Understanding investment better: insights from recent research

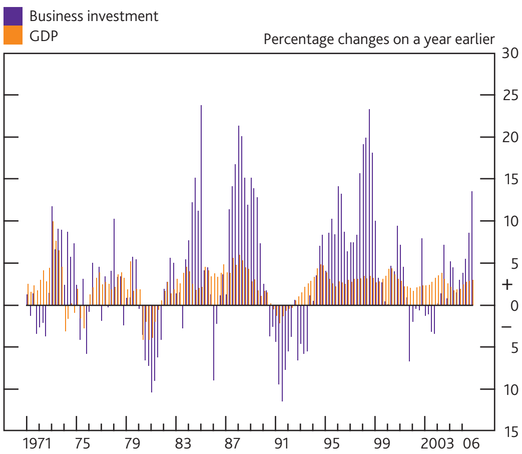
By Ursel Baumann and Simon Price of the Bank's Structural Economic Analysis Division.

Motivated by a number of puzzles about the recent behaviour of business investment in the United Kingdom (including the boom in the late 1990s and the prolonged weakness thereafter), this article brings together some of the main results of recent research on investment undertaken by the Bank and puts them into the wider context of the investment literature.

Introduction

A key influence on the short-term outlook for inflation, and thus on the decisions of monetary policy makers, is the balance of aggregate demand and supply in the economy. Business investment, currently representing around 60% of total investment, is an important component of both demand and supply.(1) Investment, net of depreciation, determines the growth rate of the capital stock — a fundamental driver of supply. On the demand side, business investment currently accounts for around 10% of GDP and the cyclical behaviour of investment makes a substantial contribution to fluctuations in GDP (**Chart 1**). Over the period 1971–2006, annual business investment growth was around twice as volatile(2) as that of household and government consumption, exports, imports and overall GDP. As a result, it is vital for policymakers to understand both long-run trends in investment and its fluctuations during the business cycle. While investment has always been a subject of intense academic research, there has been a renewed research effort over recent years by many people, including Bank staff.

Chart 1 Cyclical behaviour of business investment and GDP



The neoclassical theory (summarised in the box on page 234), is the key to understanding this area. According to that theory, investment is influenced by two main determinants in the long run: planned production levels and the real user cost of capital. Higher planned production levels lead to a rise in the desired capital stock, and thus to a rise in investment. A rise in the user cost of capital (which depends, in part, on the cost of finance) will reduce a firm's desired capital stock and thus lead to lower investment. The effect of this depends on the elasticity of substitution, which measures the ease with which firms can substitute between capital and other factors of production when producing output. Understanding these influences is important for policymakers, because monetary policy can influence investment indirectly via its influence on aggregate demand and directly via the cost of finance.

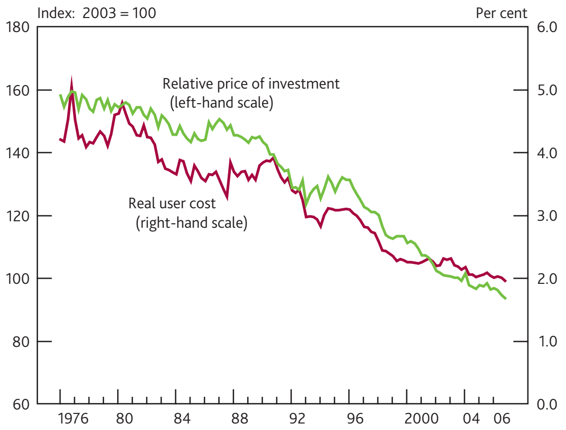
Neoclassical theory has proved very useful in understanding long-run investment behaviour. The major challenge has been to know what value the elasticity of substitution takes, and that has been the focus of much research, discussed below. But the basic neoclassical model does not completely explain the large short-run fluctuations in investment that are to be seen in the data, and understanding these is also important.(3) In the next section, there is a discussion of the ways in which the basic neoclassical theory may be extended. These are then examined in more detail in the rest of the paper.

(1) This article focuses on investment excluding inventories, the latter being another important source of business cycle fluctuations. Inventory investment is discussed in more detail in a previous article; see Elder and Tsoukalas (2006).
(2) Measured using the coefficient of variation, the standard deviation of a series relative to its mean.
(3) However, it should be noted that investment is very difficult to measure, and it is extremely susceptible to revisions, which can take several years to be finalised. While this is not considered in any depth in this article, it should be borne in mind that interpretations of the more recent data may change over time as the data are revised.

What may be missing from simple models of investment?

Basic neoclassical theory outlines long-run relationships between the underlying determinants. Long-run developments have been important in explaining investment over longer time periods, especially with respect to the user cost of capital. The user cost (the details of which are spelt out in the box on page 234) can be thought of as the amount that an owner of capital would pay if he or she were renting it, instead of owning it. In the past there was considerable controversy about the precise empirical effect on investment, but there is now more of a consensus, as explained below. **Chart 2** shows the strong trend decline in the user cost over several decades, and also illustrates how important the fall in the relative price of investment has been. Arguably, this largely explains the upward trend in the ratio of business investment to market sector output shown in **Chart 3**.(1)

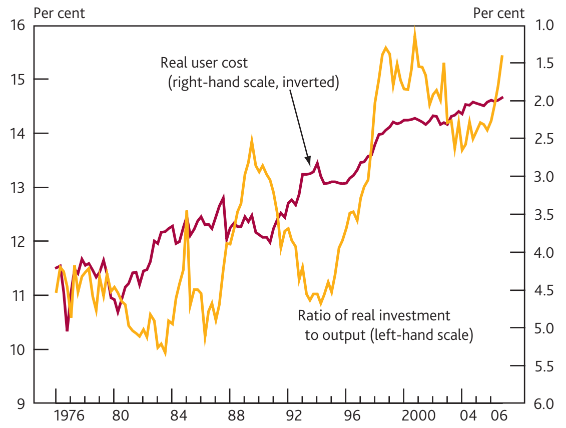
Chart 2 Real user cost of capital and relative price of investment to market sector output(a)



Sources: ONS and Bank calculations.

(a) Market sector output is a Bank estimate. The series is described in Churm *et al* (2006).

Chart 3 Real user cost of capital and ratio of business investment to market sector output



Sources: ONS and Bank calculations.

