \\
= \exp(\omega' \tilde{M} - \frac{1}{2} \omega' \tilde{\Xi} \omega)
\]

The last step follows from the need for probability densities to integrate to one. An even simpler method notes that the integral in question equals the characteristic function value \( M(1) \) for a univariate normal random variable having mean \( \omega' \tilde{M} \) and variance \( \omega' \tilde{\Xi} \omega \).

**Moments of Multivariate Lognormal Variables**

The preceding formula makes it easy to calculate the mean and covariance matrix of multivariate lognormal variables. To calculate the mean \( \mu_i \) of \( x^i = \exp(\tilde{x}_i) - 1 \), set \( \omega_i = 1 \) and zero out all other components of \( \omega \). This verifies that \( \mu_i = \exp(\tilde{\mu}_i + \frac{1}{2} \tilde{\sigma}_{ii}) - 1 \).

To calculate the expectation of \( (1 + x_i)(1 + x_j) \) set \( \omega = \omega = 1 \) and keep the rest zero. Subtract off \( (1 + \mu_i)(1 + \mu_j) \) to form the covariance:

\[
\sigma_{ij} = \exp\left(\tilde{\mu}_i + \tilde{\mu}_j + \frac{1}{2} \left(\tilde{\sigma}_{ii} + 2\tilde{\sigma}_{ij} + \tilde{\sigma}_{jj}\right)\right) - \exp\left(\tilde{\mu}_i + \frac{1}{2} \tilde{\sigma}_{ii}\right) \exp\left(\tilde{\mu}_j + \frac{1}{2} \tilde{\sigma}_{jj}\right) \\
= \exp\left(\tilde{\mu}_i + \frac{1}{2} \tilde{\sigma}_{ii}\right) \exp\left(\tilde{\mu}_j + \frac{1}{2} \tilde{\sigma}_{jj}\right) \left(\exp\left(\tilde{\sigma}_{ij}\right) - 1\right) = (1 + \mu_i)(1 + \mu_j) \left(\exp\left(\tilde{\sigma}_{ij}\right) - 1\right)
\]

When \( i = j \) this reduces to the formula, derived in the preceding appendix, for the variance of a univariate lognormal variable.
19. HIGHER-ORDER HIGH JINKS

I am not fat, thought Regretta the next morning as she looked at herself in the mirror. She turned sideways to take a better look. No, definitely not fat. Just to be sure she skipped breakfast and did a double workout. Afterwards, exhausted and hungry, she ate a chocolate bar and berated herself for letting Devlin get to her.

Realizing she had overreacted, Regretta settled down. She had to admit that she enjoyed Conway’s and Devlin’s company. Strange. Who would have imagined that? Devlin used to be such a jerk.

Suddenly the realization hit Regretta. Devlin must have a crush on her. That’s why he had stopped calling her Misery Girl and apologized so quickly about the capybara joke. Poor Devlin. He must be lonely. But no good can come out of him getting false hopes. I’ll have to put a stop to that.

Monday morning in group therapy Regretta glanced over her shoulder several times to check whether Devlin was looking at her. A few times she caught his eye. He quickly turned away as if it were an accident, but Regretta knew better.

Sure enough, after the session Devlin came up to her. “Hi Regretta, do you have a few minutes? We need to talk.”

“Yes, I think we do. Not here, though.” She didn’t want to embarrass him in front of others.

“Good idea. How about we take a walk in the park?”

Regretta agreed. Outside it was lovely, warm without being hot. They walked across the grounds into a stand of oak trees, sat down on a bench and watched the squirrels play.

Devlin cleared his throat. “Let me get straight to the point, Regretta. I think we need to take things to a higher level.”

Regretta squirmed. “Sorry, Devlin, but I don’t. I think things are fine just the way they are.”

“They are definitely not fine. But we can make them fine. We can do it three ways.”

Regretta was aghast. “Now look here, Buster. I don’t know what goes on in inside your squirrelly brain but I am not that kind of girl. Conway doesn’t strike me as that kind of guy either.”

Devlin screwed up his lips and nodded. “I think you may be right about Conway. But I’m sure you can handle it. Let’s go through it together first and introduce him later.”

“How dare you, Devlin? We hardly know each other.”

“That’s true. There’s a lot of stuff you’re not familiar with. But I can tell you’re a quick learner.”

