Optimal Hedging via Large Deviations

Michael Stutzer
Professor of Finance, University of Colorado, Boulder, CO 80302
michael.stutzer@colorado.edu

ABSTRACT

The criterion of minimizing the cumulative hedged returns’ probability of underperforming a benchmark provides a framework for evaluating short-term hedges that are rolled-over to produce longer term hedges. Large deviations theory can be used to either parametrically or nonparametrically estimate underperformance probabilities for cumulative hedged returns produced by roll-overs, providing a straightforward way to find optimal hedge ratios. Optimal hedges using soybean futures are constructed to illustrate the procedures, and their relationship to the popular hedging criteria that are motivated by normality.
1 Introduction

Futures contracts are commonly used to reduce the importance of the stochastic component inherent in risky assets. Every textbook presentation of futures contracts contains an exposition of their use in hedging e.g. Stoll and Whaley [18, pp.52-62]. In practice, it is usually impossible to find a futures position that will completely eliminate the stochastic component of returns, in which case there will be some randomness in the so-called “cross” hedged position’s return. The most widely used performance criterion of cross hedge optimality is to minimize the variance of the hedged position’s return (e.g. see Stoll and Whaley (op.cit.). Alternatively, Howard and D’Antonio [8] derive the futures position that maximizes the Sharpe Ratio (see section 2) of the portfolio. The Sharpe Ratio depends both the mean and variance of the hedged position’s return. The focus on mean and variance is usually rationalized by maintaining an assumption that the hedged returns are normally distributed.

But will hedged returns be normally distributed? In many cases short-term hedges are rolled-over to produce longer-term hedges. The longer-term hedged return is a multiplicative cumulation of the separate rolled-over hedged returns, and will likely not be normally distributed, having excess skewness and fatter-than-normal tails (i.e. positive excess kurtosis), depending on the nature of the return generating process.

So in section 3, we develop an alternative performance index for tractable minimization of the probability that the hedged position’s cumulative return will fall short of some benchmark’s cumulative return as time passes. The benchmark could be a target rate of
return, or could be the stochastic return associated with a different investment or index. The formalism will employ a heuristic from the statistical theory of large deviations, that requires application of a scalar version of the Gärtner-Ellis Large Deviations Theorem (see Bucklew [3]). Similarly motivated portfolio choice applications were developed in Stutzer [19] and generalized to continuous time in Pham [14]. Duffy, et al. [4] developed a related application. See Sornette [17] for an early investment application. The method has not been widely used by researchers in economics departments (for an exception, see Williams [20]). But it has been of growing interest among econophysicists (an early survey of the (now) vast field is found in Mantegna and Stanley [13]). Section 4 provides an empirical application of the model. Section 5 concludes.

2 Criteria Based on Mean and Variance

Following Stoll and Whaley (op.cit.), at the beginning of a data period $t$ a hedger has position of $n_s$ units in an asset costing $S_t$ per unit, and adopts a position of $n_f$ units in futures costing zero per unit. At the end of the period, the asset price is $s_{t+1}$ while the futures price is denoted $f_{t+1}$. Hence the profit or loss on the combined asset/futures portfolio (hereafter called the hedged position) is

$$\pi = n_s(s_{t+1} - s_t) + n_f(f_{t+1} - f_t)$$

Because it is assumed that establishment of the futures position is costless, the gross rate of return (i.e. one plus the net rate of return) of the hedged position over the data period is:
\[ R_{ht} = R_{st} + \frac{n_f}{n_s} \frac{f_{t+1} - f_t}{s_t} \equiv R_{st} + hR_{ft} \] 

where \( R_{st} \) is the asset’s gross rate of return over the data period, \( h \equiv \frac{n_f}{n_s} \) denotes the hedge ratio which is varied only by choosing \( n_f \), and \( R_{ft} \) denotes the (random) change in futures price over the period per dollar of the (known) beginning asset price.\(^2\)

The variance minimizing hedge ratio is defined as

\[ h_{\text{var}} \equiv \arg \min_h \left( \text{Var}[R_h] = \text{Var}[R_s] + h^2 \text{Var}[R_f] + 2h \text{Cov}[R_s, R_f] \right) \]

Set the first derivative equal to zero find

\[ h_{\text{var}} = -\frac{\text{Cov}[R_s, R_f]}{\text{Var}[R_f]} \] \hfill (3)

Typically, the covariance term is positive, so the hedge sells futures contracts to offset fluctuations in spot returns.

