Chapter 8 — Efficient Portfolios

Efficient Portfolios and Risk

In the CAPM, mean-variance analysis makes evaluating portfolios quite easy. Take any mean-variance efficient portfolio. If another portfolio has a higher expected rate of return, then it must be riskier. Conversely, if another portfolio is less risky, it must have a lower expected rate of return. This can be demonstrated easily in the figure.

This result is not so clean in general when risk is defined in a Rothschild Stiglitz sense, rather than by variance. However, some partial results do hold. In this general sense, an efficient portfolio is one that is optimal for some utility function. A rate of return, $r^*$, that is strictly riskier than that on some other portfolio, $\tilde{r}$. Then

$$r^* - \mathbb{E}[\tilde{r}^*] = \tilde{r} - \mathbb{E}[\tilde{r}] + \tilde{\varepsilon} \quad \text{or} \quad r^* = \tilde{r} + \eta + \tilde{\varepsilon} \quad \text{with} \quad \eta \equiv \mathbb{E}[\tilde{r}^* - \tilde{r}] \quad \text{and} \quad \mathbb{E}[\tilde{\varepsilon} \mid r] = 0. \quad (1)$$

The random variable $\eta$ is degenerate with no variation so if $\eta < 0$ then $\tilde{r}$ strictly second-order stochastically dominates $r^*$. In this case, the portfolio with return $\tilde{r}$ could not be optimal for any utility function as had been assumed. If there is a risk-free asset, $r_f$ can be used for $\tilde{r}$ so all efficient portfolios must have expected rates of return in excess of the interest rate.

The contrapositive of this statement is: If $\mathbb{E}[\tilde{r}] \geq \mathbb{E}[r^*]$, then $\tilde{r}$ cannot be strictly less risky. Under mean-variance analysis, this means that var[$r^*$] $\leq$ var[$\tilde{r}$]. However, in general, riskiness is not a complete ordering so it is possible to have $\mathbb{E}[\tilde{r}] \geq \mathbb{E}[r^*]$ with $\tilde{r}$ not being riskier than $r^*$ but rather non-comparable in risk. Furthermore, in general, the efficient set does not include every portfolio for which there is not another less risky portfolio with the same expected rate of return as it does in the CAPM.

Verifying the Efficiency of a Given Portfolio

In many partial equilibrium models a particular portfolio is acknowledged as being held by the representative investor and pricing results are derived. This is accomplished by identifying the stochastic discount factor as the representative investor’s marginal utility $m_{\text{rep}} \equiv u'(\tilde{R}_{\text{rep}})$ and $\mathbb{E}[m_{\text{rep}}(1 + \tilde{r})] = 1 \forall n$. For this to be a valid pricing relation the gross return on the chosen representative portfolio must be efficient — that is, optimal for some increasing concave utility function. Indeed, it must be optimal for the utility function chosen as characterizing the representative investor. But here a more basic question is being examined. Give a portfolio, is there some increasing concave utility function for which it is optimal?

Under mean-variance analysis, this is a simple matter. All mean-variance efficient portfolios are combinations of the tangency portfolio and the risk-free asset. Or if there is no risk-free asset, all efficient portfolios are combinations of any two minimum-variance portfolios. A simple rule like that is not true in general. A method for verifying the efficiency of a portfolio in the general case is given in the following theorem.

**Theorem 8.1: Efficient Portfolios.** For a given portfolio $w$, the payoff at time 1 is $x^w = Xw$. Order the states so that $x^w_1 \leq x^w_2 \leq \cdots \leq x^w_n$. The portfolio is efficient if and only if there is a vector $m$ and positive constant $\lambda$ satisfying

$$\sum \pi_n x^w_n m_n = \lambda \quad \forall n \quad \text{and} \quad m_1 \geq m_2 \geq \cdots \geq m_n > 0. \quad (2)$$

**Proof:** The first-order condition for an optimal portfolio requires that for each asset $n$

$$\sum \pi_n x^w_n u'(x^w_n) = \lambda \quad \forall n. \quad (3)$$
Identify $m_i = u'(x_i)$. The summation constraint is satisfied by construction. Because marginal utility is positive and decreasing, the ordering constraint is also satisfied.

Note that the only way a particular portfolio affects this problem is by the order it imposes on the outcomes across states. If outcome $x$ is efficient because it satisfies (2) for utility function $u$, then outcome $y$ is also efficient if it has the same state ordering as $x$ because it is optimal for utility function $v$ with $v'(y_i) = u'(x_i)$. So all portfolios that have the same ordering of outcomes are efficient if any one of them is.

Equation (2) is a set of linear equations in $m$. Assuming there are no redundant assets and no arbitrage, it imposes $N \leq S$ constraints on $S$ variables. If $N = S$, there is a unique solution for $m$ (and therefore for state prices $q_s = m_i \pi_s$) so all efficient portfolios have the same ordering with homogeneous beliefs when markets are complete. This repeats the result in Chapter 4. However, if markets are incomplete, there is more than a single vector $m$ that satisfies (2), and different valid orderings for the elements of $m$ are possible. So efficient portfolios need not all be monotonically related. A simple example is provided next.

The simplest such model has two assets and three states. There must be at least two assets for meaningful portfolio formation, and there must be more states than assets if markets are to be incomplete.

In a three-state two-asset economy, the transpose of the payoff (or gross return) matrix, $X'$, is

<table>
<thead>
<tr>
<th>X'</th>
<th>state</th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>asset</td>
<td>1</td>
<td>0.6</td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.7</td>
<td>1.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

With three states there is a potential for $3! = 6$ orderings; however, portfolios of only two assets can achieve at most only four of these. The state outcome ordering changes at the portfolio weights that make two outcomes equal. Increasing the weight in one asset makes the outcome larger in one state while increasing the other asset’s weight makes the outcome in the other state larger. In this example, the outcomes are the same in states $a$ and $b$ when the one quarter of the portfolio is invested in asset 1; $w = 0.25$. The outcomes in states $a$ and $c$ are the same when $w = 0.5$, and the outcomes in states $b$ and $c$ are the same when $w = 0.7$. So returns are ordered $(c, b, a)$ from low to high for $w < 0.25$. They are ordered $(c, a, b)$ when $0.25 < w < 0.5$. They are ordered $(a, c, b)$ when $0.5 < w < 0.7$. They are ordered $(a, b, c)$ when $0.7 < w$. With three states there can be at most three points at which two states have the same return. This divides the outcomes into four regions. Which of the four possible orderings represent efficient portfolios is yet to be determined.

The efficiency of a portfolio can be determined by solving (2) for possible SDFs. If the state probabilities are $\pi' = (0.2, 0.2, 0.6)$, then

$$0.2 \cdot 0.6 m_a + 0.2 \cdot 1.2 m_b + 0.6 \cdot 1.5 m_c = 1 \Rightarrow m_b = \frac{350 - 207m_a}{129} \quad m_c = \frac{50 + 38m_a}{129} \quad (4)$$

The figure plots the three marginal utilities as functions of $m_a$. There are four outcomes that represent possible efficient portfolios. Recalling that marginal utilities are decreasing in outcomes, the possible portfolios ordered from low to high in outcomes are $(b, c, a), (b, a, c), (a, b, c)$, and $(a, c, b)$ from left to right in the figure. The first two orderings are not possible with these assets. The final two are the portfolios with $w > 0.7$ and $0.5 < w < 0.7$. $0.5 < w < 0.7$.

1 There could be fewer than four different orderings possible if all three states had the same outcome at some portfolio weighting. Then only two outcome orderings would be possible.
With different state probabilities, different portfolio orderings become efficient. For example, if the state probabilities are equal, then solving (2) for possible SDFs gives
\[
\begin{align*}
\frac{1}{2} \cdot 0.6m_a + \frac{1}{3} \cdot 1.2m_b + \frac{1}{3} \cdot 1.5m_c &= 1 \\
\frac{1}{3} \cdot 1.7m_a + \frac{1}{3} \cdot 1.5m_b + \frac{1}{3} \cdot 0.8m_c &= 1
\end{align*}
\Rightarrow m_b = \frac{70 - 69m_a}{43}, \quad m_c = \frac{30 + 38m_a}{43}. \tag{5}
\]
For these probabilities, outcome orderings \((b, c, a), (b, a, c), (a, b, c),\) and \((a, c, b)\) represent possible efficient portfolios. The first ordering is not feasible; the second and third come from portfolios with \(w < 0.25\) and \(0.25 < w < 0.5\).

In the two examples, either all portfolios with \(w < 0.5\) or all portfolios with \(w > 0.5\) are efficient. This shows that the example was too simple. There is one portfolio that can be identified as (in the closure of) the efficient set even without specifying the state probabilities. It is the portfolio that has the largest possible worst outcome. This portfolio is the one that would be held by an investor with infinite risk aversion. In the current example, it is has \(w = 0.5\). The gross returns on this portfolio are 1.15 in states \(a\) and \(c\) and 1.35 in state \(b\). Whenever, two assets have different expected rates of return, a risk-neutral investor takes an infinitely long position in the one with the higher expected rate of return by infinitely shorting the other. An investor who is close to risk-neutral holds a portfolio with a very large position in the one with the larger expected rate of return. So when \(\bar{r}_1 > \bar{r}_2\) then all portfolios with \(w > 0.5\) are efficient and vice versa.