The likely impact of the real user cost on investment depends crucially on the value of the elasticity of substitution. The simple neoclassical model assumes one type of firm and one type of capital; but there might well be wide differences in the elasticity of substitution across firms, industries and over time. While the aggregate elasticity is most important for policy purposes, looking at disaggregated investment data and allowing the elasticity of substitution to differ across assets may improve our understanding of the behaviour of investment. This would be particularly relevant when the relative prices of various types of capital goods are changing, which is precisely what has been happening over the past few years, when the investment goods prices of information, communications and technology (ICT) goods have been in decline. This is discussed in more detail in the next section.

Although the long-run trend in business investment seems fairly well understood, **Chart 3** also suggests that deviations in the business-investment ratio from its long-run trend can be large in practice. For example, following a period of very sharp increases in investment between 1995 and 1999, growth in business investment fell back relative to that of market sector output. The investment to output ratio at constant prices fell from nearly 16% in 2000 Q4 to around 14% in 2005. Over the same period, the real user cost of capital continued to fall (although less sharply than over the 1990s), which would have been consistent with a further increase in the investment to output ratio.(2) More recently, during 2006, the ratio started to recover again. These deviations suggest there are gaps in the simple model. This is important, because the short-run dynamics of investment are a major influence on the business cycle, and thus affect the appropriate stance of monetary policy. Bank research on this issue has focused on three areas that extend the simple model.

The first relates to the detailed structure of adjustment costs, which delay the adjustment in capital (or investment) to the firm's desired level. The simplest neoclassical model is solely about the capital stock and investment simply occurs to bring capital to a new desired level after a change. But this is not realistic. It implies very large changes in investment flows. So early in the literature, adjustment costs were introduced to explain short-run dynamics in investment. But the specific way in which adjustment costs work can have important macroeconomic implications, as discussed further below.

The second research area relates to uncertainty, which may have an impact on the long run, but is likely to be more relevant for short-run dynamics. Recent years have included several periods of heightened uncertainty; for example,

(1) As the fall in the user cost was largely driven by a falling relative price of investment goods, the rise in real demand was offset in nominal terms by lower prices — more than offset, in fact, given the particular value of the elasticity of substitution, so that the ratio of *nominal* spending on investment to output has fallen.
(2) The key drivers during the 1990s behind the fall in the real user cost of capital are discussed in more depth in Bakhshi and Thompson (2002).

The neoclassical theory of capital

The neoclassical theory of capital (Jorgenson (1963)) predicts that profit-maximising firms will invest in the capital stock until the expected marginal return of a unit of capital equals its marginal cost. Assuming that capital goods are homogeneous, this condition results in a long-run 'steady-state' relationship between the firm's optimal capital stock, its two main determinants (planned production levels and the real user cost of capital), and the elasticity of substitution, as in equation **(1)**. Lower-case letters indicate logarithms, and the absence of a time subscript indicates the long run.

$$k = y + \theta - \sigma r, \text{ where } \theta \text{ is a constant} \qquad (1)$$

Higher planned production levels, y, lead to a rise in the desired capital stock, k. A higher user cost, r, means that the firm's desired capital stock falls, the extent determined by the elasticity of substitution, σ, measuring the ease with which firms can substitute between capital and other factors of production such as labour when producing a given level of output. The real user cost of capital, r, is a function of a number of other variables (equation **(2)**). It increases with the real cost of (equity and debt) finance, ρ, which represents the opportunity cost of holding capital rather than selling it and saving the proceeds. It also increases the higher the depreciation rate, δ, since the value of fast-depreciating assets falls more rapidly. An increase in the tax rate (net of subsidies) on investment, T, also leads to a rise in the cost of capital by making investment goods more expensive relative to other goods in the economy, as does a rise in the relative price of investment to output, P_k/P_y. Finally, the real user cost of capital falls with higher expected growth rates of the relative price of investment, π, capturing capital gains.

$$r = \frac{P_k}{P_Y}\left(\rho + \delta - \pi\right)\frac{1}{\left(1 - T\right)} \qquad (2)$$

To arrive at a relationship between investment and its underlying determinants, the neoclassical theory uses the fact that capital in any period is equal to the depreciated amount from the previous period plus gross investment (see equation **(3)**, the capital accumulation relationship).

$$K_{t+1} = (1 - \delta)\, K_t + I_t \qquad (3)$$

Firms bring the current capital stock back into line with its desired level by adjusting the level of investment, I. In the steady state the value of the capital stock is constant (or growing at a constant rate). Using this, and transformed into logarithms, equation **(3)** implies that equation **(4)** holds in the long run.

$$i = k + \gamma, \text{ where } \gamma \text{ is a constant} \qquad (4)$$

Equations **(1)** and **(4)** combined lead to a long-run relationship between the investment to output ratio and the real user cost of capital (equation **(5)**).

$$i - y = \theta + \gamma - \sigma r \qquad (5)$$

uncertainty associated with 9/11, the tension in the Middle East, and the large rise in oil prices. Since investment is at least partly irreversible, theory suggests that firms should delay investment when uncertainty about future pay-offs from their investment projects is high and this uncertainty is likely to be resolved at some stage in the future.[1] Events like 9/11 appear to have affected expectations of future investment growth: surveys taken by Consensus Economics shortly before and after the terrorist attacks in the United States show that the average forecast for UK whole-economy investment growth was revised down from 1.7% to 1.2% in 2001 and from 2.9% to 2.2% in 2002, perhaps because of higher uncertainty.

The third relates to the impact of financial constraints and balance sheet effects on business investment decisions. This is another factor that may help explain the weakness in investment growth during 2002–05, as companies may have been constrained by overly stretched balance sheets and pension fund deficits.

The remainder of this article summarises what we and other researchers have found in all these areas, regarding both the long and short runs, and how that has helped us to understand the behaviour of investment.

The long run: determining the elasticity of substitution

The most relevant elasticity of substitution for the policymaker is arguably that which holds in the long run.[2] As discussed further in Ellis and Groth (2003), for an elasticity of substitution below one, sustained falls in the real cost of capital will lead to proportionately small increases in the volume of investment relative to output. An elasticity of substitution of one means that a 1% fall in the real user cost of capital leads to a 1% increase in the investment to output

(1) See Dixit (1992) and Pindyck (1991) for introductions to the issues.
(2) In the short run, firms may find it hard to adjust the relative proportions of capital and labour, but their choices may widen given more time.

ratio. By contrast, if there were a zero elasticity of substitution no rise in investment volumes would occur when the real cost of capital falls, because firms are not able to substitute between factors of production.

