Microeconomic Theory
Lecture 11: Uncertainty, Time, and Asset Markets

Marcel Fafchamps
Oxford University

Readings: Varian, Chapters 11, 19, and 20.
Warning:
The material presented in this lecture is more difficult than that presented in earlier lectures. Minimum familiarity with probability theory is necessary to follow this lecture – i.e., probability distribution, expectation, mean, variance, and simple operations on mean and variance. It is also important to be comfortable with summations and indices.

1. Uncertainty

In this lecture we focus on risk and assets. We begin by defining expected utility and then turn to intertemporal preferences. Risk is very important for poor people, especially in rural areas. Being self-employed, people are subject to business risk. They also face more health risk, both because parasitic diseases are more prevalent and because the health infrastructure is weak. A proper understanding of risk is essential to development economists. How the poor deal with risk has received much attention in research and policy circles over the decade or two.

1.1. Expected utility

Economists have a favorite framework with which they model both aversion toward risk and decision under risk. This framework is called the expected utility framework. It is due to Von Neuman and Morgenstern. For those familiar with Bayesian statistical analysis, the framework has a strong Bayesian feel.

The issue we have to address is the following: how to express preferences among different risky outcomes. Suppose consumers are faced with a choice between two states of the world. This choice may result from the decision to work or not to work, to purchase a tractor or not, etc. To each state of the world is associated a distribution of consumption, that is, a vector of possible consumptions $\{c_1, c_2, ..., c_N\}$ together with a set of probabilities associated with these outcomes $\{p_1, p_2, ..., p_N\}$. Probabilities must sum to one: $\sum_{i=1}^{N} p_i = 1$. Let $p(c)$ denote the whole thing, that is, the possible consumptions with the probabilities.

The consumer has to choose among two arbitrary distributions $p_a(c)$ and $p_b(c)$, that is, between $\{c_1^a, c_2^a, ..., c_N^a\}$ with associated $\{p_1^a, p_2^a, ..., p_N^a\}$. Expected utility theory assumes that the consumer's preference between $p_a(c)$ and $p_b(c)$ can be represented by:

$$U(c_1^a, c_2^a, ..., c_N^a; p_1^a, p_2^a, ..., p_N^a) = E_a[u(c)]$$

where $E_a[.]$ denotes the expectation operator with respect to probability distribution $p_a(c)$ and $u(c)$ denotes a utility function over realized consumption c. This is an assumption about the functional form of $U(.)$. By definition of the expectations operator, we have:

$$\begin{aligned} E_a[u(c)] &= \sum_{i=1}^{N} p_i^a u(c_i^a) \text{ if the distribution is discrete} \\ &= \int_{-\infty}^{\infty} u(c) p_a(c) dc \text{ if the distribution is continuous} \end{aligned}$$

Expected utility theory assumes that $p_a(c)$ is preferred to $p_b(c)$ if and only if:

$$E_a[u(c)] \geq E_b[u(c)]$$

There are some problems with the expected utility setup. For instance, it assumes that decision makers are not confused by the way the decision problem is framed, that they are able to compute complicated probabilities, etc. Many of these assumptions are regularly violated in experimental games. In spite of these shortcomings, the setup is capable of representing the most important features of aversion toward risk at relatively low cost in terms of mathematical complexity. The expected utility framework is by far the dominant modeling framework in the economics of uncertainty.

1.2. Expected utility of income

Most of the literature on expected utility regards c not as a vector representing a consumption bundle but as a scalar representing consumption expenditures. This can be confusing at first because it obscures the relationship between expected utility and standard consumer theory. In reality, there is no contradiction between the two bodies of work. The expected utility framework can be seen as an extension of the standard consumer theory model to include uncertainty.

Perhaps the simplest way to see how the two are related is to imagine two periods. In the first period, the decision maker has to choose between two options, say a and b, without knowing his realized income. In the second period, income is known and the decision maker spends his income between various consumption goods. Conditional on income being y, the second period can thus be written:

$$V(y, p) = \max_c u(c) \text{ subject to } pc = y$$

where the standard indirect utility function is the solution to the standard utility maximization problem. If we are willing to assume that prices p are constant, the dependence of the indirect utility function on p can be suppressed from the notation. We did something similar when we introduced time. By a common abuse of notation, we can then write the utility of income (i.e., consumption expenditures) as $u(y)$. The first period problem then boils down to choosing between various distributions of income – more precisely, consumption expenditures – in period 2. If prices are also variable, the dependence on prices cannot be omitted.