You arrogant sonofa-! “What makes you so sure you have something to teach me?

“Now don’t get offended, Regretta. I’ve just had more time to work things out on my own.”
“That must be fun.”

“Most of the time, yes. But maybe others will help come up with some new twists.”

“Why don’t you and Conway start twisting on your own?”

Devlin shook his head. “Please Regretta. I know Conway. He’s a great guy, but that’s just not his forte. Maybe we should hide the really kinky stuff from him and just let him focus on results. What do you think?”

“I think you’re gross. I don’t want anything to do with the kinky stuff, either.”

“Oh, Regretta, I am so disappointed in you. Can you not just go through it once? It helps me check that everything works.”

“If you need a checkup, go see a doctor.”

“You must be kidding, Regretta. What do these doctors know about our problems anyway? They’ll just try to switch the conversation to something irrelevant, like my relationship with my mother.”

“It may be more connected than you think.”

“Ha, ha, ha. Very funny, Regretta. Let’s just get down to business, shall we? I have something I want to show you.” He reached down in his pocket.

“No, Devlin, definitely not!” She turned away.

“Relax. It’s not going to bite you.” Devlin pulled out a folded paper from his pocket and unfolded the following chart.
Regretta was stunned. “What the …!” she exclaimed and then bit her tongue. She jumped up, purple-faced.

“Are you OK?”

Regretta was too ashamed to explain. Quick, think of something. “Ouch! I must have sat on a thorn.” She brushed the back of her pants, pretending to search. “Ah, there it is.” She flicked the imaginary thorn away and then slowly straightened out her clothes, keeping her head turned away from Devlin.

“So tell me, what’s this chart supposed to represent?”

“Remember Conway’s two-regime scenario in which two assets have uncorrelated 5% and 3% returns respectively except in crisis when they both drop 30%? And the probability of crisis is 4.5%?”

“Oh yes. The first two bars must represent the portfolio mixes advised by classic mean-variance methods and your recipe. What are these refinements?”

“That’s just what I was talking to you about. They represent three different ways to refine the estimates.”

“Why not just pick one way?”
“Because I’m not sure which is best. They’re all better than my original recipe, though, just as
my recipe is better than ordinary mean-variance.”

“How do you know that?”

“Because they all rest on higher-order Taylor series approximations to valuations in the various
regimes.”

“Which valuations? The mean and variance of log returns?”

“Partly. I mean, yes. That’s what the third approach does. It assumes lognormality.”

“And the other two?”

“They focus on percentage returns.”

“If the other two focus on the same thing, then what’s the difference between them?”

“The first approach takes a fourth-order approximation to conditional expected utility. The
second approach takes a second-order approximation to conditional risk-adjusted returns.”

“Shouldn’t a fourth-order approximation be a lot better than a second-order approximation?”

“Not necessarily. The first approach focuses on outcomes close to the conditional mean of the
distribution, while the second approach integrates over a full multivariate normal distribution.”

“I thought percentage returns can’t be fully multivariate normal.”

“They can’t. I told you I’m not sure which approximation is best. Fortunately, it rarely seems to
make much difference.”

“Why do you say that? In your chart, the third refinement gives a noticeably different mix from
the others.”

“I meant: not a big difference in utility terms. No matter which refinement you think is ‘right’, the
others come close in expected utility terms. Actually my original recipe isn’t bad either. But the
ordinary mean/variance approach lags well behind, as it takes too much risk.”
Regretta studied the chart. “Is this all the refinements add relative to the quick recipe?”

“In this example, yes. In general, no. The bigger the tails, the more useful the refinements are.”

“So when are you going to show me the new recipes?”

Devlin smiled. “I was beginning to think you wouldn’t ask. Here they are.” He pulled another paper out of his pocket.
**Recipe Refinements**

**Refinement #1:** Let $m_k = \omega'M_k$ denote the mean percentage return in regime $k$ and $\hat{\omega}_k = \omega'\Sigma_k\omega(1+m_k)^2$ the variance of the return relative to its gross mean. Calculate $\mathcal{E}U_k$ as:

$$\text{sign}(1-c)\cdot(1+m_k)^{1-c} \left(1 + \frac{c}{2}(c-1)c\hat{\omega}_k + \frac{c}{2}(c-1)c(c+1)(c+2)\hat{\omega}_k^2\right).$$