Despite the importance of this elasticity, there has been some controversy about its empirical size. Chirinko (2006) reports a broad range of estimates from the literature, ranging between zero and 1.4. However, those estimates are based on different data sets and different countries. More coherent estimates based on a data set for US firms (see Chirinko *et al* (1999) and (2004)) still suggest a wide range of estimates of the elasticity of substitution, with the variation due to different estimation methods.

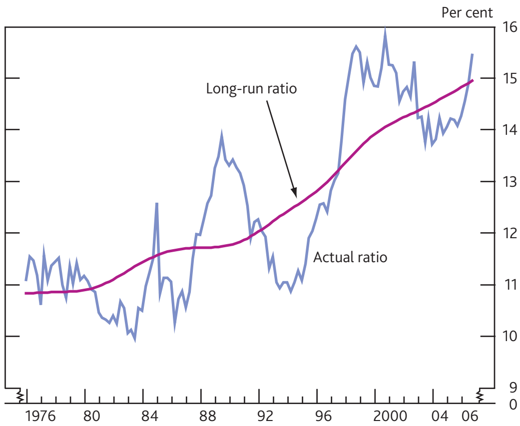
A fundamental difficulty is that the data used to estimate the elasticity may be far from the long-run 'steady state' (Chirinko (2006)). If the variation in the real user cost of capital is dominated by transitory movements, firms will expect shocks to the user cost to be quickly reversed and they will not react to such changes. As a result, this may lead to a downward 'bias' in the estimated long-run elasticity of substitution. One way this is dealt with in the literature is by exploiting long-run ('cointegrating') relationships between investment, capital, output and the real user cost of capital. Short-run deviations from the steady states are modelled by including past changes in those variables.

This approach has been widely used. Early empirical studies that tried to explain aggregate business investment within a neoclassical framework used a 'reduced-form' long-run relationship between investment, output and the real user cost of capital (as in equation **(5)** in the box on page 234), combining the long-run relationship between investment and capital and the determinants of the optimal capital stock. They generally failed to find a robust role for the user cost of capital in the United Kingdom. By contrast, Ellis and Price (2004) found a well-defined value for the critical elasticity. One innovation was that they estimated a model of both long-run relationships outlined in the box (equations **(1)** and **(4)**), using capital stock data and a long series for the cost of capital. But the main reason for their success in pinning down the elasticity was that by the time of this work there was a clear long-run trend evident in the user cost. They found a robust relationship between the real user cost of capital and aggregate business capital, with an estimated elasticity of substitution of about 0.45. Notably, that was also consistent with previous estimates for the elasticity of substitution for the United Kingdom obtained using the analogous labour demand relationship.(1)

Other approaches have also been adopted in order to avoid the downward bias. For example, Chirinko *et al* (2004) capture steady-state relationships by using time-averaged data over long periods for US firms. Recent work has followed this approach for the United Kingdom; see Barnes *et al* (2007). Additionally, in this work econometric methods are used that allow for differences in the dynamic relationships across firms to provide further evidence on the size of the elasticity of substitution. Their results lie in the region of 0.4, consistent with previous estimates based on aggregate data.

Overall, it can be concluded that recent research in this area provides relatively robust estimates of this aggregate parameter value for the United Kingdom. **Chart 4** compares the investment to output ratio to a calculated long-run equilibrium, given a smoothed series for the user cost and using the value 0.4 for the elasticity of substitution.(2) It indicates that given this value for the elasticity of substitution, the broad long-run trend in the ratio seems to be explained by the long-run decline in the real user cost.

Chart 4 Actual and long-run business investment to market sector output ratio



Sources: ONS and Bank calculations.

Despite this success at explaining the long-run trend, the assumption of a unique elasticity of substitution may be somewhat simplistic. This elasticity could, for example, be different *ex ante* (before the firm's decisions about how much to invest and how many workers to employ are taken) and *ex post* (once the capital good is installed). In the literature this assumption is known as 'putty clay'. The idea is that firms have a lot of flexibility in the choice of the proportion of capital to other factors in the planning stage (the proportion is soft 'putty') but once installed, there is much less flexibility (the proportion has hardened to 'clay'). It could also differ between different pairs of production factors (eg for capital against skilled labour, unskilled labour and oil) or between plants, firms and assets.

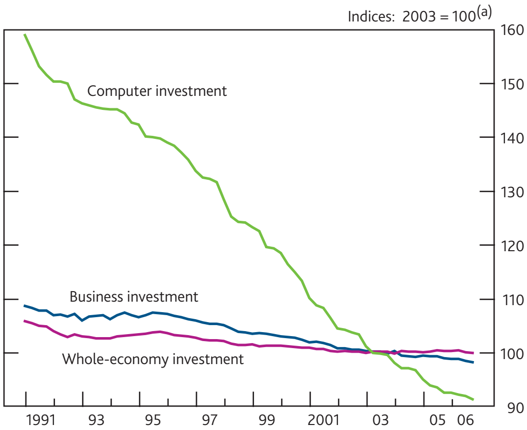
One simplification, that the elasticity is the same across all assets, has been particularly closely examined. The

(1) See for example Barrell and Pain (1997) who obtain very similar estimates for the United Kingdom.
(2) The long-run real user cost of capital is constructed using a Hodrick-Prescott filter designed to extract a smooth trend. The long-run investment to output ratio is from a simple regression of this ratio on a freely estimated constant and on the long-run user cost with an imposed elasticity of 0.4.

background is that recent research indicates that in the presence of diverging relative investment goods prices across assets, disaggregated models of investment may be superior to aggregate models, even if the elasticity is common across all assets. Tevlin and Whelan (2003) note that the increase in replacement investment associated with compositional changes in the capital stock towards assets with shorter-lives (such as computers) is not captured well by aggregate models. Bakhshi *et al* (2003) show that at the aggregate level, the standard measure of the real user cost of capital is mismeasured when there are trend declines in the relative price of investment goods across assets.

The issue of aggregation is particularly relevant for ICT compared with non-ICT assets, given that their relative prices have been diverging substantially in the past two decades. As shown in **Chart 5**, the relative price of computer investment has fallen much more sharply than that of aggregate (business or whole-economy) investment. The results by Tevlin and Whelan (2003) for the United States and (with rather weaker evidence) Bakhshi *et al* (2003) for the United Kingdom, both suggest that the elasticity of substitution differs between assets subject to rapid falls in relative prices, such as computers, and assets whose relative prices have remained more stable.[1] Both papers conclude that disaggregate models of investment can explain at least part of the investment boom of the second half of the 1990s.