1.3. Aversion to risk

Introduce the concepts of risk averse, risk neutral, and risk loving using graphical argument. By application of Jensen's inequality, we have $E[f(x)] \leq f(E[x])$ if $f(x)$ is a concave function. This shows that risk averse people – i.e., with a concave $u(y)$ – do not like risk. But how much do they dislike risk?

Arrow and Pratt used the expected utility framework to develop simple measures of aversion toward risk. These are called the Arrow-Pratt measures. It goes like this. Suppose that we face a consumer with a choice between a certain consumption of $\bar{y}-\pi$ and an uncertain consumption y. Here $\bar{y} \equiv E[y]$. We want to find the value of π that makes the consumer indifferent between the two options. This is formally similar to computing the equivalent or compensating variation (they are the same here since prices do not change, by assumption). Thus we wish to find π such that:

$$u(\bar{y} - \pi) = E[u(y)]$$

The 'trick' is to take Taylor approximations of both sides around $\bar{y}$, except that for the RHS we take a first order approximation and for the LHS we take a second order approximation. We obtain:

$$\begin{aligned} u(\bar{y}) - u'(\bar{y})\pi &\simeq E[u(\bar{y}) + u'(\bar{y})(\bar{y} - y) + \frac{1}{2}u''(\bar{y})(\bar{y} - y)^2] \\ u(\bar{y}) - u'(\bar{y})\pi &\simeq u(\bar{y}) + u'(\bar{y})E(\bar{y} - y) + \frac{1}{2}u''(\bar{y})E(\bar{y} - y)^2 \\ -u'(\bar{y})\pi &\simeq \frac{1}{2}u''(\bar{y})E(\bar{y} - y)^2 \equiv \frac{1}{2}u''(\bar{y})\sigma_y^2 \end{aligned}$$

After some straightforward manipulation, we can solve for π:

$$\pi \simeq -\frac{1}{2}\frac{u''(\bar{y})}{u'(\bar{y})}\sigma_y^2$$

π is the amount money the consumer would be willing to pay to get rid of uncertainty. It is positive as long as utility is concave, that is, as long as $u''(\bar{y}) < 0$. The larger the number, the more averse to risk the individual is. A risk neutral person would have a utility linear in total expenditures, and hence $u''(\bar{y}) = 0$.

The coefficient of absolute risk aversion A is defined as.

$$A \equiv -\frac{u''(\bar{y})}{u'(\bar{y})}$$

If we know A and σ_y^2 we can approximate the consumer's willingness to pay for the elimination of risk. The larger A is, the more the person is willing to pay, and the more averse to risk she is.

Since π is defined in currency units and σ_y^2 is defined in currency squared, A is implicitly defined in 1/currency units. It is not a unit-free measure. Consequently it is difficult to compare across countries or time periods. To eliminate this drawback, Arrow-Pratt introduced the coefficient of relative risk aversion R, which is defined as:

$$R \equiv -\bar{y}\frac{u''(\bar{y})}{u'(\bar{y})} = \bar{y}A$$

It is easy to verify that R is a unit-free measure.

It is usually believed that absolute risk aversion diminishes with average income while relative risk aversion is either constant or slightly declining with (increasing) average income. The utility function corresponding to constant relative risk aversion is:

$$u(y) = \frac{y^{1-R}}{1-R}$$

This function is used extensively in many economic models (e.g., intertemporal models, welfare function in inequality analysis, decision making under uncertainty). If $R = 1$, then replace the above with $u(y) = \log y$.

