Theory of the rate of return

This short note gives a summary of different circumstances that give rise to differences in the rate of return on different assets. We also provide a brief sketch of what macroeconomics can say about the general level around which these rates of return fluctuate.

In non-monetary models without uncertainty there is in equilibrium only one rate of return, \( r \). If in addition there is perfect competition in all markets and no capital adjustment costs, as in simple neoclassical models (like the Diamond OLG model or the Ramsey model), then the equilibrium real interest rate is at any time equal to the net marginal product of capital \( \rho = \partial Y/\partial K = \delta \) (standard notation). Moreover, under conditions ensuring “well-behavedness” of these models, they predict that the capital intensity, and thereby the marginal product of capital, adjusts over time to some long-run level (on which more below).

Different rates of return It goes without saying that the link between the real interest rate and the marginal product of capital is loosened by the need for means of payments. Or rather, this link is loosened by many factors among which money is just one. As we will see in Chapter 14, existence of convex capital adjustment costs loosens the link between \( r \) and \( \partial Y/\partial K \). The adjustment costs create a wedge between the price of investment goods and the market value of the marginal unit of installed capital. Besides the marginal product of capital, the possible capital gain in the market value of installed capital as well as the effect of the marginal unit of installed capital on future installation costs enter as co-determinants of the current rate of return on capital.

When imperfect competition on the output markets rules, prices are typically set as a mark-up on marginal cost. This implies a wedge between the net marginal product of capital and capital costs. And when uncertainty and limited opportunities for risk spreading are added to the model, a wide spectrum of expected rates of return on different financial assets and expected marginal products of capital in different production

Anticipated sectors arise, depending on the risk profiles of the different assets and production sectors. Moreover, the presence of taxation, and sometimes differential taxation on different asset returns, complicates the picture.

Table 1 reports the nominal and real average annual rates of return on a range of US asset portfolios for the period 1926–2001. The portfolio of small company stocks had an annual real return of 13.8 per cent (the arithmetic average throughout the period). This is more than that of any of the other considered portfolios. Small company stocks are also seen to be the most volatile. The standard deviation of the annual real rate of return of the portfolio of small company stocks is almost eight times higher than that of the portfolio of U.S. Treasury bills (government zero coupon bonds with 30 days to maturity), with an average annual real return of only 0.8 per cent throughout the period. Explanation in terms of risk aversion is in line with the displayed positive relation between high returns and high volatility. Yet, interpreting volatility as a rough measure of risk, the pattern is not without exceptions. The portfolio of long-term corporate bonds has performed better than the portfolio of long-term government bonds, although they have been slightly less volatile as here measured. But the data is historical, expectations are
not always met, and risk depends significantly on the correlation

of the asset’s return with the business cycle, a feature about which Table ?? has
nothing say; share prices are in fact very sensitive to business cycle fluctuations.

The need for means of payment — money — further complicates the picture. That is,
besides differences in risk and expected return across different assets, also dissimilarities
in their degree of liquidity are important, not least in times of financial crisis. The
expected real rate of return on cash holding is minus the expected rate of inflation and
is therefore negative in an economy with inflation, cf. the last row in Table ??.
When agents nevertheless hold money in their portfolios, it is because the low rate of return is
compensated by the liquidity services of money. In the Sidrauski model of Chapter 17 this
is modeled in a simple (albeit ad hoc) way by including real money holdings directly as an
argument in the utility function. Another dimension along which the presence of money
interferes with returns is through inflation. Real assets, like physical capital, land, houses,
etc. are better protected against fluctuating inflation than are nominally denominated
bonds (and money of course).

Without claiming too much we can say that investors facing these multiple rates of
return choose a portfolio composition so as to balance the need for liquidity, the wish for
a high expected return, and the wish for low risk. Finance theory teaches us that adjusted
for differences in risk and liquidity, asset returns tend to be the same. This raises the
question: at what level? This is where macroeconomics — as a theory about the economy
as a whole — comes to the fore.

**Macroeconomic theory of the trend level of rates of return** The point of departure is that market forces by and large tend to anchor the rate of return of an average portfolio to the net marginal product of capital in an aggregate production function. Some popular phrases are:

- the net marginal product of capital acts as a centre of gravitation for asset returns;
  and

- movements of the rates of return are in the long run held in check by the net marginal
  product of capital.

Though such phrases seem to convey the right flavour, in themselves they are not very
informative. The net marginal product of capital is not a given, but an endogenous vari-
able which, via changes in the capital intensity, adjusts through time to more fundamental factors in the economy.

The different macroeconomic models we have studied in previous chapters bring to mind different presumptions about what these fundamental factors are.

