Arbitrage Pricing Theory

Ross ([1],[2]) presents the *arbitrage pricing theory*. The idea is that the structure of asset returns leads naturally to a model of risk premia, for otherwise there would exist an opportunity for arbitrage profit.

Factor Model

Assume that there exists a risk-free asset, and consider a factor model for the excess return ξ on a set of assets:

$$\xi = \boldsymbol{m} + B^\top \boldsymbol{f} + \boldsymbol{e}.$$

The mean excess return $\boldsymbol{m}$ is the vector of risk premia.

Here

$$\mathrm{E}(\boldsymbol{f}) = \mathbf{0}$$

$$\mathrm{Var}(\boldsymbol{f}) = \mathsf{I},$$

and

$$\mathrm{E}(\boldsymbol{e}) = \mathbf{0}$$

$$\mathrm{Var}(\boldsymbol{e}) = D.$$

Here D is diagonal, and $\boldsymbol{f}$ and $\boldsymbol{e}$ are uncorrelated. One refers to $\boldsymbol{f}$ as *factors*. The beta coefficients B are also called *factor loadings*.

Systematic Versus Non-Systematic Risk

Assume that most of the components of B are not near zero. The diagonal elements of D are not too large, and the number of assets n is large.

Then the term $B^\top \boldsymbol{f}$ represents most of the variation in the returns. Interpret $B^\top \boldsymbol{f}$ as the systematic risk, and $\boldsymbol{e}$ as the non-systematic risk. One can argue that the non-systematic risk can be eliminated by diversification, so the beta coefficients B should determine the risk premium.

Intuitive Argument

Ross gives the following intuitive argument. Consider a portfolio $\boldsymbol{x}$. Each component denotes the fraction of wealth invested in that asset, and $1-\mathbf{1}^\top\boldsymbol{x}$ is the fraction invested in the risk-free asset. The excess return on the portfolio is

$$\xi^\top\boldsymbol{x} = \boldsymbol{m}^\top\boldsymbol{x} + \boldsymbol{f}^\top B\boldsymbol{x} + \boldsymbol{e}^\top\boldsymbol{x}.$$

Suppose that the portfolio is well-diversified: most of the components of $\boldsymbol{x}$ are non-zero. By the law of large numbers,

$$\boldsymbol{e}^\top \boldsymbol{x} \approx 0;$$

diversification eliminates the non-systematic risk. If the portfolio is chosen to eliminate the systematic risk ($B\boldsymbol{x} = \mathbf{0}$), then the resulting portfolio is nearly risk-free. Then the law of one price implies $\boldsymbol{m}^\top \boldsymbol{x} = 0$.

Ross summarizes his argument by the following:

$$B\boldsymbol{x} = \mathbf{0} \Rightarrow \boldsymbol{m}^\top \boldsymbol{x} = 0. \quad (1)$$

(Of course this argument is not valid for an arbitrary portfolio but only for a well-diversified portfolio.)

Exact Factor Model

Consider first an exact factor model, in which $\boldsymbol{e} = \mathbf{0}$ (so $D = 0$).

Remark 1 *In the exact factor model, the law of one price is equivalent to the condition (1).*

Law of One Price

For the exact factor model, the law of one price (1) says that $\boldsymbol{m}$ is orthogonal to $\mathrm{N}(B)$. By the fundamental theorem of linear algebra, $\boldsymbol{m}$ must lie in $\mathrm{R}\left(B^\top\right)$. Thus we obtain the following theorem.

Theorem 2 (Arbitrage Pricing Theory) *In the exact factor model, the law of one price holds if only if the mean excess return is a linear combination of the beta coefficients,*

$$\boldsymbol{m} = B^\top \boldsymbol{b}, \tag{2}$$

for some $\boldsymbol{b}$.

The Arbitrage Pricing Theory Versus the Capital-Asset Pricing Model

Like the capital-asset pricing model, the systematic risk embodied in the beta coefficients determines the risk premia. However the reasoning is different.

The capital-asset pricing model is derived from market equilibrium, the equality of asset demand and supply. This equality implies that the market portfolio must be efficient, and a typical investor holds the market portfolio.