**Refinement #2:** Using the same notation, calculate $\mathcal{E}U_k$ as:

$$\text{sign}(1-c)\cdot p_k \cdot (1+m_k)^{1-c} \left(1 + (1-c)\hat{\omega}_k\right)^\frac{1}{2} \exp\left(\frac{c}{2}(1-c)^2\hat{\omega}_k\right).$$

**Refinement #3:** Denote by $\tilde{A}_k = \ln\left(1 + \omega'\left(\exp(\tilde{M}_k) - 1\right)\right)$ the log portfolio return when log asset returns match their means, by $\tilde{B}_k = \text{diag}(\omega)\exp(\tilde{M}_k - \tilde{A}_k)\mathbf{1}$ the portfolio weights adjusted for different relative expected log returns, and by $\Psi_k = \text{diag}(\tilde{B}_k) - \tilde{B}_k \tilde{B}_k^\top$ an adjustment for curvature. Calculate $\mathcal{E}U_k$ as:

$$\text{sign}(1-c)\cdot|\Psi_k|^{-1/2} \exp\left((1-c)\left(\tilde{A}_k + \frac{c}{2}(1-c)\tilde{B}_k \Psi_k^{-1}\tilde{\Sigma}_k \tilde{B}_k^\top\right)\right).$$

“Oh, my,” said Regretta. “That’s a lot to swallow.”

Devlin nodded. “I agree. Fortunately, you don’t need to understand them to use them. Just wire the calculations into a spreadsheet. But are you sure you don’t want me to walk you through them once?”

“Well, if you insist…”

While Devlin is explaining things to Regretta, let me try and do the same here.

**A Quick Recapitulation**

Devlin’s initial recipe approximated percentage portfolio returns $1 + \omega'X$ as $\exp(\omega'X)$, reasoning as follows:
\[
1 + \omega' X = \exp(\ln(1 + \omega' X)) \equiv \exp(\omega' X) = \exp(\omega' \left(\exp(\bar{X}) - 1\right)) \equiv \exp(\omega' \bar{X}) .
\]

This turned conditional expected utility into a moment-generating function with a simple closed-form expression. The conditional certainty equivalent was linear in both the mean log return and variance, with a penalty per unit variance of \(\frac{1}{2}(c-1)\).

This chapter offers three ways to further refine the approximation. All rely on higher-order Taylor series expansions. The first approach creates quartic approximations to conditional expected utilities. The other two approaches create quadratic approximations to the conditional certainty equivalents, with one approach focusing on percentage returns and the other on log returns.

**A Few More Terms**

I beg readers’ indulgence to tolerate a few more terms. While I could get by without them, they will make the derivations below much easier to follow. The most useful is \(\hat{\varepsilon}\), which remeasures portfolio outcomes as a percentage deviation from their mean value. That is, given a portfolio mix \(\omega\), \(\hat{\varepsilon} \equiv \frac{\omega' X - m}{1 + m} = \frac{\omega' (X - M)}{1 + \omega' M}\). It is readily checked that \(\hat{\varepsilon}\) has mean 0 and variance \(\frac{\nu}{(1 + m)^2}\), which I will denote by another shorthand \(\hat{\nu}\). Recall from the last appendix that \(\hat{\nu}\) is close to \(\tilde{\nu}\), the log variance parameter in a lognormal portfolio having log mean \(\hat{m}\), percentage mean \(m\), and percentage variance \(\nu\).

Other new terms are \(\tilde{A}\), \(\tilde{B}\) and \(\tilde{C}\). They represent, respectively, log returns, their first derivative or gradient, and their second matrix derivative or Hessian, evaluated at the mean log return \(\bar{M}\).