Chart 5 Relative price of business, whole-economy and computer investment to GDP



(a) The chart is shown as a log index.

These encouraging results at the country level prompted additional work on whether the disaggregated approach can also better explain the recent investment behaviour in other advanced economies; see McMahon *et al* (2005). For a panel of the G7 countries and Australia, the authors estimate ICT and non-ICT investment equations. Their findings broadly confirm those conducted at the country level: while the estimated elasticity of substitution for aggregate investment and for non-ICT investment is low (between 0.0 and 0.5) and not statistically significant in some countries, the real user cost of capital proves important in determining ICT investment, with a much larger and significant estimated elasticity of 1.3. Out-of-sample forecasts of the disaggregated model of investment for the late 1990s were closer to actual outturns than the predictions by the aggregate investment model.

Further work in this field by Smith (2007) uses UK industry-level investment data. There may be an advantage to be gained from using data from a number of industries over time, as the additional cross-sectional information may improve the empirical estimates. This work finds that, while still being statistically distinct, differences in the elasticity of substitution across assets are smaller than suggested by the previous literature, but that again, the average value of these estimates is substantially less than unity.

In sum, this asset-level research appears better to explain trends in investment over the late 1990s. If more of this investment boom can be explained, this suggests that the 'capital overhang' resulting from it is likely to have been smaller than implied by aggregate models of investment, implying a period of below-average investment to come. The point here is that if high investment had led to installed capital exceeding the optimal level, there would be a period of relatively low investment while the gap between the actual and desired stocks of capital shrank. A question raised by McMahon *et al* (2005), however, is whether asset-level models can also explain the post-2000 slowdown in investment and the subsequent weakness in investment over 2002–05. Current results suggest that neither aggregate nor disaggregate models can fully explain the slowdown. This indicates that further aspects may be missing from standard models of investment, perhaps regarding the dynamic adjustment of investment in the short to medium term, and it is this which is now examined.

Short-run dynamics

Adjustment costs

An important feature of the data, omitted in the discussion above, is the relatively slow return of capital to its steady state in response to shocks. After a shift in the desired capital stock, the simple neoclassical model would predict an instantaneous jump in investment to restore capital to its equilibrium value. Instead, lagged responses are observed. To address this, the investment literature long ago introduced adjustment costs that model inertia in the adjustment of capital and lead to long periods where capital is in disequilibrium.[2] These can be

(1) It may be that firms may react more strongly to changes in the cost of ICT relative to non-ICT capital because they perceive shocks to ICT prices that result from technological innovations to be of a more permanent nature: see Tevlin and Whelan (2003).
(2) See for example Shapiro (1986), or Hamermesh and Pfann (1996) for a survey of the adjustment cost literature. The logical alternative explanation of smooth investment would be that both drivers of the capital stock (the user cost and planned output) invariably move smoothly. But this is not realistic, and does not explain the volatility in investment.

introduced into the neoclassical model by assuming that, in addition to the standard costs of hiring labour and buying or renting capital, firms face costs when the level of capital changes. Such adjustment costs can, for example, take the form of disruption costs related to temporary interruptions to production while installing new machines or moving to new premises. Another example might be the learning and implementation of new technologies that can occur while new investments are made, which thereby involve not only new plant but also new working methods and investment in human capital.(1) These costs are generally assumed to rise more rapidly as the rate of adjustment increases, resulting in rapid accumulation of capital being very costly, so that slow and continual capital adjustment is generally preferred.(2)

Previous empirical studies in this field, which have mainly focused on the United States, have found evidence of sizable costs to adjusting the level of capital (see Chirinko (1993) for an overview). Recent work by Groth (2005) provides an estimate of the size of capital adjustment costs for the United Kingdom, using industry-level data that cover both manufacturing and services sectors. She finds that the 'half-life' (the period after which 50% of the adjustment to the long run has occurred) of the adjustment in capital following a shock to the user cost of capital is about three years. This is slower than that reported for the United States by Shapiro (1986), but faster than that typically found in the Tobin's Q(3) literature, where Summers (1981), for example, finds a half-life of around 20 years (again for US data).

Groth (2005) also looks at a disaggregated set of UK industries and finds that there are significant costs of adjusting non-ICT assets, while there is less evidence of ICT capital adjustment costs.(4) The results also indicate that it may be more costly to adjust capital in the services sector than in manufacturing.(5) Given the large and increasing share of the services sector in total output, this result could also point to an increase in the importance of aggregate capital adjustment costs over time.

While capital adjustment costs can thus lead to a slow return of capital to its equilibrium following a shock, it has been argued they cannot by themselves explain other features of the observed dynamics of investment and output. One such feature is the 'hump-shaped' response of investment to monetary policy shocks that can be inferred from empirical models: following a monetary tightening, investment falls, with the peak impact occurring only after several quarters, before investment gradually returns to its pre-shock level.(6) In order to model this behaviour in macroeconomic models, the recent literature has introduced *investment* rather than *capital* adjustment costs, where there is a cost to changing the level of investment, as opposed to the level of capital. In contrast to capital adjustment costs, investment adjustment costs depend positively on the change in current relative to lagged investment. They can be interpreted as representing the inflexibility in changing the pattern of investment during the planning phase: eg, see Christiano and Todd (1996). For example, once planning permission has been obtained and architectural plans developed, a change in the investment plans would constitute considerable additional costs. Such costs induce inertia in investment itself, causing it to adjust slowly to shocks. When they are present, Christiano, Eichenbaum and Evans (2005) show that a model where prices adjust slowly can generate hump-shaped investment dynamics, consistent with the estimated response of investment to a monetary policy shock.

Against this background, Groth and Khan (2007) establish some evidence regarding the empirical importance of investment adjustment costs. Using industry data for the United States and the United Kingdom, they employ a framework that allows for both types of adjustment costs — investment and capital. The authors' findings point to little evidence in favour of investment adjustment costs for the United Kingdom, while capital adjustment costs are significant. For the United States, the results are more mixed: there is some evidence that investment adjustment costs may occur, but the effects are small. They conclude that while investment adjustment costs may appear to improve existing macroeconomic models, there is not really strong evidence for such costs in the disaggregated investment data.