In contrast, the arbitrage pricing theory is derived from an arbitrage argument, not a market equilibrium argument. The risk premia (2) follow from the factor structure of the asset returns. Asset supply is irrelevant to the argument. If some set of asset returns has the factor structure, then the conclusion follows for this set.

We next suppose that the factor model is not exact, that $\boldsymbol{e} \neq \mathbf{0}$.

Then *any* value for $\boldsymbol{m}$ is consistent with the law of one price (the only portfolio with a constant excess return is $\boldsymbol{x} = \mathbf{0}$).

Nevertheless we put forward a duality argument that (2) is a good approximation.

Weighted Least Squares

We choose $\boldsymbol{b}$ by weighted least squares.

Problem 3 (Primal)

$$\min_{\boldsymbol{b}} \left[\frac{1}{2} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right)^\top D^{-1} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right) \right].$$

The primal is a weighted regression of the mean on the beta coefficients.

The argument is that the value is small, for the optimum $\boldsymbol{b}$.

Dual

An alternate maximization problem is dual to the primal.

Problem 4 (Dual)

$$\sup_{\boldsymbol{m}^*} \left(\boldsymbol{m}^{*\top} \boldsymbol{m} - \frac{1}{2} \boldsymbol{m}^{*\top} D \boldsymbol{m}^* - \delta_{B\boldsymbol{m}^*=\mathbf{0}} \right).$$

By definition, the indicator function δ is zero if the $\boldsymbol{m}^*$ belongs to the set such that $B\boldsymbol{m}^* = \mathbf{0}$, and is ∞ otherwise.

The primal and the dual are equivalent problems, in that either one can be calculated from the other, and we explain their relationship.

The objective function of the primal is jointly convex in $\boldsymbol{b}$ and $\boldsymbol{m}$, and it follows that the value function $V(\boldsymbol{m})$ is convex. The choice variable is $\boldsymbol{b}$, and the perturbation variable is $\boldsymbol{m}$.

Conjugate

Definition 5 (Conjugate) *The conjugate of* $V(\boldsymbol{m})$ *is*

$$V^*(\boldsymbol{m}^*) := \sup_{\boldsymbol{m}} \left[\langle \boldsymbol{m}^*, \boldsymbol{m} \rangle - V(\boldsymbol{m}) \right].$$

The conjugate is a convex function.

Conjugate Duality

Proposition 6 (Conjugate Duality) *Under general conditions, the conjugate of the conjugate is the original function,*

$$V^{**}(\boldsymbol{m}) = V(\boldsymbol{m}).$$

In mathematics, *conjugate* has many definitions, but always the conjugate of the conjugate is the original; an example is the complex conjugate.

Dual

Definition 7 (Dual) *For a primal with value function $V(\boldsymbol{m})$, the dual is the maximization problem*

$$\sup_{\boldsymbol{m}^*} \left[\langle \boldsymbol{m}^*, \boldsymbol{m} \rangle - V^*(\boldsymbol{m}^*) \right].$$

By conjugate duality, the optimum value in the dual is $V(\boldsymbol{m})$.

Theorem 8 (No Duality Gap) *The minimum value in the primal is the maximum value in the dual.*

Calculation of the Dual from the Primal

Let us derive the dual problem (4) by calculating the conjugate:

$$
\begin{aligned}
&V^*(\boldsymbol{m}^*) \\
&\quad = \sup_{\boldsymbol{m}} \left[\langle \boldsymbol{m}^*, \boldsymbol{m} \rangle - V(\boldsymbol{m}) \right] \\
&\quad = \sup_{\boldsymbol{m}} \left\{ \langle \boldsymbol{m}^*, \boldsymbol{m} \rangle - \min_{\boldsymbol{b}} \left[\frac{1}{2} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right)^\top D^{-1} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right) \right] \right\} \\
&\quad = \sup_{\boldsymbol{m}} \left\{ \langle \boldsymbol{m}^*, \boldsymbol{m} \rangle + \sup_{\boldsymbol{b}} \left[-\frac{1}{2} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right)^\top D^{-1} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right) \right] \right\} \\
&\quad = \sup_{\boldsymbol{b}, \boldsymbol{m}} \left[\langle \boldsymbol{m}^*, \boldsymbol{m} \rangle - \frac{1}{2} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right)^\top D^{-1} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right) \right]
\end{aligned}
$$

