15 Theories of Investment Expenditures

Chapter 15 Contents

A. Topics and Tools

Investment expenditures play a key role in many theories of the business cycle, including Keynes's theory. Macroeconomic theorists have agreed on a basic framework that models the investment strategy of a profit-maximizing firm. However, empirical evidence has failed to provide substantial support for this model, which has been a source of frustration for those involved in investment modeling.

In this chapter, we will investigate some of the details of the investment process, including how firms raise the funds they use for investment, how models of the various kinds of investment differ, and how financial markets affect investment expenditures.

The Romer text develops the model in the form known as the q theory of investment. This is based on a measure of the desirability of investment known as Tobin's q. The q theory is easily reconciled with other approaches to investment, all of which lead to the same basic result.

B. How Firms Invest

What investment is and what it is not

The term "investment" means something different to macroeconomists than it does to most of the rest of the world. For example, if you ask your banker about investment, she will probably start talking about stocks and mutual funds that she would like you to purchase, or new kinds of deposit accounts that her bank offers. To an economist, these purchases of ***financial assets*** are not investment. From an aggregate point of view, financial assets do not represent real net wealth for the economy as a whole; instead, they reflect credit relationships within the economy. Financial assets such as loans and bank accounts represent contracts to pay interest and repay principal on borrowed money. Stocks represent partial ownership of a corporation, implying a right to vote on the governance of the corporation and to receive dividends as determined by the directors that the shareholders elect. In either case, the financial asset of one individual in the economy is offset by a financial liability of another person or corporation. Thus, when we aggregate the wealth of all members of the economy, these assets and liabilities cancel out and financial assets disappear. Thus, if you "invest" in a financial asset, someone else is "disinvesting" at the same time, so aggregate, or social, capital does not rise.

Macroeconomists reserve the term ***investment*** for transactions that increase the magnitude of ***real aggregate capital*** in the economy. This includes mainly the purchase (or production) of new real durable assets such as factories and machines.[1] The category of investment that receives the most attention is ***business fixed investment***, which is the purchase of new structures and equipment by business firms for production purposes. However, there are two other categories, and they are also important.

Inventory investment consists of increases in stocks of unsold goods or unused input materials. This kind of investment is quite different from business fixed investment because inventory capital normally has a very short life span. When inventories decrease from one period to the next, as sometimes happens even at an aggregate level, inventory investment is negative. Another unique feature of inventory investment is that it often occurs unintentionally. Unsold products are counted as inventory investment whether the firm bought them intending to build up its inventory or simply ended up selling less than it expected to sell.

Residential investment consists of purchases of new housing units, whether by firms or households. As discussed in the consumption chapter, new home purchases by households are counted as investment, with a monthly rental flow of housing services counted under consumption of services. Like inventory investment, residential investment tends to behave quite differently in some ways than business fixed investment.

Three other categories of investment are worthy of mention, though none of them is categorized as investment in the U.S. National Income and Product Accounts. One is investment in ***human capital***. Education could easily be classified as investment, but the U.S. income accountants do not do so. A second is consumer purchases of ***durable goods***. Refrigerators and automobiles have all the characteristics of capital and are so classified when purchased by firms, but in the United States they are classified under consumption rather than investment. Finally, ***government investment*** in roads, bridges, buildings, and other durable assets are classified as government spending rather than investment in the United States.

The theories discussed in Romer's text apply mainly to business fixed investment. To supplement these theories, Section D below will present prominent theories of inventory and residential investment. These theories are also examined briefly in Mankiw's text.

Financing of investment

Where do/should firms get the funds with which to make investment expenditures? The answers to this question are the subject matter of ***finance***, which is one of the main disciplines taught in academic schools of business. There are three main sources of investment funds for firms: (1) ***internal funding*** using accumulated profits,

---

[1] Note that we exclude resale transactions involving existing assets for the same reason we exclude purely financial transactions—the "investment" by the buyer is exactly offset by the "disinvestment" of the seller.

(2) ***borrowing*** either from banks or through the issue of financial assets such as (long-term) bonds or (short-term) commercial paper, and (3) ***issuing new shares*** of stock—new "equity."[2]

Each of these funding methods imposes explicit and/or implicit costs on the firm. If the firm borrows in order to finance its investment, it pays an explicit interest cost. If it uses internal funds for investment, it is forgoing other uses of those funds. Had the firm not used the internal funds for new capital, it could have earned interest on the funds by lending them or purchasing financial assets.[3] Thus, the implicit cost of each dollar of internally funded investment is (at least) the interest on forgone lending.