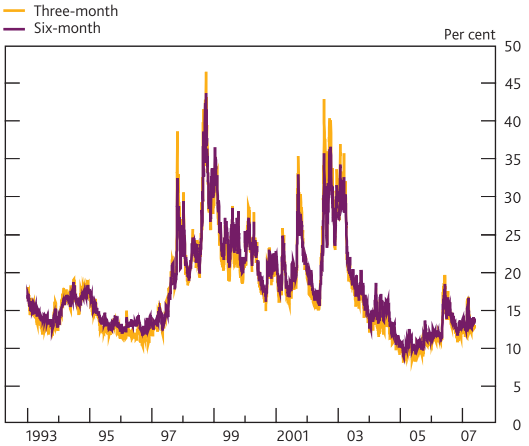
To summarise, adjustment costs can generate a slow return of capital to its steady-state values, and these can apply to either investment or the capital stock. It seems that, especially in the United Kingdom, the evidence favours capital adjustment costs over investment adjustment costs. However, it is the latter that have tended to be introduced into macroeconomic models, where they have sometimes been used to help better explain the response of investment to monetary policy shocks. The implication is that it is necessary to look elsewhere to explain this feature of the data.

(1) These would be examples of so-called 'internal adjustment costs', as the costs are internal to the firm. The literature distinguishes between those and 'external adjustment costs'. The latter assume that the firm that invests has to pay a higher price for more capital — the supply curve is upward sloping. However, from a macroeconomic perspective, the distinction between internal and external adjustment costs should not matter.
(2) See Hamermesh and Pfann (1996) for a discussion of the effects of various assumptions about the nature of adjustment costs.
(3) Tobin's 'Q' theory lays out a theoretical link between investment and expected future profitability of firms, which can be derived from the neoclassical model of investment with adjustment costs. Q is the ratio of a firm's value to the cost of replacing its capital. The theory states that a company should invest when the discounted value of future profits from an extra unit of capital exceeds the cost of acquiring it, which is equivalent to marginal Q being larger than one. Under certain assumptions, this can be approximated by the more easily measured average value of Q.
(4) These results differ from earlier results found for the United States cited in Groth (2005). They may reflect mismeasurement due to uncertainty regarding UK software investment and ICT prices.
(5) Perhaps contrary to expectations, many service industries are relatively capital-intensive.
(6) The method used is generally Structural Vector Autoregressions looking at a small number of macroeconomic variables, where theory is used to identify particular shocks and the dynamic 'impulse responses' of variables to these shocks. See Christiano *et al* (2005), for example.

Investment under uncertainty

A further feature that may lead to short-term movements in investment is changes in uncertainty. **Chart 6**, showing one market-based measure of volatility, suggests that uncertainty can vary widely over time. When making investment decisions, companies face considerable uncertainty regarding future costs and demand. Some aspects of uncertainty — those affecting the rate of return on capital required by financial markets — are implicitly captured in the neoclassical investment theory through their effect on the real user cost of capital, by increasing the required rate of return. But uncertainty can matter for companies' investment decisions beyond this effect.

Chart 6 FTSE 100 implied uncertainty[(a)]



Sources: Euronext.liffe and Bank calculations.

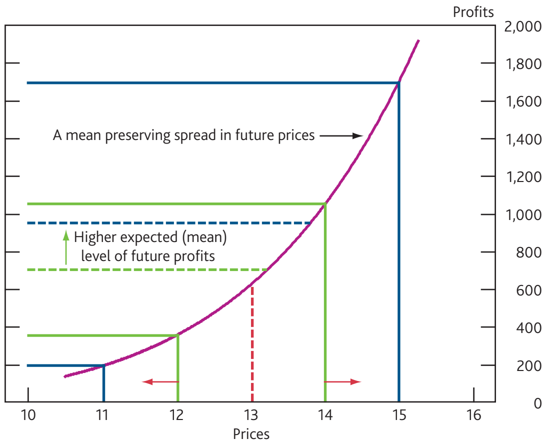
(a) Option implied volatility. Calculated from the distribution of returns implied by three-month/six-month option prices.

The long-run impact of uncertainty

We begin by clarifying what effect uncertainty is expected to have on the long-run capital stock.

If companies wish only to maximise expected (average) profit, then they will not care about risk in itself (are 'risk neutral'); it is only the average return they care about. However, Hartman (1972) showed that, for a given discount rate, in profit-maximising (and hence risk-neutral) competitive firms with constant returns to scale, increased output price uncertainty will *increase* the optimal capital stock, as long as the capital stock is fixed in the short run — which is surely realistic. This follows if profits are a convex function of prices (ie successive price changes result in more than proportional increases in profits), as they will be under perfect competition. Such a case is illustrated in **Chart 7**. To understand this, think what would happen if the price at which the firm can sell its output rose, and the firm responded by using exactly the same inputs of labour and capital as before (producing the same quantity of goods). In that case profits would rise in proportion to the rise in price — in other words, in a straight-line relationship, as costs are the same but the price is higher. But no firm would do this because, although by assumption capital is fixed, firms can hire more labour to produce more to sell at the higher price (and under perfect competition would be able to hire that labour at the same wage), so profits must rise by more than the straight-line benchmark. The more the price rises, the more the firm has this incentive. What may be less obvious is that given this shape of the curve, a widening of the distribution of prices raises expected profits. Essentially, higher prices raise profits by more than lower prices reduce them, as can be seen from the slope of the curve. So on average a wider price range raises profits.[(1)] The chart gives an example of such a mean-preserving change in the distribution of prices, in this case from an equal probability of the price being either 12 or 14, to either 11 or 15 (so the average remains at 13). Although the mean price remains constant, expected profits rise, as indicated by the dotted horizontal lines on the chart. And as the marginal profitability of capital has risen, the optimal quantity of capital stock will increase.

Chart 7 Profits as a convex function of prices[(a)]



(a) The chart shows a widening of the distribution of prices from (12, 14) to (11, 15), leaving the mean unchanged at 13. The actual profits at each of the prices are shown by the horizontal solid lines. The dotted horizontal lines show the expected levels of profits corresponding to each pair of prices. Expected profits rise as the distribution widens.