The problem with this example is that the set of all portfolios is described by a single number, \(w\) or \(\bar{r}_p\), so the efficient set must be all those with \(\bar{r}_p\) large enough. This means that the representative investor is “too easy” to find. More assets are needed to understand the problem.

---

2 It can be found as the limit of the solution to the problem \(\lim_{u \to +\infty} \max_w \mathbb{E}[-\exp(-aW(1 + w\bar{r}))].\)

3 The largest possible worst outcome portfolio is always one of those where two state outcomes are the same. Altering some portfolio weights will decrease one of the outcomes so that the portfolio no longer has the largest possible worst outcome.
Portfolio Efficiency and the Representative Agent

In the previous section we saw how to verify the efficiency of a given portfolio. This allows us to specify a representative investor. In many models, this step is skipped. The representative agent is assumed to be the average investor. That is, his consumption is per capita consumption, and his portfolio is the market portfolio. In some models like the CAPM, the specific representative agent can be found by construction. In other models, like those with effectively complete markets, we know that a representative agent exists, but determining his utility function is neither obvious nor easy. More troublesomely, in some economies a risk-averse average agent may not exist even when all agents are risk averse and have homogeneous beliefs. The following example illustrates.4

There are three assets and four states. The asset returns are

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>asset 1</td>
<td>3.3</td>
<td>2.2</td>
<td>2.6</td>
<td>2.5</td>
</tr>
<tr>
<td>asset 2</td>
<td>4.1</td>
<td>2.6</td>
<td>1.9</td>
<td>2.5</td>
</tr>
<tr>
<td>asset 3</td>
<td>3.6</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
</tr>
</tbody>
</table>

The set of all portfolio outcome orderings is shown in the figure. The x and y axes represent the portfolio weights $w_1$ and $w_2$. The third weight is $w_3 = 1 - w_1 - w_3$ and need not be illustrated. The regions of different outcome orderings are bounded by the lines on which the outcomes in two states are equal. For example, the returns in states a and b are equal when

$$3.3w_1 + 4.1w_2 + 3.6(1 - w_1 - w_2) = 2.2w_1 + 2.6w_2 + 2.4(1 - w_1 - w_2)$$

$$\Rightarrow w_2 = -4 + \frac{1}{3}w_1 .$$

There are five more such lines

$$a = c \quad a = d \quad b = c \quad b = d \quad c = d$$

These lines and the regions of distinct portfolio orderings are shown in the figure below. The x-axis is $w_1$, and the y-axis is $w_2$. The weight for the third asset is $1 - w_1 - w_2$. A portfolio that lies vertically above or below another has the same $w_1$. A portfolio that lies horizontally left or right another has the same $w_2$. A portfolio that lies exactly north-west or south-east of another on a line with slope $-1$ has the same $w_3$. A portfolio that is above that 45 degree line relative to another portfolio has a smaller $w_3$.

Portfolios to the left of the purple ($b = d$) line have higher returns in state $b$ than $d$. Portfolios above the green ($c = d$) line have higher returns in state $d$ than $c$. Portfolios above the blue ($a = b$) line have higher returns in state $a$ than $b$. So all portfolios in the region bounded by the purple ($b = d$) and green ($c = d$) line have their outcomes ordered $cdba$ from low to high. The orderings for the other regions can be similarly determined. Altogether there are 18 distinct possible portfolio outcome orderings.5

Using the method described above, the regions $CDBA$, $BDCA$, $DBC$, and $DCBA$ are identified as efficient portfolios when the states are equally probable. Possible SDFs for each region are

---

4 This example is based on Dybvig and Ross (1982).

5 There are four points where three lines intersect so only 18 regions of the possible 22 regions are created. The maximum number of regions that $n$ lines can split a plane into is given by the formula $\frac{1}{2}(n^2 + n + 2)$. 

The Portfolio Problem — © Jonathan Ingersoll 4 version: September 12, 2019
regions $BDCA$

$m' = \frac{1}{282}(1472841)$

\[
\begin{align*}
\frac{1}{4} \left( 3.3 \times \frac{1}{282} + 2.2 \times \frac{47}{282} + 2.6 \times \frac{28}{282} + 2.5 \times \frac{41}{282} \right) &= 1 \\
\frac{1}{4} \left( 4.1 \times \frac{1}{282} + 2.6 \times \frac{47}{282} + 1.9 \times \frac{28}{282} + 2.5 \times \frac{41}{282} \right) &= 1 \\
\frac{1}{4} \left( 3.6 \times \frac{1}{282} + 2.4 \times \frac{47}{282} + 2.4 \times \frac{28}{282} + 2.4 \times \frac{41}{282} \right) &= 1
\end{align*}
\]

regions $CDBA$

$m' = \frac{1}{831}(210280400390)$

\[
\begin{align*}
\frac{1}{4} \left( 3.3 \times \frac{210}{831} + 2.2 \times \frac{280}{831} + 2.6 \times \frac{400}{831} + 2.5 \times \frac{390}{831} \right) &= 1 \\
\frac{1}{4} \left( 4.1 \times \frac{210}{831} + 2.6 \times \frac{280}{831} + 1.9 \times \frac{400}{831} + 2.5 \times \frac{390}{831} \right) &= 1 \\
\frac{1}{4} \left( 3.6 \times \frac{210}{831} + 2.4 \times \frac{280}{831} + 2.4 \times \frac{400}{831} + 2.4 \times \frac{390}{831} \right) &= 1
\end{align*}
\]

regions $DBCA$

$m' = \frac{1}{27}(4131214)$

\[
\begin{align*}
\frac{1}{4} \left( 3.3 \times \frac{4}{27} + 2.2 \times \frac{13}{27} + 2.6 \times \frac{12}{27} + 2.5 \times \frac{14}{27} \right) &= 1 \\
\frac{1}{4} \left( 4.1 \times \frac{4}{27} + 2.6 \times \frac{13}{27} + 1.9 \times \frac{12}{27} + 2.5 \times \frac{14}{27} \right) &= 1 \\
\frac{1}{4} \left( 3.6 \times \frac{4}{27} + 2.4 \times \frac{13}{27} + 2.4 \times \frac{12}{27} + 2.4 \times \frac{14}{27} \right) &= 1
\end{align*}
\]

regions $DCBA$

$m' = \frac{1}{69}(14283234)$

\[
\begin{align*}
\frac{1}{4} \left( 3.3 \times \frac{14}{69} + 2.2 \times \frac{28}{69} + 2.6 \times \frac{32}{69} + 2.5 \times \frac{34}{69} \right) &= 1 \\
\frac{1}{4} \left( 4.1 \times \frac{14}{69} + 2.6 \times \frac{28}{69} + 1.9 \times \frac{32}{69} + 2.5 \times \frac{34}{69} \right) &= 1 \\
\frac{1}{4} \left( 3.6 \times \frac{14}{69} + 2.4 \times \frac{28}{69} + 2.4 \times \frac{32}{69} + 2.4 \times \frac{34}{69} \right) &= 1
\end{align*}
\]

The portfolios in the other 12 regions, those denoted by lower case letters, are inefficient; they are not optimal for any increasing, concave, state-independent utility function. The portfolio that maximizes the worst outcome is asset 3 by itself. This is located at the origin. As in the two-asset case, this portfolio lies on the border of efficient portfolios. In this case it is on two borders, both the green ($c = d$) and purple ($b = d$) lines. It is also on the yellow ($b = c$) line, but that one does is not the efficient-inefficient border. The asset’s expected gross returns are 2.65, 2.775, and 2.7 so a risk-neutral investor holds a portfolio very long in asset 2 and very short in asset 1. That
is, the risk-neutral portfolio lies in the $CDBA$ region. By themselves, these two portfolios can only identify that one region as efficient.

The set of optimal efficient portfolios have been identified. But this has revealed a potentially serious problem with identifying the representative investor. From the figure, it is clear that the efficient set is not convex. Convex combinations of some portfolio in $CDBA$ and $BDCA$ lie in $bcda$ or $cbda$, and those portfolios are not efficient. This can be demonstrated directly.

Assets 1 and 2 are efficient because their outcomes are ordered $BDCA$ and $CDBA$ so they must be efficient as already shown. Specifically, they are the optimal portfolios for two investors whose marginal utilities are characterized by

\[
\begin{align*}
 u_1'(3.3) &= \frac{1}{282} < u_1'(2.6) = \frac{28}{282} < u_1'(2.5) = \frac{41}{282} < u_1'(2.2) = \frac{47}{282}, \\
 u_2'(4.1) &= \frac{210}{831} < u_2'(2.6) = \frac{280}{831} < u_2'(2.5) = \frac{390}{831} < u_2'(1.9) = \frac{400}{831}.
\end{align*}
\]

or any scale multiples thereof. That asset 1 and 2, respectively, satisfy the first-order conditions for an optimal portfolio for these two utility functions is immediate from (8).