**REFINEMENT #1**

**Approximating CRR Utility Using Percentage Returns**

The first refinement starts with a Taylor series expansion of percentage portfolio returns \(\omega' Y\) around their conditional mean \(m_k = \omega' M_k\) in regime \(k\):

\[
(1 + \omega' X)^{1-c} = \sum_{j=0}^{\infty} \binom{1-c}{j} (1 + \omega' M_k)^{1-j} (\omega' X - \omega' M_k)^j
\]

\[
= (1 + \omega' M_k)^{1-c} \sum_{j=0}^{\infty} \binom{1-c}{j} \left(\frac{\omega' X - \omega' M_k}{1 + \omega' M_k}\right)^j \equiv (1 + m_k)^{1-c} \sum_{j=0}^{\infty} \binom{1-c}{j} (\hat{\varepsilon}_k)^j
\]

where \(\binom{n}{j} \equiv \frac{n(n-1)\cdots(n-j+1)}{j!}\) even when \(n\) is not a positive integer. Taking expectations over the various possible regimes establishes that:
\[ EU = \text{sign}(1-c) \cdot \sum_k p_k (1+m_k)^{1-c} \sum_{j=0}^{\infty} \left( \frac{1-c}{j} \right) E_k \left[ (\hat{\epsilon}_k)^j \right]. \]

**Caveat about Feasibility**

The preceding summations need not necessarily converge, since \( EU \) may not be well-defined for the distribution in question. For example, none of the \( \hat{\epsilon}_k \) can be fully normal, even conditionally, since CRR utility can’t handle a loss of more than 100%. Let’s carry out the Taylor expansion anyway and see what happens. Clearly, the odd moments of \( \hat{\epsilon}_k \) will be zero due to symmetry. Each even \( 2^j \) moment works out to \( (\hat{v}_k)^j \) times all the positive integers less than \( 2^j \); this can be checked by evaluating the central derivatives of the moment-generating function \( \mathcal{M}(n) = \exp\left(\frac{1}{2} \hat{v}_k n^2\right) \). Expected utility then reduces to:

\[ EU = \text{sign}(1-c) \cdot \sum_k p_k (1+m_k)^{1-c} \sum_{j=0}^{\infty} \frac{(c-1)c \cdots (c+2j-3)(c+2j-2)}{2 \cdot 4 \cdots (2j-2) \cdot 2j} (\hat{v}_k)^j. \]

Now look at the ratio of neighboring terms in this expansion. It equals \( \frac{(c+2j-3)(c+2j-2)}{2j} \) or roughly \( 2j\hat{v}_k \). No matter how small this ratio starts out, it will eventually exceed one, making the summation infinite.

**Expected Utility of Quasi-Normal Portfolios**

I will call a portfolio quasi-normal if both its skewness and kurtosis are close to zero. In that case \( E[\hat{\epsilon}^3] \equiv 0 \) and \( E[\hat{\epsilon}^4] \equiv 3\hat{\nu}^2 \). Note that a portfolio can be quasi-normal even if its constituent assets are decidedly non-normal. Indeed, if a portfolio is composed of enough independent, roughly equally-weighted assets, the Central Limit Theorem indicates it must be quasi-normal.

Now our model doesn’t assume quasi-normality for the portfolio as a whole. Far from it. But we can try to carve up the world into conditionally quasi-normal regimes. Evaluating the Taylor expansion out to the fourth order then indicates that:

\[ EU \equiv \text{sign}(1-c) \cdot \sum_k p_k (1+m_k)^{1-c} \left( 1 + \left( \frac{1-c}{2} \right) \hat{v}_k + \left( \frac{1-c}{4} \right) 3\hat{\nu}_k^2 \right) \]

\[ = \text{sign}(1-c) \cdot \sum_i p_i (1+m_k)^{1-c} \left( 1 + \frac{1}{2} (c-1) \hat{v}_k + (c-1)c(c+1)(c+2)\hat{\nu}_k^2 \right) \]

where again \( m_k \equiv \omega' M_k \) and \( \hat{v}_k \equiv \frac{\omega' \sum \omega}{(1+\omega' M_k)^2} \).
Sharpe Ratio Maximization as Second-Order Approximation

Suppose that all the means and variances are tiny or that we’re just too lazy to include any terms higher than second-order; indeed that we even use a second-order approximation to \((1+m_k)\frac{1}{1-c}\). Then the expression for \(EU\) will dramatically simplify, and we can simplify even further by expressing in terms of \(EU^\# \equiv \frac{|EU| - 1}{1-c}\), which implies exactly the same preferences:

\[
EU^\# = \sum_k p_k \left( m_k - \frac{1}{2} c m_k^2 \right) - \sum_k p_k \left( \frac{1}{2} c \hat{v}_k^2 \right) = m - \frac{1}{2} c \left( m^2 + \text{Var}[m_k] + E[\hat{v}_k] \right) \\
= m - \frac{1}{2} c \left( m^2 + \sigma^2 \right) \equiv \omega' M - \frac{1}{2} c \left( (\omega' M)^2 + \omega' \Sigma \omega \right) = \omega' M - \frac{1}{2} c \omega'(MM' + \Sigma) \omega
\]