Substituting $\boldsymbol{c} := \boldsymbol{m} - B^\top \boldsymbol{b}$ separates the maximization into two parts:

$$\begin{aligned}
&V^*(\boldsymbol{m}^*) \\
&= \sup_{\boldsymbol{b},\boldsymbol{c}} \left(\left\langle \boldsymbol{m}^*, \boldsymbol{c} + B^\top \boldsymbol{b} \right\rangle - \frac{1}{2} \boldsymbol{c}^\top D^{-1} \boldsymbol{c} \right) \\
&= \sup_{\boldsymbol{c}} \left(\langle \boldsymbol{m}^*, \boldsymbol{c} \rangle - \frac{1}{2} \boldsymbol{c}^\top D^{-1} \boldsymbol{c} \right) + \sup_{\boldsymbol{b}} \langle B\boldsymbol{m}^*, \boldsymbol{b} \rangle \\
&= \frac{1}{2} \boldsymbol{m}^{*\top} D \boldsymbol{m}^* + \delta_{B\boldsymbol{m}^* = \boldsymbol{0}},
\end{aligned}$$

obtaining the dual problem (4).

Primal Greater than or Equal to the Dual

For this problem, let us verify directly the basic duality properties. Always the value of the primal is greater than or equal to the value of the dual.

If $B\boldsymbol{m}^* \neq \mathbf{0}$, then $\delta_{B\boldsymbol{m}^*=\mathbf{0}} = \infty$, so the value of the dual is $-\infty$. Necessarily the value of the primal is greater than or equal to the value of the dual.

If $B\boldsymbol{m}^* = \mathbf{0}$, then the primal less the dual is

$$\frac{1}{2}\left(\boldsymbol{m} - B^\top \boldsymbol{b}\right)^\top D^{-1}\left(\boldsymbol{m} - B^\top \boldsymbol{b}\right) - \left[\langle \boldsymbol{m}^*, \boldsymbol{m}\rangle - \frac{1}{2}\boldsymbol{m}^{*\top} D \boldsymbol{m}^*\right]$$
$$= \frac{1}{2}\left[D^{-1}\left(\boldsymbol{m} - B^\top \boldsymbol{b}\right) - \boldsymbol{m}^*\right]^\top D \left[D^{-1}\left(\boldsymbol{m} - B^\top \boldsymbol{b}\right) - \boldsymbol{m}^*\right] \geq 0 \tag{3}$$

a sum of squares. Here

$$\left\langle \boldsymbol{m}^*, B^\top \boldsymbol{b}\right\rangle = \langle B\boldsymbol{m}^*, \boldsymbol{b}\rangle = \langle \mathbf{0}, \boldsymbol{b}\rangle = 0.$$

Again, the value of the primal is greater than or equal to the value of the dual.

Primal Equal to the Dual

In the primal, the first-order condition for a minimum is

$$BD^{-1}\left(\boldsymbol{m}-B^{\top}\boldsymbol{b}\right)=\mathbf{0}.$$

Given the solution $\boldsymbol{b}$ to the primal, solve the dual by setting

$$\boldsymbol{m}^{*}=D^{-1}\left(\boldsymbol{m}-B^{\top}\boldsymbol{b}\right).$$

By the first-order condition, $B\boldsymbol{m}^{*}=\mathbf{0}$, so the dual constraint is satisfied. Furthermore, the quadratic form (3) is zero. That the value of the primal equals the value of the dual proves that indeed we have the solution to both problems.

Envelope Theorem

Applying the envelope theorem to the primal yields

$$\partial V(\boldsymbol{m}) / \partial \boldsymbol{m} = D^{-1}\left(\boldsymbol{m} - B^{\top}\boldsymbol{b}\right) = \boldsymbol{m}^{*},$$

in which $\boldsymbol{b}$ is the solution to the primal and $\boldsymbol{m}^{*}$ is the solution to the dual.

This relationship is a general duality result: the solution to the dual shows how the perturbation variable affects the optimum value. The solution to the dual is a Lagrange multiplier.

First-Order Condition Obsolete

Even though this verification makes use of the first-order condition, a theme of duality theory is that the first-order condition is obsolete. Because there is no duality gap, one can solve the primal and the dual simultaneously, by setting the primal equal to the dual. A systematic procedure then finds the optimum values for the choice variables in the primal and the dual, to achieve this equality.