The coefficient of relative risk aversion is very easy to use. Let the coefficient of variation CV be defined as the standard deviation divided by the mean. The attraction of the coefficient of variation is that it is a unit-free measure. The coefficient of variation of income is:

$$CV_y \equiv \frac{\sigma_y}{\bar{y}}$$

Knowing the coefficient of variation of income and the coefficient of relative risk aversion, one can compute the proportion of average income a person would be willing to pay to reduce all uncertainty in income. We have:

$$\begin{aligned}
\frac{\pi}{\bar{y}} &\simeq -\frac{1}{2}\frac{u''(\bar{y})}{u'(\bar{y})}\frac{\sigma_y^2}{\bar{y}} \\
&\simeq \frac{1}{2}A\bar{y}\frac{\sigma_y^2}{\bar{y}^2} \\
&\simeq \frac{1}{2}R\, CV_y^2
\end{aligned}$$

So, for instance, if $R = 3$ and $CV_y = .5$, then the percentage of income the consumer is willing to pay for full insurance is $0.5 \times 3 \times 0.25 = 37.5\%$. The coefficient of variation of income for poor farmers typically reverts between 0.25 and 1.25. Estimates of the coefficient of risk aversion for poor households typically range from 0.5 to 4, with a mean of 2. This means that the willingness to pay for insurance is quite high: risk is a serious concern and people are likely to seek avenues to insure themselves against risk.

1.4. A two-period model

The above framework can be generalized to combine time and uncertainty. To illustrate this, we consider a consumer living two periods. Since there is no bequest in this world, in period 2 the individual consumes all his wealth. To focus on intertemporal decisions, we assume that consumption at time t is a scalar.[1] The consumer has an initial endowment of w_1 and can invest his wealth in two assets. One pays a certain return of R_C. The other pays a random return of $\widetilde{R}_U$. To simplify notation, these are total returns (1+ rate of return r), i.e., they include the reimbursement of the capital as well.

The consumer decides to consume c_1 in period 1. What is not consumed is invested in various assets; deciding about consumption is thus the same as deciding about saving. The consumer invests a fraction x of his wealth in the risky asset and a fraction $1-x$ in the certain asset. In this portfolio, the consumer has $(w_1 - c_1)x$ earning a return $\widetilde{R}_U$ and a fraction $(w_1 - c_1)(1-x)$ earning a return R_C. Therefore, his second period wealth – which equals c_2 – is:

$$\begin{aligned}
\widetilde{w}_2 &= \widetilde{c}_2 = (w_1 - c_1)x\widetilde{R}_U + (w_1 - c_1)(1-x)R_C \\
&= (w_1 - c_1)(x\widetilde{R}_U + (1-x)R_C) \\
&\equiv (w_1 - c_1)\widetilde{R}
\end{aligned}$$

where $\widetilde{R}$ is shorthand for the portfolio return.

[1] This is not such an outrageous assumption provided that consumption prices are constant. In this case, it is best to think of c_t as the consumption expenditures incurred at time t and of $u(c_t)$ as the indirect utility function corresponding to expenditures c_t, with the dependence on prices suppressed from the notation because prices are assumed constant.

The consumer's optimization problem can now be written:

$$\begin{aligned} &\max_{c_1,x}\ u(c_1) + \beta E[u(\widetilde{c}_2)] \text{ subject to} \\ \widetilde{c}_2 &= (w_1 - c_1)x\widetilde{R}_U + (w_1 - c_1)(1-x)R_C \end{aligned}$$

After plugging the constraint into the objective function, this yields two first order conditions for an unconstrained problem:

$$u'(c_1) = \beta E[u'(\widetilde{c}_2)\widetilde{R}] \tag{1.1}$$

$$E[u'(\widetilde{c}_2)(\widetilde{R}_U - R_C)] = 0 \tag{1.2}$$

(Be careful to put brackets around all random terms to remind yourself that random terms cannot be factored out of an expectation operator. This is because $E[ab] \neq E[a]E[b]$, unless a and b are independent random variables.) Equation (1.1) is an intertemporal optimization condition. In the literature, it is typically referred to as the *Euler condition.* Equation (1.2) is a portfolio optimization condition.

