INVESTMENT FUNCTIONS AND THE PROFITABILITY GAP

COLIN RICHARDSON and PETER ROMILLY

Imperial College London       University of Abertay Dundee


All SHE Working Papers are Peer Reviewed

All opinions expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the Society of Heterodox Economists. The Society of Heterodox Economists is not responsible for the accuracy of the data and information included in this publication, nor does it accept responsibility for their use. Any queries arising from the contents of this publication should be directed to the authors concerned.

ISSN : 1834-7444
ISBN: 978 0 7334 2577 6
ABSTRACT

We examine a variety of fixed asset investment theory approaches and show that, despite apparent differences, all contain a common ‘gene’ – the profitability gap. This finding is equally applicable to the paradigms of neoclassical general equilibrium in logical time and post-classical fully-adjusted stationary and steady states in historical time. The generic investment function and its characteristic gene appear to be one of the universals of economic science, equally at home as explanators of investment, inflation/deflation and related cumulative processes.

Keywords: investment function, profitability gap, expected rate of profit, opportunity cost of capital.

JEL classifications: D21, E12, E13, E22.

1. Introduction

Our aim is to demonstrate that all fixed asset investment functions are driven by entrepreneurial reaction to differences or ratios of one kind or another, and that all such ‘gaps’ ultimately reduce to a common variable: the profitability gap. We examine the equations of investment theories proposed by Smith, Marx, Keynes, Kalecki, Tobin, Jorgenson, and others, and show that all classical uniform profitability, Keynesian marginal efficiency, neo-Keynesian multiplier-accelerator, and neoclassical q-theory and user cost investment functions are specific expressions of a single generic investment function. By analogy with an organism, the profitability gap term this function contains is viewed as the common gene of all species belonging to the investment function genus.

The paper does not provide a detailed survey of investment theories, since space limitations preclude this and several already exist (see, inter alia, Baddeley (2006)). Nor does it try to explain why most econometric models based on these theories have not performed well against datasets recording actual investment outcomes. Our more modest aim is to examine four classes, and in some cases their sub-classes, of investment function to demonstrate the presence of a common explanator. It should come as no surprise that investment is determined by profitability or, in our formulation, a profitability gap. But investment theory has developed in diverse and sometimes complex ways from paradigms that embody very different worldviews. This heterogeneity of approach has generated a rich variety of functions that attempt to encapsulate the motive forces behind entrepreneurs’ investment decisions. It is not always readily apparent, however, whether this
diversity of functions has any single factor in common. This paper demonstrates, through both logical and mathematical reasoning, that there is.

The following section introduces some basic definitions. In Section 3 we discuss the background to gap theories of investment and the profitability gap gene. Sections 4 through 7 extract the generic function and its signature gene from a number of specific investment functions proposed by different schools of economics within the four main approaches identified above. The final section offers our concluding remarks.

2. Some Basic Definitions

The generic investment function is specified as \( I_t = f(a_t, K_{t-1}) \) where \( I_t \) is gross investment, \( K_{t-1} \) is the opening capital stock and \( a_t \) is the profitability gap. Gross investment \( I_t \) is the sum of the replacement investment \( I'_r \) needed to offset depreciation of \( \delta K_{t-1} \) and the net investment \( I''_n \) needed to make \( K_{t-1} \) grow into \( K_t \), with \( \delta \) being the rate of depreciation. The profitability gap is defined as \( a_t = r'_i - n_t \), where, for simplicity, the time period is taken to be one year. No distinction is made between the investment functions of a firm, industry, sector, or economy because the underlying motives for capital accumulation by entrepreneurs are common to all levels, and (dis)aggregation schemes exist for moving between them.

The expected profit rate \( r'_e \) is the annual rate of profit \( R'_e / K_t \) that investors, towards the end of year \( t-1 \), expect to earn on their capital stock during the coming year \( t \), with \( R'_e \) representing expected non-labour income. Alternative terms for \( r'_e \) include ‘internal rate of return’, ‘marginal efficiency of capital’ (or ‘investment’), ‘rate of return over cost’, and ‘marginal revenue product of capital’, all understood as being generated by whatever profitability expectation function governs entrepreneurs’ decision-making processes.

The entrepreneurs’ normal profit rate \( n_t \) is their annual opportunity cost of capital. Other terms for \( n_t \) include ‘hurdle rate of return’, ‘target rate of profit’ and ‘required profitability’. It is defined as \( n_t = i_t + \varphi \), where \( i_t \) is the market rate of interest on money loans and \( \varphi \) is the risk premium associated with investment in some particular firm, industry, sector, or economy. Explicitly or implicitly, the simplifying assumption that \( \varphi = 0 \) is often made, so that \( n_t = i_t \) is the hallmark of a risk-free economy. Expected profitability \( r'_e \) may, and often will, differ from realised profitability \( r_t = R_t / K_t \) for all expectation functions – except for the singular case of rational expectations, where \( r_t = r'_e \) because the absence of systematic forecasting error drives the average forecasting error to zero.\(^1\)

3. Investment and the Profitability Gap Gene

The generic investment function with its embedded profitability gap gene can be traced back through John Maynard Keynes (1936), Michal Kalecki (1933), Irving Fisher (1930), and Keynes (1930) to Arthur Spiethoff (1925). Earlier, Knut Wicksell (1898, 1906) utilised gaps between the natural and money interest rates to explain price cycles, with Henry Thornton (1802) being the first to harness this profitability gap concept. As a ‘bullionist’ opposed to the ‘real bills’ doctrine, he showed how, whenever capital stocks yield returns higher (lower) than the market rate of interest, merchants will demand too much (too little) paper currency than is consistent with maintaining price level stability, thereby circumventing controls over the quantity of real bills.
After Spiethoff, Irving Fisher was the first economist to explicitly state that investment is driven by the profitability gap. Keynes (1936, pp 140-1) acknowledged Fisher’s strong influence on his own work:

“Although he does not call it the ‘marginal efficiency of capital’, Professor Irving Fisher has given in his Theory of Interest (1930) a definition of what he calls ‘the rate of return over cost’ which is identical with my definition. ‘The rate of return over cost’, he writes, ‘is that rate which, employed in computing the present worth of all the costs and the present worth of all the returns, will make these two equal.’ Professor Fisher explains that the extent of investment in any direction will depend on a comparison between the rate of return over cost and the rate of interest. To induce new investment ‘the rate of return over cost must exceed the rate of interest’.”

Thus the generic investment function can be traced back directly through Keynes (1936) to Fisher (1930). In Section 4 we show how Kalecki (1933) utilised the same concept, which earlier had been deployed by Spiethoff (1925) in his business cycle theory.

As an explainer of the price level, however, the profitability gap gene is far older. Keynes (1930) credited Wickel’s (1898, 1906) gap between the natural and money rates of interest, which drove the Swede’s ‘cumulative processes’ of deflation and inflation, as the inspiration for his own ‘investment-saving gap’ theory of profitability and the price level proposed in the Treatise on Money. Wickell himself identified Thornton (1802) as the ultimate progenitor of this universal ‘primitive’ of investment, capital and business cycle theory.

Other gaps play a central role in linking capital and investment theory. The positive, zero or negative differences between \( K_t \) and \( K_{t-1} \) (i.e. \( \Delta K_t \)), for example, are identical with those between \( I_t \) and \( \Delta K_{t-1} \). These gaps or differences also may be expressed as ratios, since \( K_t / K_{t-1} = 1 \) is mathematically equivalent to \( K_t = K_{t-1} \) and to \( K_t - K_{t-1} = 0 \). It follows that the generic investment function may be present in gap or difference theories and in ratio theories, a possibility we confirm in Sections 4 through 7 below. Crucially, the size of these capital stock and investment gaps is directly related to the size of another gap, that between expected and normal profitability.

Quantum physicists were embarrassed by their ‘particle zoo’ until, beginning in 1964, Murray Gell-Mann and others eventually demonstrated that these 200-plus different sub-atomic particles detected in their supercolliders were built from only six quarks, six leptons and six gauge bosons (plus their anti-particles). Perhaps economists should feel similarly uncomfortable about the ‘gap zoo’ on display in their own investment models. This paper demonstrates that all the various flow and stock gaps are merely proxies for something deeper and more fundamental, viz. the difference \( \alpha_t \) between expected profitability \( r^e_t \) and the opportunity cost of capital \( h_t \).

For instance, Keynes (1936, pp 315-7) insisted that gaps between the marginal efficiency of investment (MEI) and the long-term rate of interest were responsible for fluctuations in the investment aggregate. These, he claimed, become amplified by the multiplier into output instability, the trade cycle and the crises that intermittently plague capitalist economies. The multiplier-accelerator analyses of Harrod (1936), Samuelson (1939), Hicks (1950) and Goodwin (1951) showed how such crises can be generated.

With respect to capital and growth, Edmond Malinvaud (1986, p 382) comments that
“I agree with economic historians in thinking that an essential element … is the course of business profitability … this latter is precisely a deviation from the flexprice equilibrium.

“Business profitability may be characterised by the anticipated marginal pure profit rate (excess over the real interest rate … ). Over decades and excepting cases of major shocks, this can be properly measured by the mean realized pure profit rate … The existence of a non-zero pure profit rate is inconsistent with existing flexprice growth theories … the observed differences, with at some times and places negative, at others high pure profit rates … truly reveal what is best interpreted as disequilibria of the price system.” [Italics added]3

In the next four sections we use logico-mathematical reasoning to extract from a variety of investment theories and equations the (sometimes well-hidden) generic investment function and its profitability gap gene.

4. Classical “Uniform Profitability” Investment Functions

With the exception of Karl Marx and Thomas Malthus, all classical economists, including Adam Smith (1776) and David Ricardo (1817), accepted that supply creates its own demand, in accordance with Say’s Law of Markets. Given the classical Iron Law that real wages tend to the subsistence level – the corollary being zero average saving by workers – it is saving by capitalists out of their realised profits (of $R_t$ dollars pa) that determines investment. Saving by workers is considered by most classical economists to be an unusual, occasional and temporary occurrence, usually negated by later episodes of dissaving.

*Prima facie*, there is no room here for investment to be determined by the profitability gap. At best, one component of $a_i$ – the interest rate $i_t$ – could be said to influence the amount saved out of capitalists’ profit incomes. But the really interesting question is: what determines the macroeconomic aggregate $R_t$ of profits (hence also saving and investment) in the classical model?

At the microeconomic level, classical economists were aware that industries differed with respect to the risk premium $\phi$ that capitalists had to cover, before investing part of their profit-determined saving (opportunity cost $i_t$) in some particular industry. At this level, the allocation of saving across all lines of production must be governed by the rule $r^e_t \geq n_t$, where both sides of the inequality differed across industries. However, the right-hand side only differs because the risk premium $\phi$ is specific to each industry. The left-hand side differs because expectations of profitability also are industry-specific. What is general across all investment opportunities is the basal opportunity cost $i_t$, quoted by the rentiers, of converting foregone consumption (i.e. saving) into particular stocks of working capital and fixed capital equipment.