Now consider a portfolio that is equally weighted in assets 1 and 2. Its returns are 3.7, 2.4, 2.25, and 2.5 in states $a$, $b$, $c$, and $d$, respectively. The outcome ordering of this portfolio is $cbda$ which we know to be inefficient. This can also be demonstrated directly. Not only is it inefficient for every risk averse investor, it is first order stochastically dominated so it is not optimal for any investor with increasing utility whether risk averse or not.

One portfolio that stochastically dominates it holds $3/4$ in asset 2 and $1/4$ in asset 3. This portfolio has returns 3.7, 2.55, 2.4 and 2.475. Its returns in states $a$ and $c$ are the same as the returns in states $a$ and $b$ on the first portfolio. Its returns in states $b$ and $d$ are higher than those on the first in states $d$ and $c$, respectively. Because the states are equally probable, this means that the second portfolio first-order stochastically dominates the equally weight portfolio.

Because the efficient set is not convex, it is possible that the market portfolio, which is a convex combinations of efficient portfolios, might not be efficient. If this is true, then there is no concave utility function for which the average (market) portfolio is optimal. That is, there can be no representative investor who is average. 6

The two investors’ SDFs differ as shown in the first two parts of equation (8). Because each state has the same probability, the investors’ subjective state prices are proportional to their private SDF and also misaligned.

\[
q_{a1} < q_{c1} < q_{d1} < q_{b1} \quad \text{and} \quad q_{a2} < q_{b2} < q_{d2} < q_{c2}
\]

Both investors could benefit if investor 1 traded for more return in state $b$ in exchange for giving higher return in state $d$ to investor 2. Similarly, investor 1 could trade for more return in state $c$ in exchange for either state $b$ or $d$. However, the existing assets do not allow these trades in a satisfactory manner that does not create other inefficiencies.

The violation of convexity in this example cannot be eliminated by weakening any of the perfect market assumptions. All three assets have limited liability. The two optimal portfolios and the stochastic dominating portfolio do not hold any negative positions so short sale restrictions would not help, and they have strictly positive payoffs so bankruptcy is not an issue.

This example has minimal complexity. As seen previously, the efficient set must be convex with just two assets so a nonconvex efficient set requires at least three assets. In addition, the efficient set is always convex in a complete market because all efficient portfolios have the

---

6 There are only a few models where the representative agent is typical, but not average. For example, Constantinides and Duffie (1996).
same ordering of outcomes, and lie in a single region. Each region is bounded by hyperplanes so must be convex.

Under state-dependent utility, essentially any portfolio might be efficient as marginal utility need not be higher in states with low returns. Marginal utility need only decrease within a state as consumption increases. Similarly, with heterogeneous beliefs, portfolios can be efficient for some beliefs and not for others. So with no restrictions on state-dependent utility or heterogeneous beliefs, all feasible portfolios could be efficient in which case the market portfolio would also be efficient, but little could be said about its use for pricing.

When utility is state-independent and beliefs are homogeneous, a complete market (or just two assets?) is the only general guarantee that the efficient set is convex so that the market portfolio is efficient. In the absence of a complete market, the convexity of the efficient set and the efficiency of the market portfolio can be verified if certain types of separation or mutual fund theorems hold. These are discussed in the next chapter.

The Hansen-Jagannathan Lower Bound

Testing the CAPM is straightforward. Portfolios are evaluated using Sharpe ratios. Asset’s expected excess returns should be equal to their betas multiplied by the expected excess return on the market. That does not mean testing is a simple task, only that what needs verification is obvious. In contrast, the general pricing results in this chapter are empirically vaguer. The market is not necessarily efficient so some efficient portfolio must be found, and the utility function for which it is optimal must be identified. In fact, in a large economy with many possible states, it can be quite difficult to assess even if there is arbitrage.

The Hansen-Jagannathan (1991) lower bound can be used to reject some equilibrium models. The only assumptions required to apply this bound are that the Law of One Price holds and that asset return variances are finite. Of course, the latter is always true in sample and it is reasonable to assume it must characterize beliefs about the future as well as the world is finite.

**Theorem 8.2: Hansen-Jagannathan Lower Bound.** If there is a risk-free asset, the Law of One-Price holds, \( \mathbb{E}[\tilde{m}^2] < \infty \), and the covariance matrix of returns is bounded, \(^8\) then the ratio of the standard deviation of a stochastic discount factor to its mean exceeds the Sharpe Ratio attained by any portfolio

\[
\frac{\sqrt{\text{var}[\tilde{m}]} }{\mathbb{E}[\tilde{m}]} \geq \frac{\mathbb{E}[\tilde{r}_w] - r_f}{\sqrt{\text{var}[\tilde{r}_w]}} \forall w : 1'w = 1. \tag{11}
\]

**Proof:** The Law of One Price guarantees the existence of a SDF, \( \tilde{m} \), which, as shown in equation (31) of Chapter 3, prices assets as

\[
\mathbb{E}[\tilde{r}] - r_f 1 = -(1 + r_f) \text{cov}(\tilde{m}, \tilde{r}). \tag{12}
\]

---

\(^7\) With only two assets, the feasible set is described by a single parameter, the fraction of the portfolio in asset 1. Under homogeneous beliefs, the efficient set is either all portfolios with \( w \geq \tilde{w} \) or all portfolios with \( w \leq \tilde{w} \) depending on whether the expected rate of return on asset 1 exceeds or is less than the expected rate of return on asset 2.

\(^8\) The assumption that \( \mathbb{E}[\tilde{m}^2] \) and the covariance matrix are finite are obviously necessary to make comparisons involving variances.

\(^9\) If there is no risk-free asset, the Hansen-Jagannathan lower bound is derived from equation (30) in Chapter 3.

\[
1 = \mathbb{E}[\tilde{m}(1 + \tilde{r}_w)] = \mathbb{E}[\tilde{m}]\mathbb{E}[1 + \tilde{r}_w] + \text{cov}(\tilde{m}, (1 + \tilde{r}_w)) \Rightarrow \sqrt{\text{var}[\tilde{m}]} \geq \sqrt{\mathbb{E}[\tilde{m}]\mathbb{E}[1 + \tilde{r}_w] - 1}/\sqrt{\text{var}[\tilde{r}_w]} \forall w.
\]
For portfolio, $\mathbf{w}$, $\text{cov}(\mathbf{m}, \mathbf{w}) = \text{corr}(\mathbf{r}_w, \mathbf{m})\sqrt{\text{var}[\mathbf{r}_w]}\text{var}[\mathbf{m}]$. Therefore,

$$\left(\frac{\mathbb{E}[\mathbf{r}_w] - r_f}{\sqrt{\text{var}[\mathbf{r}_w]}}\right) = (1 + r_f)\left|\text{corr}(\mathbf{r}_w, \mathbf{m})\right|\sqrt{\text{var}[\mathbf{m}]} \leq (1 + r_f)\sqrt{\text{var}[\mathbf{m}]}.$$

(13)

The inequality follows because the correlation must be less than 1 in absolute value. The left-hand side of (13) is the absolute value of the Sharpe ratio, measuring the portfolio’s risk-return tradeoff. Because the SDF prices the risk-free asset, $\mathbb{E}[\mathbf{m} \cdot 1] = (1 + r_f)^{-1}$ and (11) follows.

This bound applies for every SDF. Typically, the most convenient SDF is the payoff-SDF. As shown in Chapter 3, this can be constructed as a projection of the SDF onto the space spanned by the asset returns. Provided there is a risk-free asset, the payoff-based SDF, $\tilde{\mathbf{x}}_m$, is the one with the smallest variance. For any other SDF, write $\mathbf{m} = \tilde{\mathbf{x}}_m + \mathbf{u}$. Then

$$\text{cov}[\mathbf{u}, \tilde{\mathbf{x}}] = \mathbb{E}[\mathbf{u} \tilde{\mathbf{x}}] - \mathbb{E}[\mathbf{u}]\mathbb{E}[\tilde{\mathbf{x}}] = \mathbb{E}[\mathbf{m} - \tilde{\mathbf{x}}_m] \tilde{\mathbf{x}} - \mathbb{E}[\mathbf{m} - \tilde{\mathbf{x}}_m] \mathbb{E}[\tilde{\mathbf{x}}] = \mathbf{0}.$$

(14)

The first term is zero because the expectation of the product of each SDF with the payoffs must equal the price vector $\mathbf{p}$. The second term must be zero because the expectation of the each SDF is $1/(1 + r_f)$, the value of a constant payoff of 1.\(^{10}\) So

$$\text{var}[\mathbf{m}] = \text{var}[\tilde{\mathbf{x}}_m + \mathbf{u}] = \text{var}[\tilde{\mathbf{x}}_m] + \text{var}[\mathbf{u}] \geq \text{var}[\tilde{\mathbf{x}}_m].$$

(15)

For any SDF that differs at all from $\tilde{\mathbf{x}}_m$ the bound is strict. So $\tilde{\mathbf{x}}_m$ is the SDF with the strictly smallest variance. The simple intuition for this result is that adding any variation in the SDF that is outside the span of the payoffs cannot change the pricing relation, but will be uncorrelated with the SDF and so increase its variance.