To see what the Euler equation implies, imagine for a moment that (1) $\widetilde{R}$ is non-random and (2) that $c_2 = c_1$. In this case, the Euler equation simplifies to:

$$\begin{aligned} u'(c_1) &= \beta u'(\widetilde{c}_2)R \\ \beta R &= 1 \\ R &= \frac{1}{\beta} = 1 + \delta \\ \text{since } R &= 1 + r \text{ we have} \\ r &= \delta \end{aligned}$$

This means that in the absence of uncertainty, the (non-random) interest rate r will equal the rate of time preference δ if consumption is constant over time. For this reason, economists often assume that δ is approximately equal to the risk-free interest rate r. Since in the long run the (real realized) return on non-risky assets hovers between 2 and 5%, it is usual to assume that β takes values between 0.95 and 0.98.

Turning to the portfolio optimization condition (1.2), we can use it to show that, if the decision maker is risk averse, then it must be that $R_C < E[R_U]$. Put differently, the average return on the risky asset must be greater than the return on the risk-free asset. To show this, without loss of generality we can define a random variable $\widetilde{\theta}$ such that:

$$\widetilde{R}_U = \widetilde{\theta} E[\widetilde{R}_U]$$

We note that when $\widetilde{R}_U < E[\widetilde{R}_U]$ then $\widetilde{\theta} < 1$ and vice versa. Armed with this notation, equation (1.2) can be rewritten:

$$\begin{aligned} E[u'(\widetilde{c}_2)(\widetilde{R}_U - R_C)] &= 0 \\ E[u'(\widetilde{c}_2)\widetilde{R}_U] &= E[u'(\widetilde{c}_2)R_C] \\ E[u'(\widetilde{c}_2)\widetilde{\theta} E[\widetilde{R}_U]] &= R_C E[u'(\widetilde{c}_2)] \\ E[\widetilde{R}_U]E[u'(\widetilde{c}_2)\widetilde{\theta}] &= R_C E[u'(\widetilde{c}_2)] \\ E[\widetilde{R}_U] &= R_C \frac{E[u'(\widetilde{c}_2)]}{E[u'(\widetilde{c}_2)\widetilde{\theta}]} \end{aligned}$$

We want to show that $E[u'(\widetilde{c}_2)\widetilde{\theta}] < E[u'(\widetilde{c}_2)]$ if the decision maker is risk averse. We see that $E[u'(\widetilde{c}_2)]$ is the average of $u'(\widetilde{c}_2)$ while $E[u'(\widetilde{c}_2)\widetilde{\theta}]$ can be regarded as a weighted average of $u'(\widetilde{c}_2)$ – some values of $u'(\widetilde{c}_2)$ are multiplied by $\widetilde{\theta} < 1$ and other values of $u'(\widetilde{c}_2)$ are multiplied by $\widetilde{\theta} > 1$. Can this observation help us?

If the decision maker is risk averse (concave function $u(c)$), the marginal utility of consumption is high for low values of consumption expenditures. Put differently, one extra unit of money generates a larger gain in satisfaction at low consumption levels that at high consumption levels. Now low values of θ correspond to low values of R_U and thus of c_2. This means that when the decision maker is risk averse, high values of $u'(\widetilde{c}_2)$ are multiplied by $\widetilde{\theta} < 1$ while low values of $u'(\widetilde{c}_2)$ are multiplied by $\widetilde{\theta} > 1$. It intuitively follows that, compared to the unweighted average $E[u'(\widetilde{c}_2)]$, the weighted average $E[u'(\widetilde{c}_2)\widetilde{\theta}]$ must be smaller. This implies that $E[u'(\widetilde{c}_2)\widetilde{\theta}] < E[u'(\widetilde{c}_2)]$ if the decision maker is risk averse, and hence that:

$$E[\widetilde{R}_U] > R_C \text{ since } \frac{E[u'(\widetilde{c}_2)]}{E[u'(\widetilde{c}_2)\widetilde{\theta}]} > 1$$

This shows that in order to induce a risk averse person to hold a risky asset, the asset must earn a 'risk premium', that is, a higher return than the risk-free asset.

By extension it follows that risk averse people refrain from making a risky investment if the risk premium is not large enough. To the extent that the risky investment is, on average, more remunerative, poor people tend to remain poor because they invest in safe assets with a low return.