We can combine these facts with the insistence by most classical economists – although there were exceptions, including Smith and Marx - upon a natural tendency for the economy to gravitate towards its ‘dismal’ stationary state, in which only replacement investment occurs. One is left with a long-period equilibrium where the economy’s average $r^e_t$ has come into equality with its average $n_t$, underpinned by its common interest rate of $i_t$. The equalities $r_{t-1} = n_{t-1} = r^e_t = r_t = n_t$ are replicated, in the stationary state, year after year *ad infinitum*. At this set of uniform rates of return on capital and on risk-adjusted money loans – and with equilibrium saving out of equilibrium profits being equal to equilibrium investment – it follows that $I_t = S_t = g(R_t) = f(r^e_t - n_t)$, as per our generic investment function.
Thus it is microeconomic competition between capitalists to invest their flow of saving out of profits, in those industries they expect will yield the highest rates of return, which results in a particular macroeconomic outcome. The economy will be pushed onto its production possibilities frontier (PPF), with aggregate real income continuously maximised in a classic stationary state. So, in this equilibrium of zero wage and price inflation, it is true that $Y_t = Z$, dollars pa where $Z_t$ is the maximum annual flow of output that the economy can produce when all firms operate at their full capacity utilisation levels. By definition, the GDP is $Y_t = W_t + R_t$ dollars pa, and the wages bill is $W_t = w_t L_t$ dollars pa. With the uniform wage rate of $w_t$ dollars/worker pa stuck at the subsistence minimum, only the stock of employment ($L_t$ workers) and the flow of profits ($R_t$ dollars pa) are free to adjust.

With $w_t$ fixed by the Iron Law, it must be $L_t$ and/or $R_t$ that are the motive forces pushing $Y_t$ all the way out to $Z_t$ on the economy’s PPF. Whereas Marx assumed an industrial “reserve army of labour” in the form of the urban unemployed, most other classical economists relied on attracting low-productivity rural labour. Effectively, both scenarios result in an unlimited supply of labour at the prevailing real wage. As the level of employment $L_t$ is a purely passive variable, the sole active force is the microeconomic competitive struggle between capitalists to maximise their profitability gaps between $e_t$ and $n_t$, industry by industry. This same process therefore also maximises the macroeconomic aggregate of $R_t$ dollars pa they receive as profits, so the economy reaches its PPF and stays there for as long as the stationary state endures.

As the last of the classical economists, Marx accepted that fierce competition between capitalists tended to make realised profit rates uniform throughout the economy. What he could not stomach was the classical economists’ creed that capitalists frugally saved, then passively invested, real resources that always were limited by whatever profit incomes the market dictated. Those whom Marx criticised do not seem to have been aware that their own classical investment process, which equalised the microeconomic profitabilities across all firms, also maximised the macroeconomic aggregates of profits, hence saving, hence investment. So, in any given year when the process was active, aggregate gross investment really was determined by the profitability gap gene in the classical model. The process continued until the dismal long-period equilibrium stationary state of $I_t = 0$ and $I_t = \delta K_{t-1}$ and $a_t = 0$ was attained.

Marx, however, was a dynamic disequilibrium theorist. To him, the economy was always in traverse. “Accumulate, accumulate; that is Moses and the Prophets!” Marx (1867, p 742) exclaimed, thus ruling out the classical inevitability of the stationary state and substituting in its place a relentlessly growing and fluctuating economy, subject to intermittent crises. Positive net investment was the norm and Say’s Law inoperative. If capitalist entrepreneurs lacked sufficient current or retained profits to support their investment schemes, there were always plenty of capitalist rentiers on hand to extend money loans on the promise of future profits – provided, of course, the entrepreneurs had sufficient collateral.

Lack of collateral was the only thing preventing frugal workers from becoming capitalists. Marx, like Ricardo and others, did permit wages occasionally to fluctuate above subsistence, but workers’ savings were their only insurance against starvation during the next downturn. It was too risky for the working class to use them as collateral for rentier loans and set up as capitalists on their own account. When capital accumulation was strong (weak), wages rose (fell) and the reserve army of labour shrank (expanded). Whenever an investment boom carried the economy onto its PPF, this did not usher in a stationary state because, for Marx, this frontier was forever moving outwards, due to net investment that embodied the fruits of 19th century technical progress.
According to Ernest Mandel (1990), Marx shows that it is competition for profits which basically fuels a "... terrifying snowball logic: initial value of capital – accretion of value (surplus-value) – accretion of capital – more accretion of surplus-value – more accretion of capital, etc. ‘Without competition, the fire of growth would burn out’, (Marx, 1894, p 368).” As with his contemporaries, it is the classical uniform profitability investment function, based on the competitive struggle and containing the profitability gap gene, that Marx is using to explain capital accumulation, albeit without the stationary outcomes.

5. Keynesian “Marginal Efficiency” Investment Functions

The internal rate of return (IRR) or expected profit rate $r_t^e$ is the concept Keynes uses in the General Theory (1936, pp 135-46), where he refers to the marginal efficiency of capital (MEC) or, more accurately, of investment (MEI). In Chapter 11, Keynes argues that, if equilibrium prevails, then aggregate investment must have been pushed to the point where the economy-wide MEI has reached equality with the ruling rate of interest. Here he was abstracting from risk, assuming that $\varphi = 0$, but it is basically the same opportunity cost of capital concept, viz. $n_t = i_t + \varphi$. So, in the long-period equilibrium of a stationary state, it must be true that $r_t^e = r_t = n_t$. For the stationary state to be maintained, the equalities $r_{t-1} = n_{t-1} = r_t = r_t = n_t$ and $I_t = \delta K_{t-1}$ must remain true for all entrepreneurs, year after year.