In a "perfect capital market," where all borrowers and lenders pay and receive a uniform interest rate, the explicit interest cost of loan-financed investment equals the implicit forgone-interest cost of self-financed investment, so the cost is the same whether the firm finances through borrowing or internally.

Issuing new shares of stock creates costs for those who own existing shares. Since the new shares represent claims on the firm's future profits, they dilute the claims of existing shareholders in direct proportion to the amount of new stock issued. If there are 1000 shares in the firm initially outstanding and the firm issues 1000 additional ones, then each existing share's claim on profits is reduced by half, from 1/1000 to 1/2000. If the expected profits of the firm fail to rise (as they should if the investment is a wise one), the price of each share will fall to half its original value. Thus, owners of the firm may incur a cost in terms of reduced market value of their shares and dilution of their share of future profits unless the proceeds of the stock issue are used in ways that raise profits commensurately.

There is a fundamental difference between financing investment through borrowing and financing either with accumulated cash or by issuing new stock. Borrowing creates a *legal obligation* to repay (with interest) that is not present when investment is financed internally or with equity. Suppose that a firm's investment does not pay off. If it has financed the investment with internal funds, then those funds are simply lost. If new stock was issued, then all stockholders will see a reduction in share values and perhaps a low or zero dividend yield. However, there is no legal obligation in either case. In contrast, a firm whose investment fizzles is still legally obligated to pay interest and principal on any debt incurred. Failure to make the required payments results in bankruptcy of the firm. Thus, debt financing entails a bankruptcy risk that is not present with either internal or equity finance.

---

[2] There are also many hybrid financial instruments such as preferred stock, convertible bonds, and the like. We shall not concern ourselves with these.

[3] Alternatively, accumulated profits that are used for investment could have been distributed directly to shareholders as dividends. The value of such dividends may also be part of the opportunity cost of investment.

From the standpoint of the lender, credit transactions are very convenient. Whereas an owner of equity must be constantly vigilant about the company's performance and profits, a bond-holder need only worry if the firm's performance deteriorates to the point that bankruptcy threatens.

The decision about how much of the firm's capital stock should be financed by borrowing vs. equity or cash is sometimes called the ***leverage*** or ***gearing*** decision. A firm is said to be "highly levered" or "highly geared" if it has a lot of debt relative to the amount of its equity. Recall that levers and gears are simple machines that allow a small amount of effort to move large objects. As the following example shows, high leverage allows the owners of a firm to make a small investment (in the financial sense) of their own money but to control a large volume of assets.

Consider the gains or losses of the owners of two firms, LL (for low leverage) and HL (high leverage). Firm LL has no debt outstanding and finances all of its investment through accumulated earnings or by issuing new equity. Suppose that LL has $1000 of capital backed by 1000 outstanding shares and that it earns either $150 or $50 in profit, depending on whether it has a good or bad year. Each share represents a claim on 1/1000 of that profit, so the owners of the LL stand to earn a profit of $0.15 or $0.05 on each share invested. If good and bad years happen with equal frequency, the expected return is $0.10 per share (a 10% return) and the range of outcomes around the mean is ±$0.05.

Now consider HL, which has $1000 in capital backed by 500 shares and $500 in borrowing at a 10% interest rate. The owners of the firm have now only put up $500 of their own money, but they still control the entire $1000 worth of assets and have claim on the entire profit of the company (after interest is paid). The "lever" of debt-finance gives them the same economic clout as the LL owners with half as much personal investment. The firm now pays $50 in interest each year, so profit is either $100 or $0 depending on whether it has a good or bad year. Since each share represents a claim on 1/500 of total profit, the owners of HL will earn $0.20 per share in a good year and zero in a bad one. Although the mean return is still $0.10 (10%), the range has now doubled to ±$0.10.

Thus, high leverage allows owners of the firms to increase the variation of their returns (and their potential profits or losses) for each dollar invested. Taking leverage to the extreme, consider firm RHL (for ridiculously high leverage) that has only 1 share outstanding and $999 of debt. The owner of RHL has a total investment of $1 in the firm, but controls $1000 of assets. RHL must pay $99.90 in interest each period, so it makes a profit is $50.10 in a good year and loses $49.90 in a bad year. The mean return is still $0.10, but the range is an astounding ±$50.00!