Beyond the above observations, the solution to these two first order conditions is in general difficult to characterize without imposing additional structure either on $u(.)$ or on the probability distribution of returns $\widetilde{R}_U$ – or both. We simply note that the solution to the above system of FOCs defines an optimal choice of c_1 and x given an initial value of wealth w_1. Let us denote these optimal choice functions as:

$$\begin{aligned} c_1^* &= c(w_1) \\ x^* &= x(w_1) \end{aligned}$$

Plugging these optimal choices into the utility function yields what is called a value function:

$$V(w_1) = u(c(w_1)) + \beta E[u((w_1 - c(w_1))x(w_1)\widetilde{R}_U + (w_1 - c(w_1))(1 - x(w_1))R_C)]$$

Note the similarity between the above definition of a value function and the definition of the profit function, indirect utility function, and cost function: they are all value functions corresponding to a specific optimization problem. Function $V(w_1)$ denotes the level of utility the consumer can achieve if endowed with initial wealth w_1. In general, $V'(w_1) > 0$: more wealth is better than less. Dynamic programming takes the above observation one step further by making the above recursive. To this we now turn.

1.5. Dynamic programming

Keeping the same two assets, the model can be extended to an arbitrary number of periods:

$$U(c_1, ..., c_T) = \sum_{t=1}^{T} \beta^t E[u(c_t)]$$

Decisions about what to save $w_t - c_t$ and in which asset to invest x_t are now repeated in every period (except period T). This results in the same two first order conditions for each period – a system of $2(t-1)$ simultaneous non-linear equations.

In general, such system is extremely complicated to solve. One useful 'trick' is to note that instead of solving for consumption and portfolio decisions over all periods at once, we can solve the decision problem recursively. This trick is called dynamic programming. It is an extremely useful tool for numerical analysis and is often used in the presentation of economic models. It is much less useful for finding an algebraic solution to an economic problem, but then intertemporal optimization problems nearly never have algebraic solutions, so that this issue is moot. Dynamic programming can often help characterize some of the properties of the solution.

Varian illustrates dynamic programming using the complete intertemporal problem used in the previous sub-section. In my opinion, there is too much notation floating around so that it is difficult to catch the main idea. We focus on a simpler problem in which there is no certain asset, i.e., $x_t = 1$ for all t and $\widetilde{R}_U = \widetilde{R}$. To further simplify the notation, we define saving as

$$s_t = w_t - c_t$$

and optimize with respect to s_t instead of c_t. We have $w_t = s_{t-1}R$. The two approaches are equivalent and interchangeable.

The basic idea of dynamic programming is to reduce a multi-period optimization problem into a series of nested two-period problems. This is how it works. Consider the consumer's decision problem at $T-1$: it is nothing but the two period problem we discussed in the previous sub-section. In terms of saving, we have:

$$V_{T-1}(w_{T-1}) = \max_{s_{T-1}} \; u(w_{T-1} - s_{T-1}) + \beta E[u(s_{T-1}\widetilde{R})]$$

where, as before, $V_{T-1}(w_{T-1})$ denotes the value of the solution to this optimization problem. Let us now turn to the $T-2$ optimization problem:

$$\max_{s_{T-2}, s_{T-1}} \; u(w_{T-2} - s_{T-2}) + \beta E[u(s_{T-2}\widetilde{R} - s_{T-1})] + \beta^2 E[u(s_{T-1}\widetilde{R})]$$

This can be written equivalently as the nested problem:

$$\begin{aligned} & \max_{s_{T-2}} \; u(w_{T-2} - s_{T-2}) + \beta E\left[\max_{s_{T-1}} u(s_{T-2}\widetilde{R} - s_{T-1})] + \beta E[u(s_{T-1}\widetilde{R})]\right] \\ = \; & \max_{s_{T-2}} \; u(w_{T-2} - s_{T-2}) + \beta E[V_{T-1}(s_{T-2}\widetilde{R})] \end{aligned}$$

where we have made use of the fact that $w_{T-1} = s_{T-2}\widetilde{R}$. Once we have understood this trick, we can apply it again and define:

$$V_{T-2}(w_{T-2}) = \max_{s_{T-2}} \; u(w_{T-2} - s_{T-2}) + \beta E[V_{T-1}(s_{T-2}\widetilde{R})]$$

to solve the $T-3$ optimization and so on until time $t = 1$. This can be summarized in the following recursive equation:

$$V_t(w_t) = \max_{s_t} \; u(w_t - s_t) + \beta E[V_{t+1}(s_t\widetilde{R})] \tag{1.3}$$