Howard and D’Antonio (op.cit.) studied the hedge ratio that maximizes the Sharpe Ratio, i.e.

\[ h_{\text{Sharpe}} \equiv \arg \max_h \frac{E[R_h] - r}{\sqrt{\text{Var}[R_h]}} \] \hfill (4)

where \( r \) denotes the gross return on a riskfree asset.
Note that if the hedged return is normally distributed, the probability that the hedged return is less than the riskfree return is

\[
\text{Prob}[R_h < r] = \text{Prob} \left[ Z < \frac{r - E[R_h]}{\sqrt{\text{Var}[R_h]}} \right] = \text{Prob} \left[ Z > \frac{E[R_h] - r}{\sqrt{\text{Var}[R_h]}} \right]
\]

(5)

where \( Z \) is the standard normal random variable. Note the Sharpe Ratio on the right hand side of (5), so maximizing it minimizes the probability of underperforming the riskfree return.

But hedgers may be more interested in measuring performance relative to a benchmark other than a riskfree asset. In light of this, analysts may utilize the ratio of mean to standard deviation of the hedged position’s return in excess of the benchmark’s – a criterion known as the Information Ratio (IR) [7]. If the return in excess of the benchmark is normally distributed, the calculation (5) shows that maximizing the IR minimizes the probability of underperforming the benchmark.

### 3 General Underperformance Probability Minimization

A corporation engaged in hedging may choose futures contracts to avoid losing money on the spot commodity or asset when that is the company’s core business, suggesting a benchmark of zero net cumulative return. When it hasn’t been predetermined how long the hedge will be in place, desirable hedge ratios are those that make the probability of underperforming the benchmark decay exponentially to zero as time passes. Amongst those hedge ratios, we
define the optimal hedge ratio to be the one maximizing the underperformance probability’s rate of decay.

To formalize this, let cumulative wealth at some future time $T$, arising from initial wealth $W_0$ invested in a hedged position, be denoted $W^h_T = W_0 \prod_{t=1}^{T} R_{ht}$. Note that the validity of this expression does not depend on the length of the time interval between $t − 1$ and $t$, nor the particular times $t$ at which the random gross returns are measured. Similarly, a benchmark investment of $W_0$ that generates returns $R_{bt}$ yields $W^b_T = W_0 \prod_{t=1}^{T} R_{bt}$. The benchmark could be a fixed target rate of return, e.g. set the gross benchmark return $R_{bt} \equiv 1$ for a loss avoidance benchmark, or one could set it to some other constant. Alternatively, $R_{bt}$ could be the random gross return process from something else, e.g. a spot commodity index. Taking logs, subtracting, and then dividing by $T$ yields the following expression for the difference in respective stochastic growth rates of wealth:

$$
\log W^h_T / T - \log W^b_T / T = \frac{1}{T} \log \frac{W^h_T}{W^b_T} = \frac{1}{T} \sum_{t=1}^{T} (\log R_{ht} - \log R_{bt}).
$$

(6)

The underperformance probability associated with the hedge ratio $h$ at time $T$ is thus equivalently expressed by:

$$
Prob \left[ \frac{1}{T} \sum_{t=1}^{T} (\log R_{ht} - \log R_{bt}) \leq 0 \right].
$$