The asset based SDF is the one with the lowest variance, and all SDFs have the same expectation, $\mathbb{E}[\mathbf{m}]/1/(1 + r_f)$. Therefore, the asset-based SDF is the one most restricted by the Hansen-Jagannathan lower bound. In fact, the bound is tight for the asset-based SDF because its correlation with the tangency portfolio is $\pm 1$, and the tangency portfolio has the highest Sharpe ratio. Note that this is true by definition; it does not depend on whether the CAPM is valid. The only requirements are that there is a risk-free asset and no risk-free arbitrages so that there are no zero-risk portfolios with returns in excess of the interest rate.

The equity premium puzzle, to be introduced later, is, probably, the most studied apparent violation of the Hansen-Jagannathan lower bound. If there is not much variation in consumption, then there cannot be much variation in the SDF unless investors are very risk averse. This in turn implies that risk premiums must be small or the interest rate is large, but neither appears to be true. In the next section, the bound will be used to limit possible prices of derivative contracts.

### The Cochrane and Saá-Requejo Good-Deal Pricing Bounds

In the previous chapter, joint assumptions about probability distributions and utility functions were used to determine the values of derivative contracts. With no assumptions about utility functions, only no-arbitrage bounds on prices, like those developed in Chapter 3, are

---

\(^{10}\) This is the reason there must be a risk-free asset. With no risk-free asset it may be possible to find an SDF with a smaller variance. Here is a very simple example. There is one asset with a price of 1 and a payoff of 1 or 2 in two equally likely states. Then $\mathbf{m} = 1/\mathbb{E}[\mathbf{x}] = 2/3$, and $\tilde{\mathbf{x}}_m$ has realizations of $2/3$ and $4/3$. However, a random variable with the constant realization of $2/3$ is also a valid SDF because $1 = (1/2)(2/3)1 + (1/2)(2/3)2$. The constant SDF obviously has a smaller variance.
available. These typically provide a wide range of possible prices. The lower bound of Hansen and Jagannathan can be used to construct tighter bounds without an assumption about the representative investor’s utility function.

In the absence of arbitrage, the price of any payoff is determined by the SDF, \[ p = \mathbb{E}[\tilde{m}\tilde{x}] \]. If markets are complete, then the SDF is unique, and the value of any derivative is determined. If the existing assets already span the payoff \( \tilde{x} \), then the Law of one Price determines \( p \) even when the market is incomplete and the SDF is not unique. (The latter is the foundation of the Black-Scholes pricing technology.) If neither of these conditions hold, then an exact value for an unspanned payoff stream can only be determined by making assumptions about the utility function that pin down the SDF. However, bounds on possible valuations can be inferred using the Hansen and Jagannathan lower bound and assumptions about possible risk-return tradeoffs. Cochrane and Saa-Requejo (2000) describe these as “Good Deal” bounds.

The assumption required, beyond the absence of arbitrage and the existence of a risk-free asset, is that the reward for bearing variance is not too high. This does not mean that variance is a proper measure of risk in the Rothschild-Stiglitz sense (or any sense). The assumption is simply that the best Sharpe ratio available cannot be too large. The HJ lower bound is tight for the asset based SDF so

\[
\max_w \left[ \frac{\mathbb{E}[\tilde{r}_w] - r_f}{\sqrt{\text{var}[\tilde{r}_w]}} \right] = \sqrt{\frac{\text{var}[\tilde{x}_m]}{\mathbb{E}[\tilde{x}_m]}} = (1 + r_f) \sqrt{\text{var}[\tilde{x}_m]}. \tag{16}
\]

Therefore, under the assumption there are no investment opportunities in the economy with Sharpe ratio in excess of \( \bar{S} \), the variance of the portfolio SDF is bounded by

\[
\text{var}[\tilde{x}_m] \leq \bar{S}^2 (1 + r_f)^{-2}. \tag{17}
\]

This means the value of any cash flow in the economy, \( \tilde{x} \), must be bounded by

\[
\text{PV}(\tilde{x}) \leq \text{PV}(\tilde{x}) \equiv \max_{\tilde{m}} \mathbb{E}[\tilde{m}\tilde{x}], \text{subject to } \tilde{m} \geq 0, \mathbb{E}[\tilde{m}\tilde{x}] = p, \text{var}[\tilde{m}] \leq \bar{S}^2 (1 + r_f)^{-2}
\]

\[
\text{PV}(\tilde{x}) \geq \min_{\tilde{m}} \mathbb{E}[\tilde{m}\tilde{x}], \text{subject to } \tilde{m} \geq 0, \mathbb{E}[\tilde{m}\tilde{x}] = p, \text{var}[\tilde{m}] \leq \bar{S}^2 (1 + r_f)^{-2}. \tag{18}
\]

In each case, the Law of One Price restriction, \( \mathbb{E}[\tilde{m}\tilde{x}] = p \), is imposed for all assets whose prices are known. In practice, the restriction need be applied only to relevant assets. For example, in bounding the value of an option written on a share of stock, the restriction would apply to the risk-free asset, the stock itself, and any other option contracts on the same stock for which the prices were known. The restriction \( \tilde{m} \geq 0 \) is a no-arbitrage constraint. The variance restriction is the just introduced no-good-deals constraint. Variation in the SDF can permit variation in the resulting present values. Limiting this variation limits the possible present values. Using just the \( \tilde{m} \geq 0 \) and \( \mathbb{E}[\tilde{m}\tilde{x}] = p \) restrictions produces no-arbitrage bounds like those given in [Chapter 3](#). If the variance restriction is not binding in some application, the bound derived is just the no arbitrage bound.

The figure below shows an example of applying the good-deal bounds. The derivative contract is an at-the-money call option written on a share of stock with a price of 100. There are twenty states at the end of the year with prices of 100 \( \cdot (1.05)^s (0.95)^{20-s} \). The state index \( s \) ranges from 0 to 19. State \( s \) occurs with a probability of \( \pi_s = \frac{20!}{s!(20-s)!}(0.55)^s (0.45)^{20-s} \). These are the probabilities from the binomial distribution so along with the product for of the outcomes, the
The stock’s return distribution is approximately lognormal. The stock’s expected rate of return and standard deviation are 9.94% and 23.99%. The interest rate is 5% so the stock has a Sharpe ratio of \( \bar{S} = 0.206 \). The upper and lower bounds are computed limiting the SDF variance to 
\[
\text{var}[\tilde{X}_m] \leq k\bar{S}^2 (1 + r_f)^2
\]
The parameter \( k \) ranges from \( \infty \) to just above 1. That is the lowest possible value for \( k \) as Sharpe ratios obviously cannot be restricted to be less than that already available on the stock. The square root of \( k \) equals the ratio of the maximum Sharpe ratio allowed to the stock’s Sharpe ratio.

For \( k = \infty \), the bounds are the no-arbitrage bounds, \( 4.76 \leq C \leq 45.51 \). These bounds are not 0 and 100 because the state space is discrete and necessarily bounded. As shown in Chapter 3, the lower bound has positive values for the SDF only for the two outcomes surrounding the strike while the upper bound has positive SDF outcomes only for the two extremes.

With a very loose restriction of \( k = 100 \), implying a Sharpe ratio no more than 10 times that on the stock, the lower bound is unchanged as the constraint is not binding while the upper bound drops to 26.89. As \( k \) decreases to 1, the bounds become tighter. Both bounds converge on 10.76.

If there are additional related assets with known prices, then the bounds can be improved, sometimes substantially so. For example, suppose we wish to bound the price of a call option with strike price \( K \) and we know the value of a call option with the same maturity and a strike price \( H \).

One obvious no-arbitrage bound is that the option with the lower strike price must be more valuable, so if \( K < H \), then \( C_K \geq C_H \). Another less obvious arbitrage bound was given in Chapter 3, \( C_K \leq C_H + (H - K)/(1 + r_f) \). For example, if the value of an option with a strike of 120 and a price of 5 is added to the above example, the upper bound on the option is \( 5 + 20/1.05 = 24.05 \). This is quite a bit lower than pervious limit of 45.51. The lower bound isn’t changed much. However, there is still some effect. The positive outcomes for \( \tilde{m} \) cannot be concentrated on the two outcomes surrounding the strike of 100 as that would leave no values of the SDF that were positive when the 120-strike option was in the money. This would make that option’s value of 0 rather than 5. Furthermore, with the three constraints of pricing the stock, bond, and 120-strike option at least three of the outcomes for \( \tilde{m} \) must be positive. The second figure shows the improvements in both Good-Deal bounds when the 120-strike option is added.