The Modigliani-Miller theorem

You would expect that the leverage decision of a firm would affect the firm's attractiveness to potential buyers of its stock and to potential lenders and, in practice, investors often do pay attention to leverage ratios. However, in a perfect capital market where everyone has full information about the probabilities of good and bad years and where everyone borrows and lends at the same interest rate, the value of the firm and the attractiveness of its equity turn out to be totally independent of how it is financed.

Franco Modigliani and Merton Miller demonstrated this most remarkable result in a famous 1958 paper.[4] The Modigliani-Miller theorem demonstrates that under conditions of perfect capital markets, the cost of investment to firms is the same regardless of which of the three methods of finance it chooses.

To understand the logic of the Modigliani-Miller theorem, we return to our example firms from above. Suppose that you have $100 to invest and are considering whether to buy shares in LL or in HL. At first glance, an investment in HL appears much riskier. However, this risk is fully "diversifiable;" you can "unwind" the high leverage of HL by combining your investment in HL with a simultaneous purchase of bonds, whose return does not depend on HL's profit. Suppose that you invest the entire $100 in 100 shares of LL. This earns you $15 in dividends in a good year and $5 in dividends in a bad one.[5] Alternatively, you could invest $50 in 50 shares of HL and $50 in bonds paying 10% interest (perhaps even the same bonds issued by HL!). The bonds pay you $5 every year, while the shares in HL pay $10 in a good year and nothing in a bad one. Thus, you get $15 in a good year and $5 in a bad year, which is the same return in both good and bad years that you got from investing $100 in LL shares.

This example shows that individual investors can choose the riskiness of their portfolios independently of the financial decisions of the firms. A low-risk portfolio can either consist of lots of low-risk shares or of a few high-risk shares balanced by bonds. Further, high leverage increases both the supply of bonds (because firms are borrowing more) and the demand for bonds (when investors want more bonds to balance off the high-risk equities), hence there will be no effect on market interest rates.

The Modigliani-Miller theorem shows that under ideal conditions the decision about how much to invest is independent of the decision about how to finance that investment, since the value of the firm is the same regardless of whether the firm is-

[4] See Modigliani and Miller (1958). These two economists were awarded the Nobel Prize in economics for this (and other) work in 1985 and 1990, respectively.

[5] We are assuming that the firms pay out all profits in dividends. Modigliani and Miller (1961) showed that the firm's dividend policy does not affect our fundamental result.

sues bonds (becoming highly levered) or uses accumulated profit or the proceeds from issuing new equity. This independence allows macroeconomists to focus only on the firm's investment decision, leaving analysis of the decision about how to raise the required funds to specialists in finance.

Of course the assumptions underlying the Modigliani-Miller theorem, like those of most macroeconomic theories, are unlikely to be completely fulfilled. The world is full of information asymmetries and other capital-market imperfections that lead to some important exceptions to the Modigliani-Miller result.

Sources of finance and credit availability effects

One important exception to Modigliani-Miller results from differences between the financing options of small and large firms. While information about the prospective profit outcomes (and, therefore, about the creditworthiness) of large firms is likely to be widely known, outsiders may know little about smaller firms in the economy. Everyone knows Microsoft's reputation and many analysts and investment advisors track Microsoft's performance on a week-to-week basis. However, it would be difficult for a potential investor to get similar information about a tiny startup company. Since investors will not usually invest in firms they know nothing about, small companies must often rely on borrowing from banks or attracting investment from specialized "venture capitalists" rather than raising investment funds by selling bonds or shares directly to the public.

Banks are less reluctant than the general public to lend to small firms for several reasons. First, banks may have access to detailed information about these firms' transactions through records of their checking accounts and of other financial transactions in which the bank has participated. These records allow banks to verify information that the firms provide about their financial performance. Second, many banks have departments that specialize in small-business lending, so they have professional loan officers who have extensive experience in evaluating small firms' financial prospects. Finally, banks are large institutions that have large amounts of (their depositors') money to lend. By lending to a large number of relatively small firms, they are able to diversify their risk more effectively than most individual investors can.