When Keynes (1936, p 157) spoke of “… the forces of time and our ignorance of the future”, it was obvious he believed the universe to be non-ergodic. He did not subscribe to the ergodic axiom of neoclassical economics, whereby economic agents “… draw samples from the past or present, assume that such samples are equivalent to drawing samples from the future, and then place them into an optimising algorithm”, as Jerry Courvisanos (1996, p 164) characterises it. Economic models of investment always should disclose which particular profitability expectations function is being employed to generate the expected profit rate $r_t^e$, although Keynes avoided this duty by relying on ‘animal spirits’ and ‘the state of the news’.

Three years before the General Theory, Kalecki (1933, in Polish) had published a model “… identifying aggregate investment orders as a function of both anticipated gross profitability and interest rates”, according to Courvisanos (1996, p 15). In Kalecki’s (1933) Version I model, the investment function

$$I_t = f(r_t^e - i_t) \quad (1)$$

contains the profitability gap gene in pristine form. Kalecki went on to substitute average realised profitability – an ‘average expectations function’ – for the unobservable expected profit rate. The right-hand side of equation (1) is identical with the gap between the MEI and the interest rate that drives investment in Keynes (1936), and pre-dates it by three years, as did Kalecki’s derivation of the principle of effective demand.

In Kalecki’s (1943) Version II model, the investment function

$$I_t = f(\Delta R_t, \Delta K_t, R_t^e) \quad (2)$$

has investment responding positively to the ‘profits gap’ $\Delta R_t = R_t - R_{t-1}$ and negatively to the ‘capital stock gap’ $\Delta K_t = K_t - K_{t-1}$. The third term shows that investment also responds positively
to cash flow $R_t^c$. Kalecki included this to reflect his principle of increasing risk, whereby more internal financing means less recourse to risky external borrowing, *contra* the Modigliani-Miller theorem discussed later in Section 7.3. Kalecki’s principle was a precursor to the neoclassical ‘financial constraint’ investment theories examined in Section 7.

Dividing the first gap by the second yields the marginal profit rate $\Delta r_t = \Delta R_t / \Delta K_t$, which is equivalent to the difference between $r_t = R_t / K_t$ and $r_{t-1} = R_{t-1} / K_{t-1}$. Thus equation (2) can be recast as

$$I_t = f(r_t - r_{t-1}, R_t^c)$$

(3)

Therefore this function also contains the profitability gap gene – but only if it is true that $r_t = r_t^c$ and $r_{t-1} = n_t$.

The $r_t = r_t^c$ condition holds if $r_t$ represents an average of previously realised profit rates, the expectations function (for future profitability) specifically recommended by Malinvaud and actually used by Kalecki for his operational version of equation (1). The $r_{t-1} = n_t$ condition holds if Kalecki began his analysis from a fully-adjusted state, in which the normal profit rate is being earned on the opening capital stock. Kalecki (1954) suggests that this interpretation is correct: “… [if] we consider the rate of investment decisions in a short-period we can assume that at the beginning of this period the firms have pushed their investment plans up to the point where they cease to be profitable either because of the limited market for the firm’s products or because of ‘increasing risk’ and the limitations of the capital market …” (pp 96-7), [Italics added]

In Kalecki’s (1968) Version III model, investment depends on marginal profitability alone, so that

$$I_t = f(\Delta r_t)$$

(4)

is equivalent to equation (3) without the cash flow $R_t^c$ explanator. Kalecki realised that, because of technical progress, later vintages of capital stock tend to exceed earlier ones in productivity performance, hence also in profit potential, thus accounting for the positive gap between $r_t$ and $r_{t-1}$. Effectively, he returns to his starting point, equation (1). Consequently, Kalecki’s three Versions can all be interpreted as species of our generic investment function.

6. Neo-Keynesian “Multiplier-Accelerator” Investment Functions

Kalecki’s macroeconomic income distribution analysis is important for detecting the presence of profitability gaps/differences and ratios in the multiplier-accelerator investment theories that follow. He shows how the consumption and output aggregates favoured by neo-Keynesian investment theorists depend upon the total investment and consumption outlays of all entrepreneurs.

Kalecki expands the expenditure components of a closed economy’s gross domestic product into $Y_t = C_t^w + C_t^p + I_t$, where the first two right-hand side terms are consumption out of wage and profit incomes respectively. Then he uses the classical assumption concerning propensities to save out of wages $W_t$ and profits $R_t$ (*viz.* $s_w = 0 < s_p < 1$) to forge a link with the corresponding income components of GDP, $Y_t = W_t + R_t$. So, if the wage bill is $W_t = C_t^w$, then the profit residual $R_t = C_t^p + I_t$ must follow. Finally, Kalecki proposes $(C_t^p + I_t) \rightarrow R_t$ as the direction of causation.
This is entirely plausible since capitalists, having collateral and hence preferred access to rentier finance, can decide their own investment and consumption outlays, but never their own profits. It is ‘the market’ that decrees what profits each may subsequently earn … or so they think.

However, what no isolated entrepreneur ever perceives is that the aggregate of all capitalists’ investment outlays \( I_t \) principally determines what level of profits \( R_t \) the market will generate for them, in the form of the average rate of profit \( r_t \) they realise on the economy’s aggregate capital stock \( K_t \).