However, the convexity of the profit function may be reduced, and possibly reversed, by introducing imperfect competition (Caballero (1991)). Firms facing a downward-sloping demand curve will then only be able to increase output at the cost of lower prices, so the marginal profit following such a rise, all things being equal, is lower than under perfect competition (when prices are given). Similarly, if the firm faces an upward-sloping supply curve, wages will have to rise if the firm hires more workers, reducing marginal profits. If there are decreasing returns, these tendencies are aggravated, as then marginal productivity will decrease with scale (increasing returns will have the opposite effect). All these factors move the firm's price-profit relationship back towards the straight line, and may even push us beyond it. So there is ambiguity

(1) This is an example of 'Jensen's inequality', which has other applications in economics; for example to do with risk aversion.

about the effect of uncertainty on investment in the long run.(1) In the rest of this section the focus is therefore on the impact of uncertainty on the timing of firm's investment decisions as there is less ambiguity in the literature regarding this prediction, and since the empirical literature suggests that this is also where the major impacts on investment lie (see, eg, Bloom *et al* (2007)).

Irreversibility and timing of investment

The effects on timing are most marked when investment projects are irreversible, which is a plausible assumption. Once capital is installed, it is not easily uninstalled, and investment also has a low resale value. In the context of the previous section, irreversibility can be interpreted as representing asymmetric adjustment costs, where the cost of reducing the capital stock exceeds the costs of augmenting it. Such a feature is not captured by traditional investment models, which implicitly assume that investment projects are fully reversible. Uncertainty matters because if the decisions of whether or not to invest can be postponed to a later date, when more information about future demand and price outturns are known, firms may be better able to discriminate between profitable and unprofitable investment opportunities. Undertaking the investment destroys this valuable option, where value rises with uncertainty. Thus these models are often referred to as 'real option' models: see Dixit and Pindyck (1994) for a clear and comprehensive introduction. The point is that an immediate investment extinguishes the value of the option to wait, and this lost value should form part of the opportunity cost of investing. The main message is that more uncertainty may lead to delayed investment.

Bloom's (2007) model of firms facing investment irreversibility and adjustment costs examines this incentive to delay investment. He simulates a sharp rise in macroeconomic uncertainty. In the model a temporary slowdown in investment rates can be observed, followed by a rapid rebound, which seems to match the actual data. The empirical prediction of the real option models is thus that investment may occur in 'bursts' following periods of no investment, suggesting that at times, when business investment is below certain threshold levels, it may be unresponsive to the user cost.

Most of the aggregate empirical literature is consistent with a short-run negative effect of uncertainty on investment in the United States and United Kingdom (see Carruth *et al* (2000) for a survey), using a wide range of models and proxies for uncertainty. By contrast, the evidence using disaggregated data (which have largely focused on the United States) is less conclusive, perhaps because there is heterogeneity of effects across industries and firms, as some of the richer models suggest. Nevertheless, firm-level data may be the best way to seek insights into models of irreversible investment under uncertainty, precisely because of this heterogeneity.

Bloom *et al* (2007) follows this route for the United Kingdom, by looking at the effects of uncertainty (measured as standard deviation of daily stock returns) on investment spending of UK firms between 1972 and 1991. The findings support the predictions from the real options theory that investment responds (in a non-linear fashion) less to demand shocks at higher levels of uncertainty. The size of their estimates suggests that aggregate shocks — like the OPEC oil shocks — can seriously reduce the responsiveness of investment to demand in the short run.

Earlier work by Bond and Cummins (2004) had applied this approach to US data, but controlling for the level of expected profitability on investment.(2) The authors found that in this sample uncertainty helps explain investment over and above the level of future profitability in the United States, and that investment responds less to demand shocks in the short run when uncertainty is high. Recent work applied this methodology to the United Kingdom, using a firm-level data set of around 650 quoted non-financial companies for the period 1987–2000; see Bond *et al* (2005). The authors' empirical set-up allows the distinction between temporary effects of uncertainty (measured using the volatility of firm's stock market returns and by the dispersion of Institutional Brokers' Estimate System (IBES) analysts' earnings forecasts) on investment and long-run effects on the capital stock. The estimates are sizable, suggesting that a 10% increase in uncertainty implies a 4.4% reduction in investment rates in the short run. The authors also find that the capital stock falls in the long run if high levels of uncertainty are sustained. However, unlike other studies, they do not find that investment reacts less strongly to demand shocks when uncertainty is high, as predicted by the real options models.(3)

As the discussion above has revealed, accounting for uncertainty can have important effects on business investment spending. In particular, it could be a factor in explaining the weakness in business investment during the period 2002–05 inasmuch as a number of the shocks mentioned above may have led to sharp rises in uncertainty. Aggregate measures of uncertainty increased following some of these shocks (**Chart 6**), although volatility fell back after 2003. There is also some evidence that firm-specific uncertainty has risen since 2000; see Parker (2006).

Uncertainty may also be essential in understanding the sensitivity of investment to monetary policy changes. If the

(1) See Abel and Eberly (1999) and Caballero (1999) for further discussion of some other sources of ambiguity.
(2) Controlling for expected profits in this way means that any effects of uncertainty are in addition to the long-run effect of investment explained previously.
(3) See references quoted in Bond *et al* (2005).

channel based on the options theory outlined above is important, then there may be periods of high uncertainty when investment is not sensitive to monetary policy. Alternatively, if the uncertainty effect largely operates via its impact on the level of profitability of firms, then the monetary policy transmission to investment continues to be effective under uncertainty.

Financial constraints

A large body of the finance literature argues that business investment may also be influenced by cash flow and other balance sheet considerations that are not captured in the basic neoclassical model.(1) Firms that face financial constraints usually pay a premium for external sources of finance, and so prefer to use internal funds. As a result, firms may forego investment opportunities when faced with adverse cash flow. Similarly, the pressure experienced by firms with a large financial burden resulting from interest payments on high debt levels may also temporarily depress investment. Financial variables may thus also be relevant in explaining aggregate investment flows.

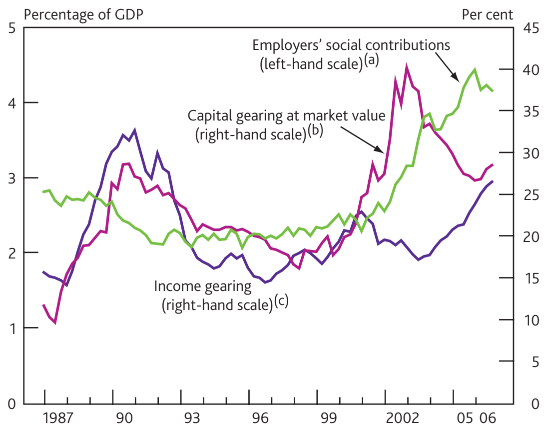
Bond *et al* (2004) investigate the importance of cash flows for investment decisions by UK firms. They estimate an equation where the ratio of investment to capital is a function of expected profitability, derived from the Tobin's 'Q' theory. In addition to expected profits, the authors include firms' cash flows in the equation, as the existence of financing constraints would imply that the level of investment also depends on the availability of internal funds. However, the findings indicate that firms' cash flows are only relevant in explaining investment when expected profitability is measured by Tobin's Q. A more direct measure of expected profits based on analysts' earnings forecasts results in cash flows becoming insignificant, suggesting that Tobin's Q does not adequately control for expected profits. Rather than providing evidence of financing constraints, cash flows may provide additional information about expected profitability that is not captured by the simple measure most easily available — the ratio of the firm's market value to the value of the capital stock.