If small firms are dependent on the banking system for credit, then they may be especially sensitive to conditions in the banking sector. It is widely believed that a tightening of monetary policy by the central bank (the Federal Reserve System in the United States) causes commercial banks to reduce the volume of their lending. While interest rates for large firms typically go up somewhat as a result of monetary tightening, these firms usually still have access to funds through financial markets. Small firms however may find their financial tap dried up completely due to a so-called ***credit crunch*** in the banking system.

As we shall study later, the banking system is quite sensitive to the availability of reserves that are supplied by the central bank. When the central bank contracts the supply of reserves relative to the demand for them, banks must reduce lending below the level they would otherwise choose. When this happens, banks typically respond with some combination of (1) raising interest rates on loans and (2) reducing credit lines and refusing to lend to marginal customers.[6]

Thus, small firms may be more sensitive to changes in credit conditions than large firms. This asymmetry lies at the heart of a debate about whether there is a "credit channel" of influence of monetary policy on the economy. Proponents of the credit channel have found some evidence that tight credit affects investment by small firms more than that of large firms.[7]

C. Investment and the Business Cycle

Economists have long recognized that investment tends to be the most volatile of the components of expenditure over the business cycle. Of course, strong correlations between investment and output only mean that the two variables tend to move together; they do not allow us to determine the direction(s) of causality. For that, we need the framework of economic theory with which to interpret the data. Some economists have inferred from the high volatility of investment that fluctuations in investment demand are a major cause of business cycles. Others have argued that the wide variation in investment over the cycle reflects consumption smoothing: investment gets squeezed out as households attempt to maintain their consumption expenditures at a high level during recessions.

Keynes and the business cycle

A satisfactory theory of the business cycle was a pressing need in the 1930s, when the Great Depression inflicted widespread economic suffering on Europe and America. John Maynard Keynes attempted to fill that need with *The General Theory*

---

[6] In a perfectly competitive market, we would expect this rationing of credit to be done entirely by raising interest rates (the price of credit) to a level that would equate supply and demand. However, credit markets are plagued by an information problem known as ***adverse selection***, which implies that increasing the interest rate on loans will mean that relatively safe borrowers will stop borrowing but very risky borrowers (who are likely to default anyway) will be attracted. Thus, if banks raised interest rates enough to reduce demand to equality with supply, they would probably end up with a substantially riskier pool of borrowers and a higher default rate.

[7] A survey article on the credit-channel literature is Bernanke and Gertler (1995).

of Employment, Interest and Money, which he wrote in 1935. Although the ambiguities in *The General Theory* have proved sufficient to sustain a huge literature attempting to interpret Keynes, one of the points on which most scholars agree is that Keynes believed that fluctuations in investment were the primary source of cyclical fluctuations.

Keynes began by rejecting the classical assumption that the economy automatically reverts to full employment quickly and reliably. Under conditions where markets do not clear, he argued, a shortage of aggregate demand may prevent the economy from producing at full capacity. Since investment is the component of aggregate demand that falls most strongly in business-cycle downturns, it was a natural candidate for Keynes in his search for the causes of these declines in demand.

The underlying principles of Keynes's theory of investment do not differ much from the theories that we study today. He used somewhat different terminology and ignored some of the subtleties that subsequent theoretical work has filled in, but his basic framework was similar both to that of classical economists and to the framework we use today. This theory asserts that investment is the result of firms balancing the expected return on new capital—we call it the marginal product of capital; he called the marginal efficiency of capital—with the cost of capital, which depends primarily on the real interest rate.

However, Keynes and classical economists emphasized different kinds of fluctuations within this similar framework. Classical (and often modern) economists usually emphasized the effect that changes in real interest rates have on investment. This effect occurs as firms move up and down on their downward-sloping investment-demand curves. Keynes believed that the large fluctuations in investment were due to shifts in the investment-demand curve itself rather than to movements along the curve.

According to Keynes, the investment-demand curve is volatile because it depends on firms' expectations of the profitability of investment. Keynes thought that the "animal spirits" of investors tended to fluctuate wildly in waves of optimism and pessimism. He viewed the business cycle as a sequence of contagious spells of over-optimism and over-pessimism. During an economic boom, businesspeople project the rapid expansion of the economy (and of the demand for their products) to continue. They respond to these favorable projections of future demand by increasing their production capacity through high levels of investment in new capital. This high spending then fuels the expansion, raising demand for the products of other firms and encouraging their optimism. (Recall that output is determined by aggregate demand in Keynes's system.)