Thus, Kalecki (1971, p 13) could state that “… capitalists, as a whole, determine their own profits by the extent of their investment and personal consumption”, an insight he attained in the 1930s. It has since been distilled into Kalecki’s dictum: ‘Workers spend what they get, but capitalists get what they spend’. Sidney Weintraub (1979, p 101) describes Kalecki’s dictum as “… a penetrating light beam that speeds us close to the real situation”. Independently, Keynes (1930) had derived his equivalent ‘widow’s cruse’ explanation of how profits are generated:

“If entrepreneurs choose to spend a portion of their profit on consumption …, the effect is to increase the profit on the sale of liquid consumption goods by an amount exactly equal to the amount of profits which have been thus expended. … Thus, however much of their profits entrepreneurs spend on consumption, the increment of wealth belonging to entrepreneurs remains the same as before. Thus profits, as a source of capital increment for entrepreneurs, are a widow’s cruse which remains undepleted however much of them may be devoted to riotous living.” (p125).

In the neo-Keynesian consumption accelerator theory of Samuelson (1939), the relevant gap is the difference between current and lagged consumption

\[
I_t = f(C_t - C_{t-1}) = f(\Delta C_t) \tag{5}
\]

where the right-hand side is a proxy for the profitability gap, as demonstrated below. This also is true of the output accelerator theory that Roy Harrod (1936) pioneered, J R Hicks (1950) extended and econometricians such as Lawrence Klein (1950) utilised, namely

\[
I_t = f(Y_t - Y_{t-1}) = f(\Delta Y_t) \tag{6}
\]

Keynes (1936) showed how \( I_t \) – via the multiplier – determines \( Y_t \), and hence also \( C_t Y_t - I_t = C_t^u + C_t^r \). Furthermore, Kalecki’s dictum shows how \( C_t^r + I_t \) determines \( R_t \) – albeit by assuming that \( C_t^w = W_t \), implying that workers do not save. Yet, regardless of the saving behaviour of workers, it remains true that \( W_t = Y_t - R_t \), and this helps us determine causation. Combining the insights of Keynes (\( I_t \rightarrow Y_t \)) and Kalecki (\( I_t \rightarrow R_t \)) leaves the wage bill \( W_t = w_t L_t \) as a pure residual. The money wage \( w_t \) might be contractual, but employment \( L_t \) is not, so it must be the entrepreneurs’ aggregate investment (determining profits, output, consumption, and employment) that drives the macroeconomic fundamentals.

Now in the current (previous) year, the capital stock \( K_t \) \( (K_{t-1}) \) is given, so the current realised profit rate \( r_t \equiv R_t / K_t \) must be implicit in both \( C_t \) and \( Y_t \), just as the lagged realised profit rate \( r_{t-1} \equiv R_{t-1} / K_{t-1} \) must be implicit in both \( C_{t-1} \) and \( Y_{t-1} \). Thus, our two neo-Keynesian accelerator formulations – the consumption gap equation (5) and income gap equation (6) above – may be
viewed as proxies for the functions containing the profitability gap, namely
\[ I_t = f(r_t - r_{t-1}) = f(\Delta r_t) \]. These are the same as equations (3) and (4), respectively, in Kalecki’s Version II and III models discussed above.

The output accelerator also can be derived using standard neoclassical profit-maximisation analysis with perfect competition. Gapinski (1982, pp 95-98), generalising an earlier derivation by Hamberg (1971, pp 21-24), shows that changes in the factor price ratio – the wage rate relative to the profit rate – can alter the capital-labour and capital-output ratios when a constant elasticity of substitution production function is assumed. Gapinski (1982, p 98) concludes that the linear form of the output accelerator “… affords substantial convenience in exposition [and] it may be regarded as providing a crude approximation to more sophisticated investment functions”.

Linearity is not a prerequisite for linking accelerator theories to the profitability gap, however. Goodwin’s (1951) non-linear accelerator model contains the gene in near-pristine form as

\[ I_t = f(r_t^e - i_{t-1}) \] (7)

with the full specification of his discontinuous (or ‘bang-bang’) investment function being

\[
\begin{align*}
I_t &= \bar{I}_t \quad \text{if} \quad r_t^e > i_{t-1} \\
I_t &= 0 \quad \text{if} \quad r_t^e = i_{t-1} \\
I_t &= -\delta K_{t-1} \quad \text{if} \quad r_t^e < i_{t-1}
\end{align*}
\]

where \( \bar{I}_t \) is the supply-constrained maximum level of investment and \(-\delta K_{t-1}\) is the depreciation-constrained maximum level of disinvestment.

7. Neoclassical Investment Functions

7.1 The q-ratio

The q-ratio theory, originating with William Brainard & James Tobin (1968) and Tobin (1969), states that net investment by a firm depends directly on \( q_t \), the ratio of the stock market valuation \( (K_t^d) \) to the replacement cost \( (K_t^r) \) of that firm, viewed as a collection of capital assets

\[ I_t = f(q_t) \] (8)

where \( q_t = K_t^d / K_t^r \). When \( q_t > 1 \) (\( q_t < 1 \)) there will be positive (negative) net investment. When \( q_t = 1 \) there is no incentive to change the firm’s stock of capital assets, so only replacement investment occurs.

When \( q_t = 1 \), the firm considers it already possesses an optimal capital stock \( (K_t^*) \), so that \( K_t^* = K_t^d = K_t^r \) must represent the outcome of successful efforts by managers to maximise the equity value of the firm to its shareholders. Associated with each possible value of capital stock is some maximum capacity to produce output \( (Z_t) \). Optimal capital stocks or production flow capacities \( (Z_t^*) \) are key concepts in the neoclassical user cost investment functions discussed below.
In microeconomic investment analysis, the firm’s opportunity cost of capital ($n_t$) is used to discount back the net proceeds expected over the life of (say) a factory from its commissioning date onwards. Call this amount $P_t^d$, the factory’s Marshallian demand price, then accumulate forward the construction outlays up to commissioning date to find $P_t^s$, its Marshallian supply price. For a firm whose only asset is such a factory, $K_t^d = P_t^d$ and $K_t^s = P_t^s$, so the $q$-ratio investment theory involves comparing the results of forward-looking and backward-looking present value calculations.