This approach is called dynamic programming and equation (1.3) is called the *Belman equation*. The same approach can be applied to more complicated problems, such as the one discussed in the previous sub-section. An excellent (albeit quite challenging) treatment of dynamic programming for economists can be found in the book by Stokey and Lucas entitled "Recursive Methods in Economic Dynamics".

2. Assets and arbitrage

2.1. Arbitrage

Varian discusses a few issues surrounding assets. The first one is arbitrage. Let R_i be the return on asset i. In a world of certainty and without transactions costs, two assets must have exactly the same return for firms or individuals to hold them. If an asset has a return higher than others, it will be the only asset held as individuals and firms arbitrage between assets. For instance, consider a world with two assets – money and treasury bonds. Money yields a zero return, bonds yield a positive return. In such a world, money will not be held. Of course, things would be different if there are transactions costs so that liquidating bonds is costly. In this case, agents may hold some money to economize on transactions costs.

In practice, the world is uncertain so that assets can have different realized returns, that is, different returns ex post. If two assets coexist so that one has a certain return and the other an uncertain return, risk averse agents will typically refuse to hold the risky asset unless it has a higher expected return. The difference between the certain return and the higher expected return of the risky asset is called the risk premium. We can define the risk premium as follows. Let there be multiple assets with different returns denoted $\widetilde{R}_a$ with $a = \{1, ..., A\}$. Further assume that there is a risk-free asset with return R_0. (If there were other risk-free assets with a different return, arbitrage would dictate that only the asset with the highest risk-free return would be used. The others can thus be ignored.) The risk premium of asset a is then defined as:

$$\text{risk premium} = E[\widetilde{R}_a - R_0]$$

2.2. The Capital Asset Pricing Model

[This section is not exam material.] The Capital Asset Pricing Model or CAPM is an attempt to derive a formula for the risk premium. This formula is obtained by assuming that the utility function takes a very specific form, namely that:

$$E[u(\widetilde{w})] = E[\widetilde{w}] - AVar[\widetilde{w}]$$

This is a very peculiar form of utility function that obtains when utility is quadratic or when $\widetilde{w}$ has a normally distributed return.[2] This set up is usually justified as an approximation to more complicated utility functions. Parameter A measures how sensitive the agent is to risk.

The CAPM model is basically a two-period model with the above utility function. We begin by defining the budget constraint. The share of total wealth invested in asset a is written x_a. Obviously we need shares to sum to 1:

$$\sum_{a=0}^{A} x_a = 1$$

Second period consumption $c_2 = w_2$ is given by:

$$\begin{aligned} c_2 &= (w_1 - c_1) \sum_{a=0}^{A} x_a \widetilde{R}_a \\ &= (w_1 - c_1) \left[R_0 + \sum_{a=1}^{A} x_a (\widetilde{R}_a - R_0) \right] \end{aligned}$$

[2] This is because the distribution of a normal variable can be summarized by its mean and variance. Consequently, preferences over different normal variables can be described as preferences over combinations of mean and variance. By extension, this reasoning applies to other two-parameter distribution functions, e.g., log-normal.