1. **Solow’s growth model** The Solow growth model (Solow 1956) leads to the fundamental differential equation (standard notation)

\[
\dot{k}_t = sf(\tilde{k}_t) - (\delta + g + n)\tilde{k}_t,
\]

where \( s \) is an exogenous and constant aggregate saving rate, \( 0 < s < 1 \). In steady state

\[
r^* = f'(\tilde{k}^*) - \delta,
\]

where \( \tilde{k}^* \) is the unique steady state value of the (effective) capital intensity, \( \tilde{k} \), satisfying

\[
sf(\tilde{k}^*) = (\delta + g + n)\tilde{k}^*.
\]

In society there is a debate and a concern that changed demography and less growth in the source of new technical ideas, i.e., the stock of educated human beings, will in the future result in lower \( n \) and lower \( g \), respectively, making financing social security more difficult. On the basis of the Solow model we find by implicit differentiation in (2) \( \partial \tilde{k}^*/\partial n = \partial \tilde{k}^*/\partial g = -\tilde{k}^*\left[\delta + g + n - sf'(\tilde{k}^*)\right]^{-1} \), which is negative since \( s f'(\tilde{k}^*) < s f(\tilde{k}^*)/\tilde{k}^* = \delta + g + n \). Hence, by (1),

\[
\frac{\partial r^*}{\partial n} = \frac{\partial r^*}{\partial g} = \frac{\partial r^*}{\partial \tilde{k}^*} \frac{\partial \tilde{k}^*}{\partial n} = f''(\tilde{k}^*) \frac{-\tilde{k}^*}{\delta + g + n - s f'(\tilde{k}^*)} > 0,
\]

since \( f''(\tilde{k}^*) < 0 \). It follows that

\[
n \downarrow \text{ or } g \downarrow \Rightarrow r^* \downarrow .
\]

2. **The Diamond OLG model** The Diamond OLG model also just concludes that \( r^* = f'(\tilde{k}^*) - \delta \). Like in the Solow model, the long-run rate of return thus depends on the aggregate production function and on \( \tilde{k}^* \), which in turn may depend in a complicated way on the lifetime utility function and the production function. The steady state of a well-behaved Diamond model will nevertheless have the same qualitative property as indicated in (3).
3. The Ramsey model  In contrast to the Solow and Diamond models, the Ramsey model implies not only that \( r^* = f'(\bar{k}^*) - \delta \), but also that the net marginal product of capital converges in the long run to a specific value given by the *modified golden rule* formula. In a continuous time framework this formula says:

\[
r^* = \rho + \theta g,
\]

where the new parameter, \( \theta \), is the (absolute) elasticity of marginal utility of consumption. Because the Ramsey model is a representative agent model, the Keynes-Ramsey rule holds not only at the individual level, but also at the aggregate level. This is what gives rise to this simple formula for \( r^* \).

Here there is no role for \( n \), only for \( g \). On the other hand, there is an alternative specification of the Ramsey model, namely the “discounted average utilitarianism” specification. In this version of the Ramsey model, we get \( r^* = f'(\bar{k}^*) - \delta = \rho + n + \theta g \), so that not only a lower \( n \), but also a lower \( g \) implies lower \( r^* \).

Also the Sidrauski model, i.e., the monetary Ramsey model of Chapter 17, results in the *modified golden rule* formula.

4. Blanchard’s OLG model  A continuous time model with OLG structure and emphasis on life-cycle aspects is Blanchard’s OLG model (Blanchard 1985). In that model the net marginal product of capital adjusts to a value within an interval:

\[
\rho + g - \lambda < r^* < \rho + g + b,
\]

where two additional parameters appear, the retirement rate \( \lambda \) (reflecting how early in life the “average” person retire from the labor market) and the crude birth rate \( b \) (\( \theta = 1 \) for simplicity). The population growth rate is the difference between the crude birth rate, \( b \), and the crude mortality rate, \( m \), so that \( n = b - m \). The qualitative property indicated in (3) becomes conditional. It still holds if the fall in \( n \) reflects a lower \( b \), but not necessarily if it reflects a higher \( m \).

5. What if technological change is embodied?  The models in the list above assume a neoclassical aggregate production function with CRS and disembodied Harrod-neutral technological progress, that is,

\[
Y = F(K,TL) \equiv TLf(\bar{k}), \quad f' > 0, f'' < 0.
\]
This amounts to assuming that new technical knowledge advances the combined productivity of capital and labor independently of whether the workers operate old or new machines.

In contrast, we say that technological change is embodied if taking advantage of new technical knowledge requires construction of new investment goods. The newest technology is incorporated in the design of newly produced equipment; and this equipment will not participate in subsequent technological progress. Both intuition and empirics suggest that most technological progress is of this form. Indeed, Greenwood et al. (1997) estimate for the U.S. 1950-1990 that embodied technological change explains 60% of the growth in output per man hour.

So a theory of the rate of return should take this into account. Fortunately, this can be done with only minor modifications. We assume that the link between investment and capital accumulation takes the form

\[
\dot{K}_t = Q_t I_t - \delta K_t, \tag{6}
\]

where \( I_t \) is gross investment \((I = Y - C)\) and \( Q_t \) measures the “quality” (productivity) of newly produced investment goods. Suppose for instance that \( Q_t = Q_0 e^{\gamma t}, \gamma > 0. \) Then, even if no technological change directly appears in the production function, that is, even if (5) is replaced by

\[
Y = F(K, L) = K^\alpha L^{1-\alpha}, \quad 0 < \alpha < 1,
\]

the economy will still experience a rising standard of living.\(^1\) A given level of gross investment will give rise to a greater and greater additions to the capital stock \( K \), measured in efficiency units. Since at time \( t \), \( Q_t \) capital goods can be produced at the same cost as one consumption good, the price, \( p_t \), of capital goods in terms of the consumption good must in competitive equilibrium equal the inverse of \( Q_t \), that is, \( p_t = 1/Q_t \). In this way embodied technological progress results in a steady decline in the relative price of capital equipment.