(7)

and we seek to rank the performance of hedge ratios inversely to (7). Because the hedger may not choose to precommit to a fixed value of $T$, we are interested in a ranking valid for suitably large $T$. Fortunately, a scalar version of the general Gärtner-Ellis Large Deviations
Theorem, proven in Bucklew [3, pp.14-20], is tailor-made for this purpose. Following Bucklew (op.cit.), define

\[ Y_T(h) \equiv \sum_{t=1}^{T} (\log R_{ht} - \log R_{bt}) \]  

and define

\[ \phi(\gamma_h) = \lim_{T \to \infty} \frac{1}{T} \log E \left[ e^{\gamma_h Y_T(h)} \right] \]  

Assumption 1  Following Bucklew (op.cit.), we restrict attention to processes for \( Y_T(h) \) for which (i) \( \phi(\gamma_h) \) exists as an extended real number (i.e. possibly \( \infty \)) and (ii) \( \phi(\gamma_h) \) is differentiable on the set of \( \gamma_h \) for which \( \phi(\gamma_h) < \infty \).

Then, direct application of the theorem shows that the exponential decay rate of the underperformance probability (7) is

\[ D_h \equiv \inf_{x \leq 0} \sup_{\gamma_h} (\gamma_h x - \phi(\gamma_h)) \]  

The optimal \( h^* \equiv \arg \max_h D_h \).

3.1 The IID Case

Suppose that the process generating \( \log R_{ht} - \log R_{bt} \) is identically and independently distributed (IID). Then (9) simplifies to

\[ \phi(\gamma_h)^{IID} \equiv \log E \left[ e^{\gamma_h (\log R_h - \log R_b)} \right] \]  

7
which is the cumulant generating function (CGF) of the distribution for $\log R_h - \log R_b$. For example, suppose $\log R_h - \log R_b$ is normally distributed with mean $c_h > 0$ and variance $\sigma_h^2$. This lognormal process underlies the traditional random walk and Black-Scholes option pricing models. Substitute the normal distribution’s CGF into (11) to yield:

$$
\phi(\gamma_h) = c_h \gamma_h + \frac{1}{2} \sigma_h^2 \gamma_h^2
$$

which clearly satisfies Assumption 1. Substitute (12) into (10), and perform the inner maximization over $\gamma_h$ to find:

$$
\gamma_h = \frac{x - c_h}{\sigma_h^2}
$$

Substituting (12) and (13) into (10) yields $D_h = \inf_{x \leq 0} \frac{1}{2} \left( \frac{x - c_h}{\sigma_h^2} \right)^2$. The infimum is achieved at $x = 0$, so the optimizing $\gamma_h < 0$ and the resulting underperformance probability decay rate is

$$
D_h^{\text{Lognormal}} = \frac{1}{2} \left( \frac{c_h}{\sigma_h} \right)^2.
$$

The performance index (14) is half the squared Information Ratio computed using log gross hedged returns instead of ordinary arithmetic net returns. We dub this variant the Log Modified Information Ratio (LMIR). Hence in the lognormal case, the performance index is indeed analogous to the ranking of positive Information Ratios that is oft-advocated in single period analyses. The need for log gross returns arises from our appropriate emphasis on measuring performance in terms of the cumulative return as time passes.
3.1.1 A More Flexible IID Process: Incorporating Skewness and Kurtosis

It is instructive to analyze a flexible parametric specification that facilitates study of the effects that the first four moments have on the optimal hedge ratio. We adopt the Variance Gamma process specification of Madan, et.al. [11] and Seneta [16] that performs well in fitting historical asset returns. This process was also used in Duffy, et al. (op.cit.).

The model has only four parameters, and they have elegantly simple relationships to the mean, variance, skewness and kurtosis of the log gross excess returns. The simplicity of those relationships facilitates analysis of the effects the moments have in determining the optimal hedge ratio, and also facilitates a simple method-of-moments estimation of the four parameter’s values from historical returns.