Figure 8.3: Good Deal Pricing Bounds for a Call Option Based on Stock and Bond Prices

This figure shows the range of possible option prices arising from a limit imposed on the variance of the SDF across states. As the variance is restricted, the range of possible option prices narrows.
Using additional options would further restrict the bounds. In some cases, an exact price might result. This would be true, of course, if there were as many assets as outcome states. In that case, the market is complete and the state prices (and the SDF) are uniquely determined. So in this example with 20 outcome states, 18 options, along with the stock and the bond, would complete the market and determine the value of any other options exactly. An exact price can arise even with fewer basis assets. For example, if the prices of the stock, the bond, and a put option are known, then put-call parity determines the price of a call option with the same strike regardless of the number of states.

Figure 5 shows the values for the SDF that achieve the upper and lower bounds using just the stock and bond when $k = 10$. For the no-arbitrage upper bound, the SDF has nonzero values only at the two extreme outcomes. The no-arbitrage lower bound only has nonzero values at the two outcomes just above and below $S(1 + r)$. This is just a repeat of the result in chapter 3 using state prices to put no-arbitrage bounds on option prices. When only two outcomes are assigned positive values for $\hat{m}$, they must be large to satisfy the Law of One Price restriction $\mathbb{E}[\hat{m} \tilde{\mathbb{X}}] = \mathbf{p}$ for
the stock and risk-free asset. However, two large values of $\tilde{m}$ with zero for the remaining values produces a high variance. When a restriction is put on the variance, positive values of $m$ must be assigned to more outcomes to keep them all and the variance low. Still, the high values of $\tilde{m}$ are kept near the midrange for the lower bound and near the extremes for the upper bound to produce a low and a high value for the call as desired.

It is obvious that these SDF are not monotonically related to the outcomes. Of course, there is no requirement that they should be for an arbitrary basis asset. The SDF must be monotonically declining with the outcome of the basis asset only when the basis asset is a claim on consumption or wealth (i.e., the market portfolio).

When relevant, imposing the restriction that the SDF must be monotonic further limits the possible option prices. As shown in the figure in the left-most position on the graph, the monotonicity restriction alone imposes lower and upper limits of 10.04 and 13.21. The variance restriction applied alone on the lower bound is never that tight. When the two restrictions are applied together, the variance restriction does not become binding until $k \approx 1.001$ where it increases the lower bound by .0015. The variance restriction applied alone does not improve on the monotonicity upper bound until $k$ is around 4. However, it is binding and the two restrictions together do produce a somewhat tighter upper bound, which decreases from 13.21 to 11.86 as $k$ decreases to 1.

When a monotonicity restriction is not appropriate, there are other ways to improve bounds as shown in the next section.

---

**Gain-Loss Restrictions and Near-Arbitrage**

One limitation of the “Good-Deal” Pricing Bounds is illustrated in Figure 3. The SDF generally takes on high values in some states and low or zero values in other states with few intermediate values. This is problematic because it seems reasonable that the SDF realizations should be more regular. The SDF should never be zero as this means that the corresponding state price is also zero, and there is an arbitrage. In addition, if $m_s = \infty$, then the state price, $q_s$, is also infinite. This too is an arbitrage as unbounded consumption today could be financed by paying 0 in the future. Just as a high Sharpe ratio indicates an investment that is better than should be found in an economy, a very high or low realization for $m_s$, and therefore for $q_s$, indicates there is
a “near arbitrage” that ought to be excluded as well.

Stated in another way, a low Sharpe ratio, as imposed in the Good-Deal bounds does not, by itself, preclude arbitrage. As a simple example, consider the payoff $\tilde{x}$ on a zero-cost portfolio. The payoff has a density function $f(x) = 2x^{-3}$ for $x \geq 1$ and 0 otherwise. This is clearly an arbitrage as the payoff is strictly positive. However, the first two moments of $\tilde{x}$ are

$$E[\tilde{x}] = 2\int_1^\infty x \cdot x^{-3} dx = 2 \quad E[\tilde{x}^2] = 2\int_1^\infty x^2 \cdot x^{-3} dx = \infty$$

(19)

so the variance is infinite. Combining this zero-cost portfolio with the risk-free asset produces a portfolio with a Sharpe ratio of 0 even though it is an arbitrage.

Bernardo and Ledoit (1997) examine pricing bounds when the SDF is uniformly bounded across all states, $m_0 \leq \tilde{m} \leq m_1$. This produces results similar to the variance bound because the maximum variance for a bounded random variable occurs when the probability of each endpoint is ½ and no other outcomes are possible; therefore, bounding the possible realizations of the SDF bounds the variance $\text{var}[\tilde{m}] \leq (m_1 - m_0)^2 / 4$. It also ensures there is no arbitrage. An arbitrage would be a zero or negative realization for the SDF. This section outlines an approach, known as the Gain-Loss Ratio, which expands upon that idea.

Consider the payoff on a zero cost portfolio, $\tilde{x}$. It can be decomposed into its positive and negative parts, $\tilde{x} = \tilde{x}^+ - \tilde{x}^-$, where $\tilde{x}^+ \equiv \text{max}(\tilde{x}, 0)$ and $\tilde{x}^- \equiv \text{max}(-\tilde{x}, 0)$. For any twice differentiable utility function and its associated SDF and optimal consumption

$$0 = E[\tilde{m}\tilde{x}] = E[\tilde{m}\tilde{x}^+] - E[\tilde{m}\tilde{x}^-] \Leftrightarrow \text{GLR} = \frac{E[\tilde{m}\tilde{x}^+]}{E[\tilde{m}\tilde{x}^-]} = 1$$

(20)

for all zero-cost $\tilde{x}$ if there is no arbitrage. The ratio of the expectations is called the gain-loss ratio, GLR.

The GLR is used in a benchmark model. The benchmark posits a SDF. It is usually constructed from an assumed distribution of consumption and utility function such as lognormal and CRRA or normal and CARA, but any strictly positive $\tilde{m}^0$ could be used. The benchmark SDF is what the benchmark investor is willing to pay per unit probability for $1$ in each state. There need be no presumption that the benchmark investor is representative or average. Neither need it be assumed that the benchmark prices all assets correctly. Deviations from price are used to test the efficiency of the model or to provide bounds on possible prices of assets not traded. The benchmark utility-adjusted probabilities are $p^0 = \pi^0 m^0$. $E^0[.]$ denotes expectations with respect to these probabilities.\(^{12}\) This relation gives the following property that is used in the proof of the next theorem.

$$E^0\left[\frac{\tilde{m} \tilde{x}}{m^0} \right] = \sum \pi^0_s \frac{\tilde{m}}{m^0} x_s = \sum \pi_s \tilde{m} x_s = E[\tilde{m}\tilde{x}]$$

(21)

If a zero-cost $\tilde{x}$ can be found with an infinite or zero GLR under the $E^0$ expectation, then that or its short sale is an arbitrage. If an $\tilde{x}$ can be found with $\text{GLR}^0(\tilde{x}) \neq 1$, then a mispricing relative to the benchmark has been identified. Taking a long or short position in $\tilde{x}$ is a “pseudo arbitrage” relative to the benchmark. It is not a true arbitrage because scaling it does not produce the infinite GLR that a true arbitrage has. The GLR is invariant to scaling as both the numerator and denominator change in the same proportion.

**Theorem 8.3: The Gain-Loss Ratio and Pseudo Arbitrage.** Let $X$ be a set of zero-cost payoffs under consideration. It excludes the null payoff $\tilde{x} \equiv 0$ and any arbitrages, $\tilde{x} \geq 0$ and $\tilde{x} \leq 0$.

\(^{12}\) These utility-adjusted probabilities are one possible set of risk-neutral probabilities.
Let $\mathcal{M}$ be a set of SDFs that price all payoffs in $\mathcal{X}$ correctly. I.e., for all $\tilde{m}$ in $\mathcal{M}$, $\tilde{m} \gg 0$ and $\mathbb{E}[\tilde{m} \tilde{x}] = 0 \ \forall \tilde{x} \in \mathcal{X}$. Then

$$\max_{x \in \mathcal{X}} \frac{\mathbb{E}^o[\tilde{x}^+]}{\mathbb{E}^o[\tilde{x}^-]} = \min_{m \in \mathcal{M}} \max_s \left( \frac{m_s}{m_s^o} \right) \min_s \left( \frac{m_s^o}{m_s} \right). \quad (22)$$

**Proof:** For every $\tilde{x}$ in $\mathcal{X}$,

$$\mathbb{E}^o[\tilde{x}^+] \times \min_s \frac{m_s}{m_s^o} \leq \mathbb{E}^o[\tilde{x}^+] = \mathbb{E}^o[\tilde{m} \tilde{x}^+] = \mathbb{E}^o[\tilde{m} \tilde{x}^-] = \mathbb{E}^o[\tilde{x}^-] \leq \max_s \frac{m_s^{o\star}}{m_s} \leq \mathbb{E}^o[\tilde{x}^-] \times \min_s \frac{m_s}{m_s^o} \forall \tilde{x} \in \mathcal{X}. \quad (23)$$

The two inequalities in the first line must obviously hold. The first and third equalities follow from (21). The second equality follows from (20). As the second line in (23) holds for all $\tilde{m} \in \mathcal{M}$, the left-hand side of (22) cannot be bigger than the right-hand side.