where the variables without subscripts are aggregate, unconditional means and variances. The expected utility approximation will be maximized at:

\[
\omega^* = \frac{1}{c} \left( \Sigma + MM' \right)^{-1} M = \frac{1}{c} \left( \frac{1}{1+M'\Sigma^{-1}M} \right) \Sigma^{-1} M = \frac{1}{c} \left( \frac{1}{1+S^2} \right) \Sigma^{-1} M
\]

where \(S\) denotes the Sharpe ratio. This recalls an approximation derived in the previous chapter except for the multiplier of \(c\) on the \(MM'\) term. To verify the second equality, pre-multiply both sides by \(\Sigma + MM'\). The solution, which is valid for all \(c\), always achieves the maximum Sharpe ratio \(\left( M'\Sigma^{-1}M \right)^{1/c}\), implying the last equality. The only difference from previous Sharpe-based solutions is that this formula recommends less leverage.

Contributions of Higher Orders

The more risk averse the investor is, the more important it is to include higher-order terms. For the plausible case \(c=3\), the fourth-order expression for expected utility reduces to:

\[
EU_k = \frac{1}{(1+m_k)^2} \left[ 1 + 3 \hat{v}_k + 15 \hat{v}_k^2 \right] \\
\]

The higher-order terms in this expression don’t matter much for \(|m|\) less than 10% or \(\hat{v}\) less than 20%. For \(\hat{v}\) above 35%, this expression probably isn’t high-order enough, but such high downside risk makes quasi-normality suspect anyway.

Incentives to Mitigate High Risks

For the general quasi-normal case, the risk-adjusted return for each regime works out to:

\[
CE_k = \left[ EU_k \right]^{1/c} - 1 \equiv \mu_k + \left( 1 + \frac{1}{2} (c-1)c \hat{v}_k + \frac{1}{2} (c-1)c(c+1)(c+1) \hat{v}_k^2 \right)^{1/c} \\
\equiv m_k + \frac{1}{2} c \hat{v}_k^2 - \frac{1}{2} c(3c+2) \hat{v}_k^2
\]

Osband, ICEBERG RISK
where the last step follows from a Taylor expansion after dropping all terms in \( \hat{v}_k^2 \) and higher. The negative sign on \( \hat{v}_k^2 \) imposes an extra penalty on high variance over and above the standard mean/variance formulation. Moreover, because \( \hat{v}_k^2 \) itself measures variance relative to the conditionally expected return, variance in bad regimes get penalized more than variance in good regimes.

**REFINEMENT #2**

**A Different Approximation to CRR Utility**

Refinement #2 returns more to Devlin’s original recipe. It rewrites \((1 + \omega'X)^{1-c}\) as \(\exp\left((1-c) \cdot \ln\left(1 + \omega'X\right)\right)\), approximates the logarithm with a more tractable expression, and then draws on moment-generating functions to evaluate the expectation. However, the second approach does not approximate \(\ln\left(1 + \omega'X\right)\) as \(\omega'X\). Instead, it takes a second-order Taylor series expansion around the mean:

\[
\ln\left(1 + \omega'X\right) \approx \ln(1 + \omega'M) + \frac{\omega'X - \omega'M}{1 + \omega'M} - \frac{1}{2} \left(\frac{\omega'X - \omega'M}{1 + \omega'M}\right)^2 \equiv \ln(1 + m) + \hat{e} - \frac{1}{2} \hat{e}^2.
\]