The impact of corporate balance sheet adjustments on investment and financial decisions by UK firms in a broader sense is the topic of a paper by Benito and Young (2002). The authors examine the behaviour of dividends, new equity issuance and investment at the firm level as a function of company financial characteristics, assuming that firms are bound by a budget constraint that links the sources of their funds with their uses. The findings suggest a significant effect of financial pressure, defined as the ratio of interest payments to profits, in reducing investment.

Further evidence on the impact of a specific source of financial pressure — the contributions to company pension schemes — on investment (and dividends) of UK non-financial companies is provided by Bunn and Trivedi (2005). The advantage of using pension contributions to test this mechanism, is that companies are committed to raising these in line with regulatory requirements when the value of assets or liabilities change, thus providing a source of financial pressure that is independent of the firm's other decisions about its capital structure. The results are consistent with some impact of increased pension contributions in reducing both dividends and investment, although the effect on investment is only just significant.(2) Some survey evidence suggesting that the effect was small was given by the Bank's regional Agents in 2006.(3) The Agents also reported that small firms were affected more, which is consistent with the idea that the external finance premium is larger for small companies.

So financial constraints faced by firms can depress investment. But were they a factor in explaining the weak investment of 2002–05? **Chart 8** shows that some indicators of financial pressure on firms rose quite sharply after 1999–2000, which may have been a contributing factor in the weakness in investment during 2002–05. But other measures of financial conditions remained buoyant: overall firm liquidity (eg growth in M4 deposits of private non-financial companies) has been relatively high, and the overall cost of capital has been low by historical standards. Consistent with this business investment growth was strong in 2006.(4) So if anything, overall financial conditions appear to have supported, rather than constrained, investment growth more recently.

Chart 8 Indicators of company financial pressure



(a) Employers' contributions to social insurance schemes such as pensions. These data exclude National Insurance contributions.
(b) Private non-financial corporations' debt, net of liquid assets as a percentage of companies' market valuation.
(c) Private non-financial corporations' interest payments as a percentage of gross operating surplus excluding the alignment adjustment.

(1) See Myers (2001) for a survey of the literature on corporate capital structure.
(2) The evidence is weaker than in Rauh (2006) for the United States, which may be accounted for by differences in the quality of measurement of financial pressure used in the papers.
(3) Bank of England *Inflation Report*, August 2006, pages 14–15.
(4) Bank of England *Inflation Report*, May 2007, pages 18 and 20.

Conclusions

Some advances have been made in understanding investment in the past decade. First, it is possible to be reasonably confident that the business investment rate in the United Kingdom does not react one-for-one with changes in the real user cost of capital in the long run. Instead, the aggregate elasticity is well below unity, at about 0.4. Given this, broad trends in the business investment to output ratio can be well explained by changes in the real user cost of capital. This judgement is quite different to what would have been concluded two decades ago. At that point there was a scarcity of evidence that the user cost affected investment, thus shedding doubt on a key part of the monetary transmission mechanism. Second, when the prices of investment goods are diverging across assets, such as ICT and non-ICT, disaggregated models of investment may be empirically superior to aggregate models. By being better able to evaluate where investment and capital are relative to their equilibrium values using these long-run models, it is possible to have a better understanding of the sustainability of current investment trends. Third, short-run factors determining investment are vital in explaining the slow return of capital to its equilibrium. The empirical evidence suggests that capital adjustment costs can lead to capital disequilibria persisting over many years in the United Kingdom. Option-based theories have shown that higher uncertainty can lead to short-run adverse effects on investment, while financial constraints can also be relevant.

It should be said that the factors discussed in this paper are probably not sufficient entirely to explain the weakness in business investment during 2002–05. The MPC have highlighted a number of considerations that could have influenced business investment, not all of which have been discussed here.(1) One such factor is that there may well be future revisions to the investment data, which is particularly susceptible to revision. So it may be discovered that recent developments differed from what is currently believed. Another data-measurement issue, albeit somewhat more subtle, is that there may be some spending on intangibles investment which is not currently recorded in the official statistics. HMT and the ONS are currently working to create estimates for the United Kingdom. There may also be effects running from globalisation; it is possible, for example, that multinational firms might have decided to allocate more of their investment spending to overseas projects. So no one would claim to have all the answers; but, as should be clear from this article, several of the unresolved issues are now better understood.

Overall, the research described above has substantially improved understanding of recent investment trends, and therefore of the balance between aggregate demand and supply, a key factor behind changes in inflation. To give a concrete example, there was an investment boom during the late 1990s in both the United Kingdom and the United States. In the aftermath, it was suspected that there was a sizable ‘capital overhang’ requiring a prolonged period of below-average investment. But the disaggregated models of business investment predicted at least part of that boom. This makes it clear how useful such analysis may be for the policymaker. And as another example, the weakness in investment during 2002–05 may partly have been caused by higher uncertainty, perhaps due to volatile energy prices in the latter part of this period (although if so this did not appear in implied stock market volatilities), and perhaps by the change in firms’ balance sheets.

(1) See, for example, *Minutes of Monetary Policy Committee meeting 11 and 12 January 2006* at www.bankofengland.co.uk/publications/minutes/mpc/pdf/2006/mpc0601.pdf.

References

Abel, A B and Eberly, J C (1999), 'The effects of irreversibility and uncertainty on capital accumulation', *Journal of Monetary Economics*, Vol. 44, pages 339–77.

Bakhshi, H, Oulton, N and Thompson, J (2003), 'Modelling investment when relative prices are trending: theory and evidence for the United Kingdom', *Bank of England Working Paper no. 189*.

Bakhshi, H and Thompson, J (2002), 'Explaining trends in UK business investment', *Bank of England Quarterly Bulletin*, Spring, pages 33–41.