The advantage of the second formulation is that it is expressed in terms of risk premium. The expression in square brackets is the portfolio return. Plugging into our utility function, we have:

$$\begin{aligned} E[w_2] &= E[c_2] \\ &= E\left[(w_1 - c_1)\sum_{a=0}^{A} x_a \widetilde{R}_a\right] \\ &= (w_1 - c_1)E\left[\sum_{a=0}^{A} x_a \widetilde{R}_a\right] \\ &= (w_1 - c_1)\sum_{a=0}^{A} x_a E[\widetilde{R}_a] \end{aligned}$$

$$\begin{aligned} Var[w_2] &= Var\left[(w_1 - c_1)\sum_{a=0}^{A} x_a \widetilde{R}_a\right] \\ &= (w_1 - c_1)^2 Var\left[\sum_{a=0}^{A} x_a \widetilde{R}_a\right] \\ &= (w_1 - c_1)^2 \sum_{a=0}^{A}\sum_{b=0}^{A} x_a x_b Cov(\widetilde{R}_a, \widetilde{R}_b) \end{aligned}$$

where we have made use of the fact that $Var[x + y] = Var[x] + Var[y] + 2Cov(x, y)$.

The idea of the CAPM model is that the agent wishes a portfolio that yields a specified expected return $\bar{R}$ but has the lowest possible variance:

$$\min_{x_0,\ldots,x_A} \sum_{a=0}^{A}\sum_{b=0}^{A} x_a x_b Cov(\widetilde{R}_a, \widetilde{R}_b) \text{ subject to} \tag{2.1}$$

$$\begin{aligned} \sum_{a=0}^{A} x_a E[\widetilde{R}_a] &= \bar{R} \\ \sum_{a=0}^{A} x_a &= 1 \end{aligned}$$

Note that x_a can either be positive or negative. If $x_a < 0$, it means that the agent is selling securities. Let λ be the Lagrange multiplier for the first constraint and μ for the second constraint. First order conditions take the form:

$$\begin{aligned} 2\sum_{b=0}^{A} x_b Cov(\widetilde{R}_a, \widetilde{R}_b) - \lambda E[\widetilde{R}_a] - \mu &= 0 \\ 2\sum_{b=0}^{A} x_b \sigma_{ab} - \lambda \bar{R}_a - \mu &= 0 \end{aligned} \tag{2.2}$$

where $\sigma_{ab} \equiv Cov(\widetilde{R}_a, \widetilde{R}_b)$ and $\bar{R}_a = E[R_a]$. This is a complicated system of equations. Varian tackles it in three steps.

Step 1: an expression for the risk premium

Let $(x_1^e, ..., x_A^e)$ be a portfolio made entirely of risk assets that is known to be mean-variance efficient, that is, to minimize the above problem for some $\bar{R}$. This asset is like a mutual fund. Given that this portfolio, by definition, has minimum variance, it is optimal for the agent to hold only two assets: the riskless asset R_0 and the mutual fund. Put differently, $x_a = 0$ except for $a = 0$ and $a = e$. We end up with two first order conditions:

$$\begin{aligned} -\lambda R_0 - \mu &= 0 \\ 2\sigma_{ee} - \lambda \bar{R}_e - \mu &= 0 \end{aligned}$$

Solving for λ and μ, we get:

$$\begin{aligned} \lambda &= \frac{2\sigma_{ae}}{\bar{R}_e - R_0} \\ \mu &= -\frac{2\sigma_{ae} R_0}{\bar{R}_e - R_0} \end{aligned}$$

Other first order conditions have the form:

$$2\sigma_{ae} - \lambda \bar{R}_a - \mu = 0$$

Substituting for λ and μ in the above and rearranging yields:

$$\bar{R}_a = R_0 + \frac{\sigma_{ae}}{\sigma_{ee}}(\bar{R}_e - R_0)$$

This equation says that the expected return on any asset equals the risk-free return plus a risk premium that depends on the covariance between the asset and the mutual fund e.

Step 2: the efficient portfolio

In step 1, we have assumed that an efficient portfolio exist. But what does it look like? The answer to this question is quite complicated. The standard treatment is to turn to a graphical argument. On a graph with the variance of returns on the x-axis and the mean return on the y-axis, we plot *all* possible portfolios of risky assets. This is a formidable set, with an infinite number of possible combinations. It can be shown that this set is compact – i.e., it has a well defined boundary. It can also be shown that the compact set lies entirely above both axes. [Show graph – Figure 20.1 in Varian.]