Seneta [16] derives the following relationships between the log gross return moments and the four parameters of the model, notated $c$, $\sigma^2$, $\theta$, and $\nu$:

\begin{align*}
E[\log R_h - \log R_b] &\approx c_h + \theta_h \\
Var[\log R_h - \log R_b] &\approx \sigma_h^2 \\
Skew[\log R_h - \log R_b] &\approx \frac{3\theta_h \nu_h}{\sqrt{\sigma_h^2}} \\
Kurt[\log R_h - \log R_b] &\approx 3(1 + \nu_h) 
\end{align*}

As in the lognormal specification, we maintain the assumption that $E[\log R_h - \log R_b] \approx c_h + \theta_h > 0$, so that $\gamma_h < 0$ and the infimum in (10) is achieved at $x = 0$. In addition, it will prove convenient to perform the substitution $\gamma_h := -\gamma_h > 0$. Then
\[ D_h^{Variance\ Gamma} = \max_{\gamma_h} c_h \gamma_h + \frac{1}{\nu_h} \log \left[ 1 + \theta_h \nu_h \gamma_h - \sigma_h^2 \nu_h \gamma_h^2 / 2 \right] \]  

(16)

The Envelope Theorem of unconstrained optimization permits us to ignore the maximization operation in (16) when computing the partial derivatives of \( D_h \). So from (16) and \( \gamma_h > 0 \) we see that an increase in \( c_h \) increases the performance index, i.e. lowers the underperformance probabilities as time passes, as it (desirably) does in the lognormal case previously analyzed. An increase in \( \sigma_h^2 \) will lower the performance index, undesirably resulting in higher underperformance probabilities. This is also in accord our analysis of the lognormal case. But asset returns (including the log gross hedged excess returns we have examined) typically exhibit excess kurtosis, i.e. \( \nu_h > 0 \). If in addition \( \theta_h > 0 \), (15) shows that skewness is positive, in which case (16) shows that an increase in the positive skewness increases the performance index. But the effect of an increase in excess kurtosis (i.e. an increase in \( \nu_h \)) is ambiguous when \( \theta_h > 0 \).

Now suppose \( \theta_h < 0 \). (15) shows that there is negative skewness, in which case (16) shows that an increase in the positive \( \nu_h \) (i.e. an increase in excess kurtosis) lowers the performance index and hence raises the probability of underperformance. This is because an already fatter-than-normal left hand tail is further fattened by higher kurtosis. Finally, if \( \theta_h = 0 \), there is no skewness and (16) shows that an increase in the positive \( \nu_h \) (and hence excess kurtosis) has a similarly undesirable effect, as the distribution becomes more peaked with more weight in the left-hand tail (as well as the right-hand tail).

In summary, this parametric sensitivity analysis is entirely in accord with intuition:
higher mean and positive skewness is good; higher variance and negative skewness is bad; while the effect of higher excess kurtosis is bad when returns are negatively skewed and dependent on specific parameter values when returns are positively skewed.

Moreover, the Variance Gamma process nests the lognormal specification analyzed earlier. To see this, note that the log in (16) is one plus the difference of two terms that depend on the process parameters and the maximizing $\gamma$. When the difference of the two terms is small relative to 1, one may employ the approximation $\log(1 + y) \approx y$. Then (16) is approximately:

\[
D_h \approx \max_{\gamma_h} c_h \gamma_h + \frac{1}{\nu_h} \left[ \theta_h \nu_h \gamma_h - \sigma_h^2 \nu_h \gamma_h^2 / 2 \right]
\]
\[
= \max_{\gamma_h} c_h \gamma_h + \left[ \theta_h \gamma_h - \sigma_h^2 \gamma_h^2 / 2 \right]
\]
\[
= \frac{1}{2} \left( \frac{c_h + \theta_h}{\sigma_h} \right)^2.
\]  (17)