**Evaluating Expected Utility**

To evaluate the expectation, the second approach assumes conditional normality of percentage returns \(\omega'X\) rather than just conditional quasi-normality. It follows that:

\[
sign(1-c) \cdot EU_k = \int_{-\infty}^{\infty} \exp\left((1-c)\left(\ln(1+m_k) + \hat{e}_k - \frac{1}{2} \hat{e}_k^2\right)\right) \cdot (2\pi \hat{v}_k)^{-\frac{1}{2}} \cdot \exp\left(-\frac{\hat{e}_k^2}{\hat{v}_k}\right) d\hat{e}_k
\]

\[
= (1+m_k)^{1-c} \int_{-\infty}^{\infty} \exp\left((1-c)\hat{e}_k\right) \cdot (2\pi \hat{v}_k)^{-\frac{1}{2}} \cdot \exp\left(-\frac{\hat{e}_k^2}{H}\right) d\hat{e}_k
\]

\[
= \frac{H^{\frac{1}{2}}}{\hat{v}_i} (1+m_k)^{1-c} \int_{-\infty}^{\infty} \exp\left((1-c)\hat{e}_k\right) \cdot (2\pi H)^{-\frac{1}{2}} \cdot \exp\left(-\frac{\hat{e}_k^2}{H}\right) d\hat{e}_k
\]

\[
= \frac{H^{\frac{1}{2}}}{\hat{v}_i} (1+m_k)^{1-c} \exp\left(\frac{1}{2} (1-c)^2 H\right)
\]

The third line is just a rearrangement of terms, while the fourth evaluates a moment-generating function. The only step in question is the second, which requires that \(\frac{1}{H} = 1-c + \frac{1}{\hat{v}_k}\) or

\[
H = \frac{\hat{v}_k}{1 + (1-c)\hat{v}_k}.
\]

It follows that:
A Caveat
The integral is not well-defined if \( H \) is negative, as happens when \( c > 1 \) and \( \hat{v}_k > \frac{1}{c-1} \). However, unless the investor is extremely risk averse, this limit falls well beyond the bounds where a normal approximation makes sense. For example, for \( c = 3 \) the standard deviation must exceed 70% of the conditional gross return, implying significant odds of losing more than all your wealth.

Risk-Adjusted Returns
In each regime, the log risk-adjusted return works out to:

\[
\overline{CE}_k = \ln (1 + m_k) + \frac{1}{2} \left( \frac{1-c}{1+(1-c)\hat{v}_k} \right) \ln \left( 1 + (1-c)\hat{v}_k \right) \cdot \left( \frac{1}{c-1} \right).
\]

These results look very familiar. If we approximate \( \ln(1+m_k) \) by \( \tilde{m}_k + \frac{1}{2} \hat{v}_k \) and \( \hat{v}_k \) by \( \hat{v}_k \), and drop higher-order terms, we generate Devlin’s \( \tilde{m}_k + \frac{1}{2} (1-c)\hat{v}_k \) recipe again. Moreover, both this formula and the \( CE \) formula from Refinement #1 include a term that’s quadratic in both \( \hat{v}_k \) and \( c \). Indeed, when \( CE \) is converted into \( \overline{CE}_i \equiv \ln(1 + CE) \) terms, it yields the same approximation as above.

REFINEMENT #3

CRR Utility in Terms of Log Returns
Refinement #3 comes closest to Devlin’s original recipe. Like Refinement #2, it rewrites \( (1 + \omega'X)^{-c} \) as \( \exp\left( (1-c) \cdot \ln (1 + \omega'X) \right) \) and takes a second-order Taylor approximation to the logarithm around its mean. However, it carries out the approximation in terms of log returns rather than percentage returns:

\[
\ln (1 + \omega'X) \equiv \ln \left( 1 + \omega' \left( \exp \left( \tilde{X} \right) - 1 \right) \right) \equiv \tilde{A} + \tilde{B}'(\tilde{X} - \tilde{M}) + \frac{1}{2}(\tilde{X} - \tilde{M})'\tilde{C}(\tilde{X} - \tilde{M})
\]

Calculating \( \tilde{A} \)
To determine \( \tilde{A} \), evaluate log portfolio returns at \( \tilde{X} = \tilde{M} \):

\[
EU \equiv \text{sign}(1-c) \cdot \sum_k p_k \left( 1 + m_k \right)^{-c} \left( 1 + (1-c)\hat{v}_k \right)^{-\frac{1}{2}} \exp \left( \frac{1}{2} \frac{(1-c)^2\hat{v}_k}{1+(1-c)\hat{v}_k} \right).
\]
\[
\widetilde{A} = \ln \left(1 + \omega^\prime \left(\exp \left(\bar{M} \right) - 1\right)\right) \neq \ln \left(1 + \omega^\prime \bar{M} \right) = \ln \left(1 + m \right)
\]

In words, \(\widetilde{A}\) equals the log portfolio returns evaluated at the mean log return on assets, not at the mean percentage return.