This prediction is confirmed by the data. Greenwood et al. (1997) find for the U.S. that the relative price of capital equipment has been declining at an average rate of 0.03 per year in the period 1950-1990, a trend that has seemingly been fortified in the wake of the computer revolution.

---

\(^1\)We specify \( F \) to be Cobb-Douglas, because otherwise a model with embodied technical progress in the form (6) will not be able to generate balanced growth and comply with Kaldor’s stylized facts.
Along a balanced growth path the constant growth rate of $K$ will now exceed that of $Y$, and $Y/K$ thus be falling. The output-capital ratio in value terms, $Y/(pK)$, will be constant, however. Embedding these features in a Ramsey-style framework, we find the long-run rate of return to be

$$r^* = \rho + \theta \frac{\alpha \gamma}{1 - \alpha}.$$ 

This is of exactly the same form as (4), if we define $g = \alpha \gamma/(1 - \alpha)$.

### 6. Adding uncertainty and risk of bankruptcy

Although absent from many simple macroeconomic models, uncertainty and risk of bankruptcy are significant features of reality. Bankruptcy risk may lead to a conflict of interest between share owners and managers. Managers may want less debt and more equity than the share owners because bankruptcy can be very costly to managers who lose a well-paid job and a promising carrier. So managers are unwilling to finance all new capital investment by new debt in spite of the associated lower capital cost (there is generally a lower rate of return on debt than on equity). In this way the excess of the rate of return on equity over that on debt, the equity premium, is sustained.

A rough, behavioral theory of the equity premium goes as follows. Firm managers prefer a payout structure with a fraction, $s_f$, going to equity and the remaining fraction, $1 - s_f$, to debt (corporate bonds). That is, out of each unit of expected operating profit, managers are unwilling to commit more than $1 - s_f$ to bond owners. This is to reduce the risk of a failing payment ability in case of a bad market outcome. And those who finance firms by loans definitely also want debtor firms to have some equity at stake.

We let households’ preferred portfolio consist of a fraction $s_h$ in equities and the remainder, $1 - s_h$, in bonds. In view of households’ risk aversion and memory of historical stock market crashes, it is plausible to assume that $s_h < s_f$.

As a crude adaptation of for instance the Blanchard OLG model to these features, we interpret the model’s $r^*$ as an average rate of return across firms. Let time be discrete and let aggregate financial wealth be $A = pK$, where $p$ is the price of capital equipment in terms of consumption goods. In standard versions of the above models we have $p \equiv 1$, but under item 5 above relative price. Anyway, given $A$ at time $t$, the aggregate gross return or payout is $(1 + r^*)A$. Out of this, $(1 + r^*)As_f$ constitutes the gross return to the

---

2 See Appendix.

3 This builds on Baker et al. (2005).
equity owners and \((1 + r^\ast)A(1 - s_f)\) the gross return to the bond owners. Let \(r_e\) denote the rate of return on equity and \(r_b\) the rate of return on bonds.

To find \(r_e\) and \(r_b\) we have

\[
(1 + r_e)A s_h = (1 + r^\ast)A s_f,
\]
\[
(1 + r_b)A(1 - s_h) = (1 + r^\ast)A(1 - s_f).
\]

Thus,

\[
1 + r_e = (1 + r^\ast)\frac{s_f}{s_h} > 1 + r^\ast,
\]
\[
1 + r_b = (1 + r^\ast)\frac{1 - s_f}{1 - s_h} < 1 + r^\ast.
\]

We may define the *equity premium*, \(\pi\), by \(1 + \pi \equiv (1 + r_e)/(1 + r_b)\). Then

\[
\pi = \frac{s_f(1 - s_h)}{s_h(1 - s_f)} - 1 > 0.
\]

Of course these formulas have their limitations. The key variables \(s_f\) and \(s_h\) will depend on a lot of economic circumstances and should be endogenous in an elaborate model. Yet, the formulas may be helpful as a way of organizing one’s thoughts about rates of return in a world with asymmetric information and risk of bankruptcy.

There is evidence that in the last decades of the twentieth century the equity premium had become lower than in the long aftermath of the Great Depression in the 1930s.\(^4\) A likely explanation is that \(s_h\) had gone up along with rising confidence; the computer and the World Wide Web have made it much easier for individuals to invest in stocks of shares. On the other hand, the recent global financial and economic crisis, the Great Recession 2007–, and the associated rise in mistrust may have halted and possibly reversed this tendency for some time.

---

\(^4\)Blanchard (2003, p. 333).