Using the first line in (15) to interpret the numerator in (17) shows that the expression is identical to the lognormal underperformance probability decay rate (14). This approximation will be valid in many applications, including the hedging application examined empirically in the next section, even if the log gross excess returns have significant skewness and excess kurtosis. To see why, divide lines three and four of (15) to find:

\[
\frac{\text{Skew}[\log R_h - \log R_b]}{\text{Kurt}[\log R_h - \log R_b] - 3} = \frac{\theta_h}{\sigma_h}
\]  (18)

that is, the ratio of skewness to excess kurtosis is the ratio of $\theta_h$ to $\sigma_h$. In a “good” cross-hedging situation, the spot-futures correlation is high enough to drive the standard deviation $\sigma_h$ much lower than the unhedged standard deviation $\sigma_0$. Because the ratio of skewness
to excess kurtosis is rarely very highly positive or very highly negative, the value of $\theta_h$ is something on the order of the small $\sigma_h$ achieved by hedging. Moreover, the log gross hedged returns rarely have a large mean, so $c_h$ is small, too. As a result, the approximate value of the maximizing $\gamma_h \approx \frac{c_h + \theta_h}{\sigma_h^2}$ isn’t too large to invalidate the requirement that $y \equiv \theta_h c_h \gamma_h - \sigma_h^2 c_h \gamma_h^2 / 2$ is near enough to zero to justify the approximation $\log(1 + y) \approx y$ used to derive (17).

In summary, in applications where a “good” hedge exists, nonzero levels of skewness and kurtosis in the hedged returns typically will not result in a radically different optimal hedge ratio $h^*$ maximizing (16) than that produced by maximizing (14).

4 An Empirical Illustration

Foster and Whiteman [6] exhaustively studied a typical cross-hedging problem: hedge the spot price of soybeans produced in one area (North Central Iowa) using near-term futures contracts that deliver in another location (the Chicago Board of Trade’s Soybean Futures contracts). To avoid market microstructure influences, Stoll and Whaley (op.cit.) advocate estimation of optimal hedge ratios based on the statistics of weekly returns rather than daily returns. To estimate the decay rate maximizing hedge ratio (16), first calculate the historical log gross weekly hedged returns’ mean, variance, skewness and kurtosis.$^4$ Note from (15) that method of moments parameter estimates for each value of the hedge ratio $h$ are immediately found by setting $\sigma_h^2$ equal to the historical variance of the log gross hedged returns. The value of $\nu_h$ is determined by the kurtosis.$^5$ The estimate of $\theta_h$ equals the historical skewness
times the sample standard deviation divided by $3\nu_h$, and finally the estimate of $c_h$ is simply the historical mean minus the already determined value of $\theta_h$. Then, a numerical maximizer is used to find the values of $h$ and $\gamma_h$ that jointly maximize the estimate of (16). Table 1 reports the estimated maximizing $h^* = -0.42$, and the corresponding maximizing $\gamma_{h^*} = 0.14$.

In addition to the small size of the estimated $\gamma_{h^*}$, Table 1 shows that the corresponding estimated values of $c_h, \theta_h, \text{and } \sigma_h$ are quite small. As a result, the reasoning used to justify the approximation (17) holds, so it’s not surprising to find that the maximizing the latter also produces $h^* = -0.42$ to the nearest hundredth. For comparison, the estimated minimum variance hedge ratio (3) is $h_{var} = -0.61$.


<table>
<thead>
<tr>
<th>h</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Excess Kurtosis</th>
<th>c</th>
<th>$\sigma$</th>
<th>$\theta$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4.9 bps</td>
<td>9.5 bps</td>
<td>-0.08</td>
<td>4.22</td>
<td>27 bps</td>
<td>3.1%</td>
<td>-27 bps</td>
<td>1.4</td>
</tr>
<tr>
<td>-0.61</td>
<td>0.5 bps</td>
<td>4.1 bps</td>
<td>-0.75</td>
<td>6.41</td>
<td>2.4 bps</td>
<td>2.0%</td>
<td>-2.4 bps</td>
<td>2.14</td>
</tr>
<tr>
<td>-0.42</td>
<td>0.7 bps</td>
<td>5.1 bps</td>
<td>-0.69</td>
<td>5.14</td>
<td>31.3 bps</td>
<td>2.3%</td>
<td>-30.5 bps</td>
<td>1.71</td>
</tr>
<tr>
<td>-1</td>
<td>-1.1 bps</td>
<td>4.2 bps</td>
<td>-0.38</td>
<td>11.15</td>
<td>5.9 bps</td>
<td>2.0%</td>
<td>-7.0 bps</td>
<td>3.72</td>
</tr>
</tbody>
</table>