**Calculating \(\tilde{B}\)**

To determine \(\tilde{B}\), differentiate log portfolio returns with respect to log asset returns and then evaluate at \(\bar{X} = \bar{M}\). Let’s be careful and do this one element at a time:

\[
\tilde{b}_i = \frac{d}{d\bar{x}_i} \ln \left(1 + \omega^\prime \left(\exp \left(\bar{X} \right) - 1\right)\right) \bigg|_{\bar{X} = \bar{M}} = \frac{\omega_i \exp \left(\bar{x}_i \right)}{1 + \omega^\prime \left(\exp \left(\bar{X} \right) - 1\right)} \bigg|_{\bar{X} = \bar{M}} = \frac{\omega_i \exp \left(\bar{M} \right)}{1 + \omega^\prime \left(\exp \left(\bar{M} \right) - 1\right)} = \omega_i \exp \left(\bar{M} - \tilde{A} \right)
\]

where the subscript \(i\) refers to vector components rather than regimes. Hence, \(\tilde{b}_i\) represents the asset’s portfolio weight if all assets achieve their expected log return over the period. To express the aggregate result in matrix form, I make use of \(\text{diag}(\omega)\), the diagonal matrix of \(\omega^r\):

\[
\tilde{B} = \text{diag}(\omega)^\prime \exp \left(\bar{M} - \tilde{A} \mathbf{1} \right).
\]

**Calculating \(\tilde{C}\)**

To determine \(\tilde{C}\), differentiate log portfolio returns twice with respect to log asset returns and then evaluate at \(\bar{X} = \bar{M}\). Again let’s do this an element at a time. Provided \(j \neq i\),

\[
\tilde{c}_{ij} = \frac{d}{d\bar{x}_j} \left( \frac{\omega_i \exp \left(\bar{x}_i \right)}{1 + \omega^\prime \left(\exp \left(\bar{X} \right) - 1\right)} \right) \bigg|_{\bar{X} = \bar{M}} = -\frac{\omega_i \exp \left(\bar{x}_i \right) \cdot \omega_j \exp \left(\bar{x}_j \right)}{\left(1 + \omega^\prime \left(\exp \left(\bar{X} \right) - 1\right)\right)^2} \bigg|_{\bar{X} = \bar{M}} = -\frac{\omega_i \exp \left(\bar{M} \right) \cdot \omega_j \exp \left(\bar{M} \right)}{\left(1 + \omega^\prime \left(\exp \left(\bar{M} \right) - 1\right)\right)^2} = -\omega_i \exp \left(\bar{M} - \tilde{A} \right) \cdot \omega_j \exp \left(\bar{M} - \tilde{A} \right)
\]

In addition,
Reconstituting the matrix \( \tilde{C} \), it becomes apparent that:

\[
\tilde{C} = \text{diag}(\tilde{B}) - \tilde{B}\tilde{B}' .
\]

\( \tilde{C} \) adjusts the curvature to fine-tune the optimization. Basically it says that you should apply the adjusted weights \( \tilde{b}_i \) not to the deviations \( \tilde{x}_i - \tilde{\mu}_i \) alone but also to half of \( \tilde{x}_i - \tilde{\mu}_i \) squared, and then subtract half the square of the portfolio deviation as a whole.

**Expected Utility**

Once we’ve done our \( \tilde{A}\tilde{B}\tilde{C} \)'s, the rest of the EU calculations are exact assuming lognormality. Denoting \( \tilde{X} - \tilde{\mu} \) by the shorthand \( \tilde{X} \), and recalling that parallel lines around a matrix denote its determinant, calculate:

\[
EU_k = \text{sign}(1-c) \cdot \int_{-\infty}^{\infty} \exp\left((1-c)\left(\tilde{A}_k + \tilde{B}_k' \tilde{X} + \frac{1}{2} \tilde{X}' \tilde{C}_k \tilde{X}\right)\right) \cdot \frac{1}{\sqrt{2^n \pi^n \tilde{\Sigma}_k}} \cdot \exp\left(-\frac{1}{2} \tilde{X}' \tilde{\Sigma}_k^{-1} \tilde{X}\right) d\tilde{X}
\]