Consider a zero-cost payoff that has a single positive payoff in state $s$ and a single negative payoff in state $\sigma$. This $\tilde{x}$ has a cost of zero by assumption so $\pi_\sigma m_s x_s = -\pi_\sigma m_\sigma x_\sigma$ and

$$\frac{\mathbb{E}^o[\tilde{x}^+] - \pi_\sigma m_\sigma x_\sigma}{\mathbb{E}^o[\tilde{x}^-] - \pi_\sigma m_\sigma x_\sigma} = \frac{\pi_\sigma m_s x_s}{\pi_\sigma m_\sigma x_\sigma} = \frac{\pi_\sigma m_s^o x_s}{\pi_\sigma m_\sigma^o x_\sigma} \frac{m_\sigma/m_\sigma^o}{m_s/m_s^o}. \quad (24)$$

This must be true for every $\tilde{m}$ in $\mathcal{M}$. And for each $\tilde{m}$, it must be true for every choice of the states $s$ and $\sigma$ so it must be true for the two states that maximize the numerator and minimize the denominator for that $\tilde{m}$. The GLR $^o\text{°}$ that maximizes the left-hand side must be at least as large as this specific one so

$$\max_{x \in \mathcal{X}} \frac{\mathbb{E}^o[\tilde{x}^+]}{\mathbb{E}^o[\tilde{x}^-]} \geq \max_s \left( \frac{m_s}{m_s^o} \right) \min_s \left( \frac{m_s^o}{m_s} \right) \forall \tilde{m} \in \mathcal{M}. \quad (25)$$

This proves that the right-hand side of (22) cannot be bigger than the left-hand side. As we have already seen that the opposite is true, both sides must be equal.

A special case of this theorem is a risk-neutral benchmark. The benchmark SDF, $m^o$ is constant, and $\pi^o = \pi$ so $\mathbb{E}^o[\cdot] = \mathbb{E}[\cdot]$. In this case (22) is

$$\max_{x \in \mathcal{X}} \frac{\mathbb{E}[\tilde{x}^+]}{\mathbb{E}[\tilde{x}^-]} = \min_{m \in \mathcal{M}} \max_s \left( \frac{m_s}{m_s^o} \right) \min_s \left( \frac{m_s^o}{m_s} \right). \quad (26)$$

So a bound on the GLR is a bound on the range of the SDF similar to Bernardo and Ledoit (1997) formulation. However, the GLR only bounds the logarithmic range without provided a specific lower and upper bound.

The major distinction between the GLR bound on one hand and the Good-Deal or Bernardo and Ledoit bound is that high or low values for the SDF are not inherently bad. They are only bad if they are high or low relative to the benchmark values in a given state. So, for example, if the SDF is generated by CRRA utility, $\tilde{m}^o$ ranges from 0 to $\infty$. That means that $\tilde{m}$ can as well, provided it is not too much larger or smaller than $\tilde{m}^o$ in any state.

The GLR can be used as a test of the accuracy of a given pricing assumption. If the GLR is 1 for all portfolios that can be formed from a set of assets, then the benchmark model is a complete representation of their pricing. If the GLR is much higher (or lower) than 1 for some subset of assets, those are the ones that the benchmark model does not price well. When the GLR is high, the right-hand side of (22) indicates which states are the problem. This indicates appropriate changes in the benchmark utility function. For example, if the benchmark is CRRA.
and the ratio \( m_s/m^2 \) is high for low values of \( m_s \), it indicates that the benchmark marginal utility declines too quickly with outcomes so decreasing relative risk aversion might provide a better fit.

The GLR provides a natural continuum between the loose no-arbitrage bounds and the specific prices of the benchmark. [To be finished]

General Properties of Portfolios with One Safe and One Risky Asset

Many models in Finance reduce to investors holding the market portfolio levered to one extent or another. This is true, for example, of representative investor models. The analysis in this section can be thought of in that light; namely, the comparative statics of the demand for the market portfolio.

It is assumed here that time-1 utility is state-independent and that time-0 consumption is fixed. That is, \( W_0 \) is to be interpreted here as the wealth remaining to be invested after time-0 consumption. In this analysis, portfolios are characterized by a single number, either the fraction, \( w \), of wealth invested in the risky asset or the dollar amount, \( D \equiv wW_0 \). Time-1 wealth is \( W_1 = W_0R_f + D\tilde{x} \), where \( R_f \equiv 1 + r_f \) is the gross return per dollar on the safe asset and \( \tilde{x} \) is the rate of return on the risky asset (market) in excess of the interest rate. In these terms the first-order condition of the previous chapter is

\[
0 = \frac{\partial u}{\partial D}_{D=D^*} = \mathbb{E}[u'(W_0R_f + D\tilde{x})]\tilde{x} = \mathbb{E}[u'(\tilde{W})\tilde{x}].
\]

(27)

It is useful to define the function \( H \)

\[
H(D,W_0,r_f,\tilde{x},F) \equiv \frac{\partial \mathbb{E}^F[u'(\tilde{W})]}{\partial D} = \frac{\partial \mathbb{E}^F[u'(W_0R_f + D\tilde{x})]}{\partial D} = \mathbb{E}^F[u'(\tilde{W})\tilde{x}].
\]

(28)

\( F \) is the cumulative distribution function of \( \tilde{x} \). This extra notation will be suppressed except when needed for clarity. The first-order condition for this simple problem (37) is

\[
H(D^*,\cdot) = \mathbb{E}[u'(\tilde{W}^*)\tilde{x}] = 0.
\]

(29)

Furthermore,

\[
H'(D^*,\cdot) = \mathbb{E}[u''(\tilde{W}^*)\tilde{x}^2] < 0,
\]

(30)

because utility is concave and \( \tilde{x}^2 > 0 \). Therefore, the global effects of changes in wealth, the interest rate, or the distribution of the risky asset can be determined by looking only at the local changes in the optimal investment, \( D^* \). For example, if the derivative of \( H \) with respect to \( r_f \) is negative when evaluated at \( D^* \) for the current \( r_f \), then the optimal holding of the risky asset must be smaller at all higher interest rates.

The first question to answer is under what conditions is the demand for the risky asset positive or negative. If the portfolio is invested only in the safe asset, \( H(0) = W_0u'(W_0R_f)\mathbb{E}[\tilde{x}] \). This is positive or negative depending only on the sign of \( \mathbb{E}[\tilde{x}] \). So when holding none of the risky asset, the investor’s utility would be increased by moving marginally towards a long position if the risky asset has a positive expected return rate. Similarly, when \( \mathbb{E}[\tilde{x}] < 0 \), a marginal move to a short position increases expected utility. As already shown these marginal improvements imply global improvements as well so any investor holds a long position in the risky asset when \( \mathbb{E}[\tilde{x}] > 0 \). Only this case is analyzed below. Most results are affected in the obvious way but it makes little sense to examine that case. If the only risky asset has a negative expected return rate, then every investor would want a negative position in it so the asset would not exist in equilibrium.

---

13 Throughout this section \( \tilde{x} \) denotes a rate of return in excess of the interest rather than a payoff.
There are two exceptions to this general result for a positive demand. As noted in Chapter 2, a utility function with first-order risk aversion is not differentiable at the reference point so an investor with a reference point at $W_r$ might not care to own even a little bit of a risky prospect. Also an investor with utility function that can achieve $-\infty$ will refrain from any risk that could result in such an outcome. For example, an investor with logarithmic utility would not take even a small position long or short in the risky asset whose return is normally distributed regardless of the expected rate of return or variance. Equations (37) and (39) do not hold, because expected utility is not defined for this utility when holding any of the risky asset. These two cases are excluded in all the discussion below though some examples do use utility functions that are unbounded below.

Arrow (1970) proved that decreasing (or increasing) risk aversion over the entire relevant range implies that the risky asset is a normal (or inferior) good; that is, as wealth increases, more (or less) of it is purchased. Applying the implicit function theorem to $H$ gives the comparative static effect of a change in wealth

$$\frac{\partial D^*}{\partial W_0} = -\frac{\partial H(D^*; \cdot)}{\partial D} \frac{\partial\bar{\bar{W}}(D^*; \cdot)}{\partial W_0} = -\mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}] R_f,$$

(31)

The denominator is negative so the sign of the effect is the same as the sign of the numerator. Under DARA

$$A(\bar{\bar{W}}^*) \leq A(W_r) \quad \text{for} \quad \bar{x} \geq 0,$$

$$A(\bar{\bar{W}}^*) \geq A(W_r) \quad \text{for} \quad \bar{x} \leq 0.$$

(32)

Therefore, in each case multiplying by $-u'(\bar{\bar{W}}^*) \bar{x}$ gives

$$-u'(\bar{\bar{W}}^*) \bar{x} A(\bar{\bar{W}}^*) = u^*(\bar{\bar{W}}^*) \bar{x} > -A(W_r) R_f u'(\bar{\bar{W}}^*) \bar{x}.$$

(33)

The inequalities are both the same because the first line in (42) is multiplied by a negative quantity and the second line is multiplied by a positive one. Therefore,

$$\mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}] \geq -A(W_r) R_f \mathbb{E}[u'(\bar{\bar{W}}^*)\bar{x}] = -A(W_r) R_f H(D^*) = 0.$$

(34)

This means that the demand for the risky asset increases with wealth under DARA; that is, it is a normal good. The proof that the risky asset is an inferior good under IARA is identical with a change of inequalities.