Efficient portfolios of risky assets are on the boundary of this set; other portfolios achieve similar expected returns with a higher variance, or similar variance with a lower expected return. Now we remember about the risk-free asset R_0. Draw the tangency line. The point of tangency with the efficient portfolio set is the efficient risky portfolio for that level of R_0. Let the return of this portfolio be denoted $\widetilde{R}_e$.

Since agents are assumed able to borrow and lend asset R_0 without constraint, it is clear that utility is maximized along the tangency line. Combinations of mean and variance above the line are not feasible. Combinations below the line can be improved upon by combining the risk free asset R_0 with the efficient portfolio $\widetilde{R}_e$. Any point along the tangency line can be achieved with the correct combination of R_0 and $\widetilde{R}_e$. Points beyond $\widetilde{R}_e$ can be achieved by borrowing money at R_0 and investing in $\widetilde{R}_e$.

The implication is that the solution to optimization problem (2.1) boils down to choosing the point along the tangency line that corresponds to the desired expected return $\bar{R}$. It immediately follows that *all* investors hold the same portfolio of risky assets $\widetilde{R}_e$. Investors only differ in the emphasis they put on risk, that is, the share of their wealth they hold in the risk-free portfolio and the risky portfolio $\widetilde{R}_e$.

This is quite a remarkable result – if only because it is wildly violated in practice. People, for instance, purchase a home or invest in education without being able to hedge all the location or sector-specific risk inherent in these investment decisions. This did not stop the CAPM model from enjoying widespread success. This success is explained by the next step.

Step 3: valuing market assets

The CAPM model can be used to value all risky assets as follows. To do so, we make use of the prediction that all agents hold the same portfolio – which we call the market portfolio and denote $(x_1^m, ..., x_A^m)$. For agent i, we have:

$$x_a^m W_i = p_a X_{ia}$$

where W_i is the wealth of agent i, p_a is the market price of asset a, and X_{ia} is the number of units of asset a held by agent i. Summing over all agents, we obtain:

$$x_a^m = \frac{p_a \sum_{i=1}^N X_{ia}}{\sum_{i=1}^N W_i}$$

In other words, x_a^m is nothing but the capitalization value of asset a relative to total capitalization value/total wealth.

It then follows that:

$$\begin{aligned}\bar{R}_a &= R_0 + \frac{\sigma_{am}}{\sigma_{mm}}(\bar{R}_m - R_0) \\ &= R_0 + \beta_a(\bar{R}_m - R_0)\end{aligned}$$

where σ_{am} is the covariance between asset a and the total market value of all risky assets and $\beta_a \equiv \frac{\sigma_{am}}{\sigma_{mm}}$. Since σ_{am} is observable, it becomes possible to compute the risk premium that corresponds to an efficient market outcome. In the finance literature, the ratio $\frac{\sigma_{am}}{\sigma_{mm}}$ is called the asset's β. All that is needed to price an asset a is β_a. The β_a of an asset is obtained by simply regressing the asset's return on the market return.

Needless to say, a massive number of assumptions are hidden in this result. Besides, if the market is as efficient as the CAPM model predicts, why would we ever need the formula? The market would automatically compute it! The CAPM model must be taken with a grain of salt, as some kind of approximation to an efficient situation. It has nevertheless been extremely influential and helps clarify portfolio issues tremendously.

3. Applications

There have been countless applications of the concepts presented here to development problems. In microeconomics, current research on poverty focuses on vulnerability and risk coping strategies, and hence is organized on the basis the intertemporal models with uncertainty. Portfolio issues crop up everywhere in development, e.g., when households must choose how many heads of livestock to keep as opposed to grain stocks, or when parents choose the education level of boys and girls.

The same models and concepts crop up constantly in macroeconomics. Commodity price shocks generate macro uncertainty, and hence the need to factor risk in. Much of the literature on foreign direct investment (FDI) and emerging markets reverts around some sort of risky portfolio argument. Inflation changes the relative returns of various assets in a way that dramatically affects the choice of efficient portfolio. With high and uncertain inflation, a risk-free asset no longer exist and the CAPM model collapses, wrecking havoc in financial markets as individuals with different preferences seek to hold different portfolios of risky assets.