The option with the highest net present value ($NPV$) is also the one with the greatest excess of the internal rate of return (IRR) or expected profit rate $r_t^e$ over the normal profit rate $n_t$. So, if all managers are striving to maximise the $NPV$s of their firms, the $q$-ratio theory reduces to the profitability gap theory. Note that the use of stock market valuations in the numerator of the $q$-ratio implies that the opinions of those who own the firm (its shareholders) are assumed to be identical with the knowledge of those who control the firm (its managers), this symmetry being a common situation in the 19th and early 20th centuries when most business firms were managed by their owners. Modern corporations, however, are characterised by asymmetry between shareholding outsiders and managerial insiders. Section 7.3 examines the implications of this asymmetry in more detail.

Estimates of both internal rate of return ($r_t^e$) and Marshallian demand price ($P_t^d = K_t^d$) are prospective, based on the expectation functions that guide investment behaviour. In this respect Keynes (1936, pp 156-8) contrasts ‘enterprise’ with ‘speculation’, noting that a 20th century phenomenon – the separation of ownership from control – encourages speculation and reduces enterprise. Rentier share trading, he remarked, was comparable to the farmer who, having tapped his barometer, withdraws all his capital from agriculture during a few days of expected bad weather. Finally, Keynes (1936, p 159) warns “Speculators may do no harm as bubbles on a steady stream of enterprise. But the position is serious when enterprise becomes the bubble on a whirlpool of speculation.”

The word “bubble” has entered the theoretical lexicon of neoclassical economists, for example in the paper by Robert Chirinko and Huntley Schaller (2001). After much advanced econometric analysis and testing, they conclude: “The data suggest that there was a bubble that had an economically important and statistically significant effect on fixed investment in Japan.”

Along the way, they note that

“A variety of theoretical work has called the simple present value model of stock prices into question. Empirical studies have provided evidence that stock prices may vary too much relative to dividends, that investors may overreact, that there may be fads in stock market prices, and that there may be a tendency to be overly optimistic about the future performance of stocks that have done well in the recent past …” (p 663).4

So, with empirical estimates of the Marshallian demand price numerator of the $q$-ratio most often sourced from stock market company valuations, rather than from internal company valuations based on the underlying fundamentals and made by better-informed managers, it is unsurprising that this investment function has not performed well under econometric testing.

Andrew Abel (1983) argues that the marginal $q$-ratio is a more relevant measure than the average $q$-ratio discussed above. Marginal $q_t$ is defined as the ratio of the market value of an additional piece of capital equipment to its replacement cost. It is difficult to obtain data on, or even proxies for, this...
theoretically superior concept, although Hayashi (1982) cites cases where marginal $q$ is proportional to average $q$, such as when the operating profit function and the augmented adjustment cost function are of the same degree of homogeneity.

7.2 User Cost

Tobin’s $q$-theory is the bridge linking the neo-Keynesian multiplier-accelerator and neoclassical user cost investment functions with the Keynes/Kalecki marginal efficiency approach. In the neoclassical optimal investment theories inspired by Dale Jorgenson (1963), the relevant gaps are those between last year’s and this year’s optimal capital stocks

$$I_t^* = f(K_t^* - K_{t-1}^*)$$  \hspace{1cm} (9)

or between the corresponding optimal production flow capacities

$$I_t^* = f(Z_t^* - Z_{t-1}^*)$$  \hspace{1cm} (10)

along a steady-state growth path. Trygve Haavelmo (1960, p 216) earlier had pointed out that “…the demand for investment cannot simply be derived from the demand for capital. Demand for a finite addition to the capital stock can lead to any rate of investment, from almost zero to infinity, depending on the additional hypothesis we introduce regarding the speed of reaction of capital-users”. Thus, Jorgenson had to invoke various ad hoc “delivery lags” and “adjustment costs”, modelled by distributed lags, to explain why capital stock gap closure does not occur instantaneously.

In the Preface to Volume I of his Collected Works, Jorgenson (1998) reminisces that he had “…defined the user cost of capital as the rental price of capital services, representing this price as the product of the price of investment goods and the cost of capital … I reserved the term ‘cost of capital’ for the sum of the rate of return, the rate of depreciation and the rate of capital loss, adjusted for the taxation of capital income”. His ‘rate of return’ is simply the profit rate, net of depreciation, which entrepreneurs expect to earn over the service life of the investment scheme they are contemplating. In a world with no taxes on capital and no changes in the price of investment goods relative to that of all goods, Jorgenson’s ‘cost of capital’ and ‘user cost of capital’ concepts would be identical, as we now demonstrate.

Consider a neoclassical analysis of the underlying determinants of capital stock adjustment. In a risk-free world ($\varphi = 0$, so that $n_i = i_i$), entrepreneurs react to gaps between their existing and optimal capital stocks by adjusting the former until the marginal product of capital ($MP_k$) in the ruling production function equals the rate of interest. The underlying process depends on estimating the net present value of the annual flow of gross profits that entrepreneurs expect to earn on their capital stock. The nominal capital stock is the real capital stock multiplied by the price index of investment goods $s_t$, so that $K_t = s_t K_t^*$ dollars.