**4.1 A Robustness Check**

It is important to examine whether or not the above results are sensitive to the probability distribution specified for the log gross hedged returns. Perhaps the distribution does not arise from the Variance Gamma process specification. Fortunately, there is a simple nonparametric way to estimate the optimal hedge ratio when the unknown process is IID. We have historical time series of the spot and futures returns that are used to construct historical hedged excess returns. Because this is an IID process, an analog estimator (see Manski [12]) is to substitute
an historical time average for the expectations operator in (11), and substitute that into (10) to produce the estimated decay rate $D_h$:

$$D_h^{IID} = \max_{\gamma} \gamma - \log \frac{1}{T} \sum_{t=1}^{T} e^{-\gamma (\log R_{ht} - \log R_{bt})}$$

(19)

where we again redefine $\gamma := -\gamma$. The sum in (19) is over the $T = 595$ weeks in the historical sample of weekly soybean spot and futures returns. Jointly maximizing (19) over $h$ and $\gamma$ produces the same results reported in Table 1, to the nearest hundredth.

4.2 Discussion

The results complement the findings in Brooks, et.al. [2]. They studied hedge ratios maximizing a single-period expected utility, which in principle depend on skewness, kurtosis and all higher than 2nd order moments. But in realistic empirical applications with plausible degrees of hedger risk aversion, they showed that the estimated optimal hedge ratios are also close to those obtained without the use of higher than 2nd order moments.\(^6\)

Statistical tests sometimes reject the IID assumption in favor of more complex (albeit ad-hoc) specifications of parametric statistical models for hedged returns (e.g. GARCH). In those models, the optimal hedge ratio varies over time, conditional on model-based estimates of the conditional return distributions at each point in time. But it is important to note that statistically significant deviations from the IID assumption need not lead to time-varying optimal hedge ratios that deliver *economically significant* hedging benefits. Perhaps the most comprehensive corroboration of this is found in Poomimars, Cadle, and Theobald
They contrasted both the in-sample and out-of-sample variance reductions achievable by eight different dynamic hedging models with the in-sample and out-of-sample variance reductions achieved by the optimal hedge ratio \( h = -1 \). The comparisons were made for seven different assets: three stock indices, two currencies, and the precious metals gold and silver. They documented very little intertemporal variation in the potentially time-varying hedge ratios, and hence found little performance improvement from dynamic hedging. Similar findings were reported in Alexander and Barbosa [1].

Finally, we could substitute (9) into (10) to analyze optimal static hedges in non-IID situations satisfying Assumption 1. Efficient nonparametric estimators are more complicated to derive when the process isn’t IID. The block smoothing technique developed in Kitamura and Stutzer [9] may be used for weakly dependent processes commonly assumed in economics. A similar estimator was studied by Duffy and Metcalf [5]. But the Gärtner-Ellis Theorem may not be applicable in some situations. If hedged returns have sufficiently heavy tails, there may be not be exponential decay, invalidating its use. Empirical estimates for the degree of heavy-tailedness must be done with some care, as noted by Lux [10].
Notes

1For the purpose of exposition, we adopt the common textbook pedagogical assumption that ignores any opportunity cost of margins and commissions.

2Hence $R_{ft}$ should not be interpreted as a rate of return. The rate of return for the futures contract itself is undefined, because unlike the spot security, the futures contract has no initial cost to use as the base for calculating a conventional rate of return.

3Recall that capital letters are used to denote gross returns, i.e. one plus net returns.

4We thank Doug Foster for providing the weekly data series needed for this purpose. Foster and Whiteman (op.cit.) used a subset from a database assembled by Sergio Lence.

5Unlike in Seneta (op.cit.), some statistical software reports kurtosis as an excess (positive or negative) from the normal distribution’s value of 3. If so, (15) implies that one should estimate $v_h$ to be $1/3$ of the reported excess kurtosis.

6e.g. see the first three columns in their Table V.
References