Barnes, S, Price, S and Sebastia-Barriel, M (2007), 'The elasticity of substitution: evidence from a UK firm-level data set', presented at 2007 AEA Conference, www.aeaweb.org/annual_mtg_papers/2007/0105_1015_1303.pdf.

Barrell, R and Pain, N (1997), 'Foreign direct investment, technological change, and economic growth within Europe', *Economic Journal*, No. 107, pages 1,770–86.

Benito, A and Young, G (2002), 'Financial pressure and balance sheet adjustment by UK firms', *Bank of England Working Paper no. 168*.

Bloom, N (2007), 'The impact of uncertainty shocks: a firm-level estimation and a 9/11 simulation', www.stanford.edu/~nbloom/ImpactUncertaintyShocks.pdf.

Bloom, N, Bond, S and Van Reenen, J (2007), 'Uncertainty and investment dynamics', *Review of Economic Studies*, Vol. 74, pages 391–415.

Bond, S and Cummins, J (2004), 'Uncertainty and investment: an empirical investigation using data on analysts' profits forecasts', Board of Governors of the Federal Reserve System, *Finance and Economics Discussion Paper no. 20*.

Bond, S, Klemm, A, Newton-Smith, R, Syed, M and Vlieghe, G (2004), 'The roles of expected profitability, Tobin's Q and cash flow in econometric models of company investment', *Bank of England Working Paper no. 222*.

Bond, S, Moessner, R, Mumtaz, H and Syed, M (2005), 'Microeconometric evidence on uncertainty and investment', IFS, *mimeo*, www.ifs.org.uk/publications.php?publication_id=3278.

Bunn, P and Trivedi, K (2005), 'Corporate expenditures and pension contributions: evidence from UK company accounts', *Bank of England Working Paper no. 276*.

Caballero, R J (1991), 'On the sign of the investment-uncertainty relationship', *American Economic Review*, Vol. 81, pages 279–88.

Caballero, R J (1999), 'Aggregate investment', in Taylor, J B and Woodford, M (eds), *Handbook of macroeconomics*, Vol. 1B, Amsterdam, North-Holland.

Carruth, A, Dickerson, A and Henley, A (2000), 'What do we know about investment under uncertainty?', *Journal of Economic Surveys*, Vol. 14, No. 2, pages 119–53.

Chirinko, R S (1993), 'Business fixed investment spending: modeling strategies, empirical results, and policy implications', *Journal of Economic Literature*, Vol. 31, No. 4, pages 1,875–911.

Chirinko, R S (2006), 'σ: the long and short of it', prepared for the Goethe University Frankfurt/CES Conference, 'A bright future at the age of 50 — The CES production function in the theory and empirics of economic growth', *mimeo*, Emory University.

Chirinko, R S, Fazzari, S M and Meyer, A P (1999), 'How responsive is business capital formation to its user cost?: An exploration with micro data', *Journal of Public Economics*, Vol. 74, pages 53–80.

Chirinko, R S, Fazzari, S M and Meyer, A P (2004), 'That elusive elasticity: a long-panel approach to estimating the capital-labor substitution elasticity', *CESifo Working Paper no. 1240*.

Christiano, L J, Eichenbaum, M and Evans, C (2005), 'Nominal rigidities and the dynamic effects of a shock to monetary policy', *Journal of Political Economy*, Vol. 113, No. 1, pages 1–45.

Christiano, L J and Todd, R M (1996), 'Time to plan and aggregate fluctuations', *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 20, No. 1, pages 14–27.

Churm, R, Mahajan, S, Maitland-Smith, F, Srinivasan, S, Thomas, R and Tily, G (2006), 'Measuring market sector activity in the United Kingdom', *Bank of England Quarterly Bulletin*, Vol. 46, No. 4, pages 404–14.

Dixit, A K (1992), 'Investment and hysteresis', *Journal of Economic Perspectives*, Vol. 6, No. 1, Winter, pages 107–32.

Dixit, A K and Pindyck, R S (1994), 'Investment under uncertainty', Princeton, Princeton University Press.

Elder, R and Tsoukalas, J (2006), 'Investing in inventories', *Bank of England Quarterly Bulletin*, Summer, pages 155–60.

Ellis, C and Groth, C (2003), 'Long-run equilibrium ratios of business investment to output in the United Kingdom', *Bank of England Quarterly Bulletin*, Summer, pages 177–87.

Ellis, C and Price, S (2004), 'UK business investment and the user cost of capital', *Manchester School*, Vol. 72, pages 72–93.

Groth, C (2005), 'Estimating UK capital adjustment costs', *Bank of England Working Paper no. 258*.

Groth, C and Khan, H (2007), 'Investment adjustment costs: an empirical assessment', *Bank of England Working Paper*, forthcoming.

Hamermesh, D S and Pfann, G A (1996), 'Adjustment costs in factor demand', *Journal of Economic Literature*, Vol. 34, No. 3, pages 1,264–92.

Hartman, R (1972), 'The effects of price and cost uncertainty on investment', *Journal of Economic Theory*, Vol. 5, pages 258–66.

Jorgenson, D (1963), 'Capital theory and investment behaviour', *American Economic Review*, Vol. 53, pages 247–56.

McMahon, M, Sterne, G and Thompson, J (2005), 'The role of ICT in the global investment cycle', *Bank of England Working Paper no. 257*.

Myers, S (2001), 'Capital structure', *Journal of Economic Perspectives*, Vol. 15, pages 81–102.

Parker, M (2006), 'Diverging trends in aggregate and firm-level volatility in the UK', *Bank of England, External MPC Unit Discussion Paper No. 16*.

Pindyck, R S (1991), 'Irreversibility, uncertainty and investment', *Journal of Economic Literature*, Vol. 29, pages 1,110–48.

Rauh, J (2006), 'Investment and financing constraints: evidence from the funding of corporate pension plans', *Journal of Finance*, Vol. 61, No. 1, pages 33–71.

Shapiro, M (1986), 'The dynamic demand for capital and labor', *Quarterly Journal of Economics*, Vol. 101, pages 513–42.

Smith, J (2007), 'That elusive elasticity and the ubiquitous bias: is panel data a panacea?', *Bank of England Working Paper*, forthcoming.

Summers, L (1981), 'Taxation and corporate investment: a q-theory approach', *Brookings Papers on Economic Activity*, Vol. 1, pages 67–140.

Tevlin, S and Whelan, K (2003), 'Explaining the investment boom of the 1990s', *Journal of Money, Credit and Banking*, Vol. 35, pages 1–22.