The price index of all goods is $p_t$, where $p_t = s_t = 1$ is maintained over time in a world of zero inflation and no relative price changes. It follows that

$$NPV \equiv R_0^g + \frac{R_1^g}{1 + r} + \frac{R_2^g}{(1 + r)^2} + ... + \frac{R_T^g}{(1 + r)^T} - s_t K_t^*$$
where \( R_g^t = p_t Y_t^r - w_t L_t \) dollars pa. \( R_g^t \) is that part of operating revenue annually accruing to entrepreneurs, i.e. gross operating surplus. Net operating surplus, \( R = R_g^t - \delta K_t \) dollars pa, is not appropriate because this calculation spans the service life of \( K_0^t \) (viz. \( T \) years) and the nominal capital stock is subtracted as a lump sum, not in a series of annual depreciation charges. The price and quantity of output are denoted by \( p_t \) and \( Y_t^r \) respectively, whilst \( w_t \) is the wage rate and \( L_t \) the labour force. The future values of all these variables are expectations, but their \( e \)-superscripts are dropped to avoid cumbersome notation.

The rate of return and the life of the capital asset are given by \( r \) and \( T \) respectively. The production function \( Y_t^r = f(L_t, K_t^r) \) has positive but diminishing marginal products for both inputs. Assuming output consists of a single good that may be consumed or accumulated, \( p_t \) must be identical to \( s_t \). Given perfect competition, Phelps (1963, pp 270, 272) and Gapinski (1982, pp 167-169) show that the optimal capital stock is achieved when

\[
R_g^t = (\delta + r)s_t K_t^r \quad \text{dollars pa} \tag{11}
\]

where \((\delta + r)s_t = (\delta + r)\) when \( s_t = 1 \), can be interpreted as Jorgenson’s cost of capital, or user cost of capital, or annual capital rental. Subtracting annual depreciation expense \( \delta K_t^r \) from the left-hand side of equation (11) converts expected gross profit \( R_g^t \) into expected net profit \( R_t \), the metric of most interest to entrepreneurs forming their investment plans. Performing the same operation on the right-hand side of equation (11) involves subtracting \( \delta \), the depreciation rate, leaving only \( r \) inside the brackets. So, in net terms the user cost of capital is also the expected average profit rate \((r_t \equiv R_t / K_t)\). Using this result it is evident that the firm’s capital stock is optimised when the expected profit rate, hence also the profitability gap \( r_t^e - i_t \), is maximised.

A fuller statement of the firm’s optimal investment function, based on the gap between two adjacent optimal capital stock values – and shorn of its \textit{ad hoc} distributed lags structure – would be

\[
I_t = f(\Delta Y_t^r, \Delta p_t, \Delta r_t) \tag{12}
\]

So, with expected changes in output prices \( p_t \) and in user cost \( r_t \) (identical with the firm’s expected profit rate) determining the optimal capital stock, this neoclassical investment theory already resembles our generic investment function with its profitability gap gene. Jorgenson’s investment function also includes an output \((\Delta Y_t^r)\) accelerator from neo-Keynesian theory, and it was shown in Section 6 that all multiplier-accelerator equations contain the gene.

But resemblance is not enough. Jorgenson’s \textit{ad hoc} adjustment costs were soon separated by Eisner & Strotz (1963), Lucas (1967) and Gould (1968) into intrinsic factors (i.e. costs of installation) and extrinsic factors (i.e. rising Marshallian supply price), then formalised as a convex function of the firm’s capital stock to reflect \textit{marginal} adjustment costs. Thus was the neoclassical investment model ‘perfected’: it yields an entire optimal adjustment path for the scale of the firm and, on the representative agent assumption, the entire economy.

Several commentators, including Hayashi (1982) and Abel (1990), have shown that the Eisner-Strotz-Lucas-Gould neoclassical model with marginal adjustment costs is formally equivalent to Tobin’s (marginal) \( q \)-ratio theory of investment under certain assumptions, e.g. that the firm’s cash
flows are a linear homogeneous function of its capital and labour inputs and its investment outlays. This equivalence implies that neoclassical user cost investment functions also contain the profitability gap gene.

7.3 Financial Constraint and Real Option Value

The perfected neoclassical (user cost or $q$-ratio) investment function has proved to be flexible, even amoebic, readily absorbing such critiques as the influential ‘financial constraint’ and ‘real option value’ approaches to explaining investment. Jensen & Meckling (1976), Stiglitz & Weiss (1981), Myers & Majluf (1984), and Chirinko (1987) initiated financial constraint investment theory by showing how easily the well-known ‘MM theorem’ – proposed by Franco Modigliani & Merton Miller (1958, 1963) – can break down in real-world financial markets. One implication of the MM theorem, which Jorgenson and Tobin accepted, is that the opportunity cost of capital for a firm is independent of both its financial structure (i.e. debt-equity ratio) and its chosen mix of retained earnings, bond issues and share floats for financing new investment.

The financial constraint theories may be seen as confirming Kalecki’s principle of increasing risk, in that they imply a pecking order among sources of finance. Perched at the top of this financing hierarchy are retained earnings (least risky and cheapest), then share floats (which dilute equity) and, finally, bond issues (most risky and dearest). The key finding of the MM theorem was that firms can never increase their own capital value through purely financial operations because, if this were possible, rentiers could profit through arbitrage by replicating such operations in their own portfolios. To do this, the rentiers would need to possess precisely the same data as the managers of corporations.