Economic Interpretation of the Dual

The dual has an economic interpretation. The choice variable $\boldsymbol{m}^*$ is a portfolio of investments in the risky assets, with $1 - \boldsymbol{m}^{*\top}\mathbf{1}$ as the investment in the risk-free asset.

The constraint $B\boldsymbol{m}^* = \mathbf{0}$ says that the portfolio is chosen to be uncorrelated with the factors; the variability of the return arises solely from the non-systematic risk.

Thus $\boldsymbol{m}^{*\top}\boldsymbol{m}$ is the excess return on the portfolio, and $\boldsymbol{m}^{*\top}D\boldsymbol{m}^*$ is the variance of the return.

The dual resembles the derivation of the separation theorem with small risks, in which the objective function is a linear function of the mean excess return and the variance.

Just as for the separation theorem, the solution to the dual maximizes the ratio of the mean excess return to the standard deviation,

$$\frac{\boldsymbol{m}^{*\top}\boldsymbol{m}}{\sqrt{\boldsymbol{m}^{*\top}\boldsymbol{D}\boldsymbol{m}^{*}}}, \tag{4}$$

here subject to the constraint that the portfolio return is uncorrelated with the factors. Furthermore, the square of the maximum value of this ratio is the optimum value of the dual.

Efficient Frontier

Define $\tilde{s}$ as the maximum value of the ratio (4). It is the slope of an "efficient frontier," subject to the constraint that the portfolio return is uncorrelated with the factors.

The slope s of the efficient frontier is of course greater than or equal to $\tilde{s}$.

Upper Bound to the Weighted Sum of Squares

As there is no duality gap,

$$\min_{\boldsymbol{b}} \left[\frac{1}{2} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right)^\top D^{-1} \left(\boldsymbol{m} - B^\top \boldsymbol{b} \right) \right] = \widetilde{s}^2 \leq s^2. \qquad (5)$$

Thus the slope s of the efficient frontier provides an upper bound to the weighted sum of squares. The slope provides an upper bound to how far the predicted mean $B^\top \boldsymbol{b}$ can deviate from the actual mean $\boldsymbol{m}$.

Many Assets

A key property is that this upper bound is independent of the number n of assets.

Conclusion 9 (Arbitrage Pricing Theory) *If the number of assets is large, it follows that*

$$B^\top \boldsymbol{b} \approx \boldsymbol{m} \tag{6}$$

for most assets.

Otherwise the upper bound would be violated.

The approximation may be poor for a *few* assets, but for *most* assets the approximation must be excellent.

Trivial Case

Note that this conclusion holds even for the trivial case $B = 0$, for which $B^\top \boldsymbol{b} = 0$.

Then the duality relation (5) says that

$$\frac{1}{2}\boldsymbol{m}^\top D^{-1}\boldsymbol{m} = \widetilde{s}^2 \leq s^2,$$

so

$$\boldsymbol{m} \approx \boldsymbol{0}$$

is a good approximation. For *most* assets, the mean excess return is near zero.

Irrelevance of Non-Systematic Risk?

Ross's point of view is that the error $\boldsymbol{e}$ is non-systematic risk, and this risk should be eliminated by portfolio diversification. Hence the non-systematic risk should have no effect on mean returns. *If* this point of view is true, then $\tilde{s}$ should be small, small even if s is large.

By the duality relation (5), it would then follow that the approximation (6) would be extremely good.

Diagonal Variance

That the variance of the error $\boldsymbol{e}$ is diagonal is important, and allows one to see the error as non-systematic risk. The duality relation (5) holds regardless of whether D is in fact diagonal.

If the components of $\boldsymbol{e}$ were highly correlated, then the approximation (6) might be poor for *many* assets. For example, for the trivial case $B = 0$, there is no presumption that most of the components of $\boldsymbol{m}$ should be near zero.

References

[1] S. Ross. Return, risk, and arbitrage. In I. Friend and J. L. Bicksler, editors, *Risk and Return in Finance*, pages 189–218. Ballinger, Cambridge, MA, 1977. HG4539R57.

[2] S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, December 1976. HB1J645.