The difference between the wealth elasticity of demand for the risky asset and 1 can be used to determine how the portfolio weight $w^*$ changes.

$$\frac{dw^*}{dW_0} = \frac{1}{W_0} \frac{dD^*/dW_0}{dW_0} = D^* \left( \frac{W_0}{D^*} \frac{dD^*/dW_0}{dW_0} - 1 \right).$$

(35)

So the portfolio weight increases (or decreases) with wealth if the elasticity is greater (or less) than 1. The wealth elasticity for the risk asset is

$$\frac{dD^*/dW_0}{dD^*} = \frac{W_0(dD^*/dW_0) - D^*}{dD^*} = \mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}] R_f W_0 + D^* \mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2]$$

$$= \frac{-D^* \mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2]}{-D^* \mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2]} = \frac{\mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2]}{-D^* \mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2]} - \frac{\mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2]}{-D^* \mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2]} = \mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2] - \mathbb{E}[u^*(\bar{\bar{W}}^*)\bar{x}^2].$$

(36)

where $\text{RRA}(W)$ is the Arrow-Pratt measure of relative risk aversion. Again the denominator is positive so the final ratio is positive or negative depending on the sign of the numerator. The
same analysis used in equations (42) through (44) shows that the elasticity exceeds 1 so that the proportional portfolio increases with wealth if and only if relative risk aversion is decreasing.

Changes in the investment opportunities also affect demand. We might suspect that an increase in the interest rate would increase demand for the safe asset. Similarly, an increase in the expected return on the risky asset might be presumed to increase the demand for it. However, the results are not this simple. Changes in the returns have two opposing types of effects. The one just described is essentially the substitution effect from standard demand theory—namely, the demand for a good increases as its price falls or the price of the other good rises. Here an increase in an expected rate of return is a decrease in the price of a future payoff. But there is also an effect equivalent to the income effect. Current wealth is not changed, but an increase in either expected rate of return increases time-1 wealth making more consumption available on average. This affects the allocation between first and second period consumption were we to be considering such, but this increase in time-1 “income” can also shift demand between the assets.

The comparative static effect of a change in the interest rate, leaving the distribution of the excess return constant is

\[
\frac{\partial D^*}{\partial r_f} = - \frac{\partial H(D^*; \cdot)}{\partial r_f} = \frac{\partial \mathbb{E}[u'(\tilde{W}^*)(\tilde{r} - r_f)]/\partial r_f}{-\mathbb{E}[u'(\tilde{W}^*)]}. \tag{37}
\]

The denominator is positive everywhere so the sign of the effect is the same as the sign of the derivative in the numerator.

\[
\frac{\partial \mathbb{E}[u'(\tilde{W}^*)(\tilde{r} - r_f)]}{\partial r_f} = \mathbb{E}[u'(\tilde{W}^*)(W_0 - D^*)\tilde{x}] - \mathbb{E}[u'(\tilde{W}^*)]. \tag{38}
\]

The second term, the pure interest rate partial derivative, is the cross substitution effect and is negative as usual. An increase in the price of the safe asset (a decrease in the interest rate) leads to an increase in the demand for the risky asset which is a decrease in the demand for the safe asset.

The first term, the interest rate effect that comes through a change in \( \tilde{W}^* \), is an income-like effect and can have either sign. Substituting \( u'(\tilde{W}^*) = -A(\tilde{W}^*)u'(\tilde{W}^*) \), the “income” effect is

\[
-(W_0 - D^*)\mathbb{E}[A(\tilde{W}^*)u'(\tilde{W}^*)\tilde{x}] \tag{39}
\]

If absolute risk aversion is constant, then this income effect is zero because the expectation is proportional to the first-order condition (39). The income effect is also zero if the investor is currently fully invested in the safe asset, \( D^* = W_0 \).

If absolute risk aversion is decreasing, which is the typical case, then \( A(\tilde{W}^*) \leq A(W_0R_f) \) for \( \tilde{x} \geq 0 \). The expectation in (49) is then

\[
\mathbb{E}[A(\tilde{W}^*)u'(\tilde{W}^*)\tilde{x}] < A(W_0R_f)\mathbb{E}[u'(\tilde{W}^*)\tilde{x}] = 0. \tag{40}
\]

The inequality follows because the positive values of \( u'(\tilde{W}^*)\tilde{x} \) have been weighted by \( A(W_0R_f) > A(\tilde{W}^*) \) while the negative values have been given less weight in the expectation on the left-hand side. The final expectation is zero because it is the first order condition. So the expectation in (49) is negative under DARA, and the income effect leads the investor to increase his holding of the risky asset at a higher interest rate rises provided \( D^* < W_0 \). If \( D^* > W_0 \), the opposite is true. That is, when the interest rate is higher, the optimal holding of the risky asset is closer to 100% of wealth. With increasing absolute risk aversion, the opposite is true.

Changes in the distribution of the risky asset’s return also affect the optimal portfolio, but there are many possible changes. The simplest change is a translation of the entire distribution.
keeping its shape the same so the expectation, $\bar{r}$, increases but all central moments are unchanged. For this change in the distribution, the change in the optimal portfolio holding is
\begin{equation}
\frac{\partial D^*}{\partial \bar{r}}|_{\bar{r}, \bar{\chi}} = -\frac{\partial H(D^*; \cdot)}{\partial \bar{r}} = \mathbb{E}[u'(\bar{W}^*)\bar{\chi}D^*] + \mathbb{E}[u'(\bar{W}^*)] - \mathbb{E}[u''(\bar{W}^*)\bar{\chi}^2].
\end{equation}

The second term is again the substitution effect. The translated return lowers the time-0 price of time-1 wealth leading to an increase in demand for the risky asset. The first term is the income-like effect. It can be analyzed as in (50). An increase in the risky asset’s expected rate of return leads to an increase in its demand whenever risk aversion is constant or decreasing. Recall that the asset being held long, $D^* > 0$, is a maintained assumption.

It might seem that a stronger result should be true, namely that any change in the distribution of the return on the risky asset that is first-order stochastically dominating should increase its demand when the investor has constant or decreasing risk aversion, but this is not correct. One simple counterexample was shown in the previous chapter. The increase of the better of two options can lead to a decrease in the amount allocated to the risky asset in order to improve the outcome in the poor state.

A first-order dominating change in $\bar{r}$ does increase the demand for the risky asset for any risk-averse investor, not just a DARA investor, when the only changes in the distribution are increases in the outcomes below the risk-free rate, $r_f$.

Let the two distributions of excess returns be $F$ and $G$ with $F$ the dominating distribution satisfying $F(x) = G(x)$ for $x > 0$ and $F(x) \leq G(x)$ for $x \leq 0$. So $F$ dominates $G$ due only to superior performance in returns below the risk-free rate; the distribution of their (excess) gains are the same. If the derivative of expected utility under the $F$ distribution evaluated at the optimal portfolio holding under the $G$ distribution is positive, then the investor holds more of the risky asset in the portfolio optimized for the $F$ distribution.

\begin{equation}
\mathbb{E}^F[u(\bar{W})] = \mathbb{E}^G[u'(W_0 R_f + D^G \bar{\chi})\bar{\chi}] - \mathbb{E}^G[u'(W_0 R_f + D^G \bar{\chi})\bar{\chi}] = 0 \text{ by first-order condition}
\end{equation}

\begin{equation}
= \int_{-\infty}^{\infty} u'(W_0 R_f + D^G \bar{\chi})[dF(x) - dG(x)].
\end{equation}

The second expectation can be added because it is zero by the first-order condition under the $G$ distribution. Integrating by parts and using $F(\chi) = G(\chi)$ for $x > 0$

\begin{equation}
\frac{\partial \mathbb{E}^F[u(\bar{W})]}{\partial D} = 0 - W_0 \int_{-\infty}^{0} [u'(W^G) + u''(W^G) D^G x][F(x) - G(x)]dx
\end{equation}

\begin{equation}
= W_0 \int_{-\infty}^{0} u'(W^G)[1 - A(W^G) D^G x][G(x) - F(x)]dx \geq 0
\end{equation}

where $W^G = W_0 R_f + D^G x$. The first term of the parts integral is zero because $F(\infty) = G(\infty) = 1$ and $F(-\infty) = G(-\infty) = 0$. The final expression is positive because $u'$, $D^G$, $A$ and $G - F$ are positive while $x$ is negative.