\[
= \text{sign}(1-c) \cdot \exp\left((1-c)\tilde{A}_k\right) \cdot \frac{1}{\sqrt{\tilde{\Sigma}_k \tilde{\Sigma}_k^{-1} + (c-1)\tilde{C}_k}} \cdot \int_{-\infty}^{\infty} \exp\left((1-c)\tilde{B}_k' \tilde{X}\right) \cdot \exp\left(-\frac{1}{2} \tilde{X}' \left(\tilde{\Sigma}_k^{-1} + (c-1)\tilde{C}_k\right) \tilde{X}\right) d\tilde{X}
\]

\[
= \text{sign}(1-c) \cdot \exp\left((1-c)\tilde{A}_k\right) \cdot \frac{\exp\left((1-c)\tilde{B}_k\right)}{\sqrt{\det(\tilde{C}_k \tilde{\Sigma}_k)}} \cdot \exp\left(\frac{1}{2} (1-c)^2 \tilde{B}_k' \left(\tilde{\Sigma}_k^{-1} + (c-1)\tilde{C}_k\right)^{-1} \tilde{B}_k\right)
\]

\[
= \text{sign}(1-c) \cdot |\Psi_k|^{1/2} \exp\left((1-c)\left(\tilde{A}_k + \frac{1}{2} (1-c) \tilde{B}_k' \tilde{\Sigma}_k^{-1} \tilde{C}_k \tilde{B}_k\right)\right)
\]

for \( \Psi_k \equiv \det(\tilde{C}_k \tilde{\Sigma}_k) = I^{(c-1)}\tilde{C}_k \tilde{\Sigma}_k \).

**Risk-Adjusted Returns**

The conditional log risk-adjusted returns work out to:
\[
\tilde{C} E_k = \tilde{A}_k + \frac{1}{2}(1-c)\tilde{B}_k \Psi^{-1}_k \tilde{\Sigma}_k \tilde{B}_k + \frac{1}{2} \frac{\ln |\Psi_k|}{c-1} .
\]

Here \( \tilde{A}_k \) serves as an average log return, \( \tilde{B}_k \) as the adjusted portfolio weights (which unlike \( \omega \) vary by regime), and \( \Psi^{-1}_k \) adjusts for curvature. For a bit more insight we can rewrite \( \Psi^{-1}_k \) as a matrix series expansion:

\[
\Psi^{-1}_k \equiv \left( I - (1-c) \tilde{C}_k \tilde{\Sigma}_k \right)^{-1} = I + (1-c) \tilde{C}_k \tilde{\Sigma}_k + \left( (1-c) \tilde{C}_k \tilde{\Sigma}_k \right)^2 + \cdots
\]

This implies that the effective portfolio covariance, dropping the subscript \( k \) for convenience, equals:

\[
\tilde{B}^T \Psi^{-1} \tilde{\Sigma} \tilde{B} = \tilde{B}^T \tilde{\Sigma} \tilde{B} + (1-c) \tilde{B}^T \tilde{C} \tilde{\Sigma}^2 \tilde{B} + (1-c)^2 \tilde{B}^T \tilde{C}^2 \tilde{\Sigma}^3 \tilde{B} + \cdots
\]

The first term in the expansion is just the covariance of a portfolio with weight \( \tilde{B} \) and covariances \( \tilde{\Sigma} \), but I don’t have much intuition for the other terms or for the \( \ln |\Psi_k| \) adjustment above. Fortunately, we can use the recipe without totally understanding it.

**Switching Between Lognormality and Normality**

To avoid \( \tilde{C} \) calculations, switch from multivariate lognormal to multivariate normal specifications and apply refinements #1 or #2. To make the switch, apply the previously derived formulas:

\[
\mu_i = \exp\left( \bar{\mu}_i + \frac{1}{2} \bar{\sigma}_{ii} \right) - 1
\]

\[
\sigma_{ij} = (1 + \mu_i) (1 + \mu_j) \left( \exp\left( \bar{\sigma}_{ij} \right) - 1 \right).
\]

If for some reason you want to convert from normal to lognormal specifications, calculate:

\[
\bar{\sigma}_{ij} = \ln \left( 1 + \frac{\sigma_{ij}}{(1 + \mu_i) (1 + \mu_j)} \right)
\]

\[
\bar{\mu}_i = \ln \left( 1 + \mu_i \right) - \frac{1}{2} \bar{\sigma}_{ii}.
\]