However, just as in George Akerlof’s (1970) used car markets, access to key information in the financial markets is asymmetric. To compensate for their lack, or mistrust, of what information is available on the real investment opportunities confronting firms, rentiers tend to raise the price of external finance above the opportunity cost to managers of using cash flows generated within their own firms. Basically, rentiers cannot know the full range of risk classes (possible adverse selection), what action the firm’s managers will take (possible moral hazard with hidden action) or what outcomes are revealed by the firm’s monitoring of its own investment projects (possible moral hazard with hidden information), so they add a ‘lemons premium’ to the normal market-clearing borrowing rate.

In hindsight, it was Kenneth Arrow (1968) who initiated the option value investment theory by introducing the concept of irreversibility, whereby capital goods either cannot subsequently be resold to other firms or can be resold only at a significant loss. Thus investments which are more or less firm-specific may be classified as completely or partially irreversible. It was another 18 years before McDonald & Siegel (1986) highlighted the existence of a close analogy between the decision to make an irreversible real investment and the decision to exercise a financial option. Avinash Dixit & Robert Pindyck (1994) provide a systematic exposition of this neoclassical investment theory.

They point out that a call option gives its rentier owner the right to buy a financial asset at some predetermined price: once exercised, the option is ‘killed’ and becomes worthless. By analogy, a firm’s managers own the option to take advantage of an irreversible investment opportunity at any time after careful analysis of its time-profile of expected net proceeds has shown that $r_t^* \geq n$, or, equivalently, that $q_t \geq 1$. To build or purchase the necessary capital equipment immediately that opportunity is identified would kill the real option value of waiting, i.e. the benefits of postponing the investment until more information concerning future market conditions becomes available.
According to the ‘bad news principle’ of Bernanke (1983), good news is irrelevant to the real option value of an investment opportunity. In a world of uncertainty, there are positive probabilities of future upward or downward revisions to the profitability expected from any eligible investment project. But the option value of avoiding losses by waiting must increase if there is bad news. Good news has no effect on the option value because all it does is confirm the wisdom of investing now – which kills the option anyway.

Dixit (1992, p 123) used the bad news principle to explain why American companies were less aggressive investors than Japanese firms during that nation’s era of industry protection. The former faced downside risk – hence their option value of waiting to invest was always positive – whereas the latter were protected from losses by government supports. With an option value near zero, any Japanese firm that had identified an investment opportunity never waited.

The existence of a real option value of waiting drives a wedge into the right-hand side of the NPV rule. It may be, for instance, that a firm will maximise its value by investing in projects until \( r^* = 2n \), or, equivalently, until \( q_t = 2 \). For reasonable parameter values, McDonald & Siegel (1986) show that it is optimal to defer investing until the present value of a project’s benefits is twice as large as its capital cost. This represents an upper threshold for investment to occur immediately, say via entry of new firms, but the theory also posits a lower threshold such as \( r^* = 0.5n \), or \( q_t = 0.5 \) for disinvestment to commence.

Both the financial constraint and real option value theories are valuable for explaining why, contra the MM theorem, managers constantly worry about the financial structure of their firms, favour internal finance and only accept investment projects for which \( r^* >> n \), or \( q_t >> 1 \). These new approaches, therefore, are embellishments of, rather than replacements for, the perfected Jorgenson/Tobin neoclassical investment theory. As such, the insights they afford are equally relevant to all other investment theories that contain the profitability gap gene. The required embellishment is simply: alter (e.g. double when investing and halve when disinvesting) the risk premium \( \phi \) on the right-hand side of the gap between expected profitability and the opportunity cost of capital.

8. Concluding Remarks

Investment theory, particularly in relation to the explanation of business cycles, has long been a central concern of economic theorists ranging from Smith, Marx, Spiethoff, Keynes and Fisher, to Kalecki, Tobin, Jorgenson and beyond. It has evolved in diverse ways and generated a rich variety of perspectives. We have examined four influential approaches to investment theory - classical uniform profitability; Keynesian marginal efficiency; neo-Keynesian multiplier-accelerator; and neoclassical q-theory and user cost – to determine whether, amongst this profusion of ideas, they share a common trait. Depending on how one specifies the profitability expectation function of entrepreneurs, the conclusion of this paper is that they do. Their common gene is the profitability gap.

The generic investment function and its signature gene appear to qualify as one of the universals of economic science, equally at home as an explanator of investment, inflation/deflation and related cumulative processes. The gene is common to the ergodic neoclassical universe of general equilibrium in logical time and the non-ergodic post-classical universe of fully-adjusted stationary and steady states, plus disequilibrium traverse phenomena, in historical time. Thus, the differences between our various investment theory approaches are more apparent than real. A further implication is that future investment function proposals should be all the more critically examined if
they do not contain the profitability gap, since our research suggests it is universally considered to be one of the necessary explanators of fixed capital formation.
References


Endnotes

1 Under ‘rational’ expectations, the zero expected value of forecasting errors shows that all calculable risks have been taken into account; the same cannot be said of the irreducible uncertainties that enter into investors’ expectations of future profitability. Darwinian selective pressure implies a kind of inverse Gresham’s Law, with firms that are good forecasters of their own future profitability driving out the bad ones.

2 Keynes (1936, pp 135-46) called it the marginal efficiency of capital (MEC), but the context makes clear that he really meant the marginal efficiency of investment (MEI). This was pointed out by Abba Lerner (1944, 1953) and reinforced by Luigi Pasinetti (1974, pp 60-4).

3 Malinvaud’s ‘deviation’, ‘excess’, ‘pure profit rate’, and ‘difference’ are alternative names for the profitability gap: \( a_i = r^* - n_i \% \text{pa} \).

4 The ‘dot com’ stock market bubble that burst in 2000 is a prime example of these remarks.