The holding of the risky asset is also larger under an $F$ distribution that likelihood-ratio dominates the $G$ distribution. Recall that this type of domination means that the ratio of the densities, $f(x)/g(x)$, is increasing in $x$, which is a stronger condition than first-order stochastic dominance. The expression in (52) is
Without the final factor in brackets, the integral would be the first-order condition under the $G$ distribution and be equal to zero. As the ratio $f(x)/g(x)$ is increasing in $x$, the actual integral places more weight on the larger values and is positive.

The effect of an increase in risk is also ambiguous in sign. Suppose distribution $F$ second order stochastically dominates distribution $G$. The dollar holding is smaller under the $G$ distribution if

$$H(D^F; W_0, r_f, \tilde{x}, G) = \mathbb{E}^G[u'(W_0 R_f + D^F \tilde{x})] \leq 0. \quad (45)$$

Define $\psi(x) \equiv u'(W_0 R_f + D^F x) x = u'(W^F)x$. Then (55) will hold by Jensen's inequality for every dominated $G$ distribution if and only if $\psi$ is a concave function. The first and second derivatives of $\psi$ are

$$\psi'(x) = [u'(W^F) + u''(W^F)D^F x]$$

$$\psi''(x) = D^F [2u''(W^F) + u'''(W^F)D^F x]. \quad (46)$$

Recall that the measure of prudence is $P(W) \equiv -u''(W)/u'(W)$. So $\psi$ is concave and $D^G \leq D^F$ if and only if

$$P(W^F(x)) \geq \frac{2}{D^F x} \quad \forall \xi. \quad (47)$$

This result has little intuitive content; manipulating (56) gives a sufficient condition that is easier to understand. By assumption $D^F$ is positive so if $\psi$ is to be concave everywhere $\psi''(x)/D^F$ must be negative

$$\psi''(x)/D^F = 2u''(W^F) + u'''(W^F)D^F x = 2u''(W^F) + u'''(W^F)(W^F - W_0 R_f)$$

$$= u''(W^F) + \frac{[u''(W^F)]^2 W^F}{u'(W^F)} - \frac{2u''(W^F)W_0}{u'(W^F)}\left[\frac{[u''(W^F)]^2 W^F}{u'(W^F)} - u'''(W^F)W_0^2 r_f\right] \quad (48)$$

Because

$$A'(W) = \frac{-u''}{u'} + \frac{u''^2}{u'^2} \quad \text{and} \quad \text{RRA}'(W) = A + WA' = -\frac{u''}{u'} + W\left(\frac{-u''}{u'} + \frac{u''^2}{u'^2}\right) \quad (49)$$

this can be re-expressed as

$$\frac{\psi''(x)}{D^F} = [1 - \text{RRA}(W^F) + A(W^F)W_0]\mu''(W^F)$$

$$+ [A'(W^F)W_0 - \text{RRA}'(W^F)]\mu'(W^F) - u'''(W^F)W_0^2 r_f. \quad (50)$$

where $A(W)$ and $R(W)$ are the Arrow-Pratt measures of absolute and relative risk aversion. So if relative risk aversion is less than one and increasing and absolute risk aversion is decreasing (which also guarantees that $u''' > 0$), $\psi$ must be concave, and a second-degree stochastically dominating improvement in the distribution will increase the holding of the risky asset.
Further Notes

Lower Bounds on Higher Moments: Bounds like the Hansen-Jagannathan limit on the variance of the SDF can be determined for higher moments as well. If the Law of One-Price holds, then for any portfolio \( w \),

\[
1 = \mathbb{E}[\tilde{m}(1 + \tilde{r}_w)] \leq \mathbb{E}[\tilde{m}^{\delta}]^{1/\delta} \cdot \mathbb{E}[(1 + \tilde{r}_w)^{\rho}]^{1/\rho} .
\]

The inequality is an application of Hölder's inequality. It is valid for any strictly positive random variables provided the expectations exist and \( \delta^{-1} + \rho^{-1} = 1 \). Rearranging terms gives

\[
\mathbb{E}[\tilde{m}^{\delta}]^{1/\delta} \geq \mathbb{E}[(1 + \tilde{r}_w)^{\rho}]^{1/\rho} .
\]

The SDF, \( \tilde{m} \), is strictly positive by the Law of One Price so only the restriction \( 1 + \tilde{r}_w > 0 \) need be applied. This must hold for all portfolios with strictly positive payoffs so it must hold for the highest such value.

The Growth Optimal Lower Bound on the SDF

Another useful bound on the SDF can be constructed from the growth optimal portfolio.

**Theorem 8.4 Growth-Optimal Bound.** If there no arbitrage opportunities, then the SDF is bounded by

\[
\mathbb{E}[\ell n \tilde{m}] \geq \mathbb{E}[\ell n(1 + \tilde{r}_{GO})] \quad (53)
\]

where \( \tilde{r}_{GO} \) is the rate of return on the growth-optimal portfolio, which is the optimal portfolio for an investor with logarithmic utility.

**Proof:** By the Law of One Price guarantees the SDF, \( \tilde{m} \), prices any portfolio of assets as \( \mathbb{E}[\tilde{m}(1 + w'\tilde{r})] = 1, \forall w \). So

\[
\tilde{m}(1 + w'\tilde{r}) = 1 + \tilde{e} \quad \text{with} \quad \mathbb{E}[\tilde{e}] = 0 \quad \forall w .
\]

Taking logs and the expectation of both sides

\[
\mathbb{E}[\ell n \tilde{m}] + \mathbb{E}[\ell n(1 + w'\tilde{r})] = \mathbb{E}[\ell n(1 + \tilde{e})] \leq \ell n(\mathbb{E}[1 + \tilde{e}]) = 0 .
\]

The weak inequality follows from Jensen's Inequality. The SDF is positive by the absence of arbitrage so \( \ell n \tilde{m} \) is defined. The second term is defined for all portfolios whose returns have limited liability so that \( 1 + w'\tilde{r} > 0 \), and its logarithm is defined. The tightest possible bound is

\[
-\mathbb{E}[\ell n \tilde{m}] \geq \max_w \mathbb{E}[\ell n(1 + w'\tilde{r})] \quad \text{subject to} \quad w'w = 1, w'\tilde{r} > -1
\]

\[
\equiv \mathbb{E}[\ell n(1 + \tilde{r}_{GO})] .
\]

The growth optimal portfolio always has a positive payoff because the log investor never risks a complete loss resulting in utility of \( -\infty \).

The log utility portfolio is the one with the highest compound growth rate. So equation (66) says that the reciprocal of the marginal rate of substitution must grow at least as fast as the highest possible growth rate of wealth. This interpretation is quite different from that of the Hansen-Jagannathan bound. However, they can be related.

A SDF is marginal utility, and the marginal utility for the log utility is \( \tilde{m} = (1 + \tilde{r}_{GO})^{-1} \), so \( \mathbb{E}[\tilde{m}] = \mathbb{E}[(1+\tilde{r}_{GO})^{-1}] \). These can be added to the left- and right-hand sides of (66), respectively

\[
\ell n \mathbb{E}[\tilde{m}] - \mathbb{E}[\ell n \tilde{m}] \geq -\mathbb{E}[\ell n ((1+\tilde{r}_{GO})^{-1})] + \ell n \mathbb{E}[(1+\tilde{r}_{GO})^{-1}] .
\]

(57)
If both $\tilde{m}$ and $1+\tilde{r}_{GO}$ are lognormally distributed, this simplifies to $\text{var}[\ell n \tilde{m}] \geq \text{var}[\ell n(1 + \tilde{r}_{GO})]$, which is the Hansen-Jagannathan lower bound in logs.

The expectation of the SDF is also $\mathbb{E}[\tilde{m}] = (1 + \tilde{r})^{-1}$. So adding this to both sides of (66)

$$
\mathbb{E}[\ell n (1 + \tilde{r}_{GO})] - \ell n (1 + \tilde{r}) \leq -\mathbb{E}[\ell n \tilde{m}] + \ell n \mathbb{E}[\tilde{m}] = -\mathbb{E}[\ell n (\tilde{m}/\mathbb{E}[\tilde{m}])] = -\mathbb{E}[\ell n (1 + \tilde{m}/\mathbb{E}[\tilde{m}]-1)].
$$

(58)

Using a Taylor expansion

$$
\mathbb{E}[\ell n (1 + \tilde{r}_{GO})] - \ell n (1 + \tilde{r}) \leq \mathbb{E} \left[ \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left( \frac{\tilde{m}}{\mathbb{E}[\tilde{m}]} - 1 \right)^n \right].
$$

(59)

As the expression in parentheses obviously has an expectation of zero, a lower bound has been constructed in which small odd moments and large even moments of $\tilde{m}$ facilitate meeting the bound. That is, negatively skewed SDF’s more easily achieve the bound than positively skewed SDF’s. Similarly a larger kurtosis assists reaching the bound. Market crashes produce a negative skewness in returns and therefore a positive skewness in marginal utility which is the SDF so finding a SDF that satisfies the bound is more difficult when crashes are likely or large.

Cherny & Madan

MPPM