Because these optimistic expectations eventually run ahead of the economy's ability to sustain the expansion, disappointment is inevitable. When the economy begins to turn downward, many firms find that they have substantial excess capacity, both because demand is now falling and because their high rates of investment have left them with the capacity to produce an unrealistically high volume of output. Firms faced with this excess capacity stop investing, which lowers aggregate demand and accentuates the downward pressure on the economy. As demand and output decline, firms become even more pessimistic, keeping investment near zero during the contraction phase of the cycle.

The cycle eventually starts back upward when firms in some industries find their capital stocks depreciated to the extent that they need to buy some new capital goods to sustain their (low) current levels of production. This initial trickle of investment starts aggregate demand on the road to recovery. Optimism gradually begins to replace pessimism, and the expansion phase of the next cycle begins.[8]

The accelerator theory of investment

Among the earliest empirical investment models was the acceleration principle, or accelerator.[9] In modern textbooks, the accelerator model survives as a theory of inventory investment, as discussed on page 481 of Mankiw's text. The accelerator is a simple model that incorporates the kind of feedback from current output to investment that Keynes saw occurring through the effect of current output on investors' expectations.

The accelerator model begins with an assumption that firms' desired capital-output ratio is roughly constant.[10] This implies that the desired capital stock for any period t is proportional to the level of output in t, $K_t^* = \sigma Y_t$, where σ (the lower-case Greek letter sigma) is the desired capital-output ratio. Suppose that firms invest in period t in order to bring their capital stocks to the desired level K_{t+1}^* in period $t + 1$. Then, if depreciation is zero for simplicity, $I_t = K_{t+1}^* - K_t$. But since $K_t^* = K_t$, that means that $I_t = \sigma\,(Y_{t+1} - Y_t)$. Thus, the simplest accelerator model predicts that investment is proportional to the increase in output in the coming period.

Firms, of course, do not observe future output with certainty, so the Y_{t+1} term must be interpreted as an expectation. The dependence of investment on expectations is both realistic and central to Keynes's ideas. However, since we cannot observe expectations of firms about future output, this feature of the accelerator model posed problems for those who wished to implement it. The most common way of resolving this difficulty was to assume that firms expect the change in output in the

[8] A relatively lucid interpretation of Keynes's business cycle theory is presented in Hicks (1950).

[9] The origins of the principle go back at least to Carver (1903). The name "acceleration principle" seems to have been coined by Clark (1917).

[10] Since we are focused on short-run business-cycle fluctuations here, it is reasonable to ignore changes in K/Y that may be associated with long-run advances in technology.

coming period to be equal to the change in the current period. In mathematical terms, they assume that $E_t\,(Y_{t+1} - Y_t) = Y_t - Y_{t-1}$. While modern theorists who are accustomed to using rational expectations will find fault with this myopic theory of expectations, it reflects quite reasonably what Keynes thought was happening in the 1930s—that firms observed a rise or decline in output and extrapolated that change into the future in determining their investment spending.

Since the capital-output ratio in most economies is larger than one (often three or more in advanced economies), moderate expected changes in output are capable of triggering relatively large changes in investment in the accelerator model. This is one of the reasons that this theory gained great popularity after the Great Depression as a model of investment.

The multiplier-accelerator model

The accelerator model can be combined with Keynes's theory of the consumption multiplier to produce a simple model of cyclical behavior. The multiplier, which Keynes actually borrowed from R.F. Kahn, was among the clearest concepts in *The General Theory*. Assuming that output is determined by aggregate demand, which includes consumption and investment, the multiplier shows that changes in consumption will amplify the effect of any change in investment on total output and income.

Suppose that investment increases for some exogenous reason. This will raise aggregate spending and cause output to rise. Since output equals income in the economy, aggregate income rises as the producers of the new investment goods enjoy higher sales and incomes. According to Keynes's fundamental psychological law, these people will spend part but not all of the increase in their incomes. This leads to a secondary (but smaller) increase in aggregate demand, which raises the incomes of those who produce the products that the first wave of new consumers buys. As their incomes go up, they will in turn increase consumption spending, but by less than their incomes rose. Thus, an increase in investment sets off a never-ending sequence of ever-smaller increases in consumption demand that augment or "multiply" the effect of investment on income. By simple algebra, the sum of these effects, or the ***Keynesian multiplier***, can be shown to converge in the limit to 1/(1 – MPC). Keynesian economists generally reckoned the MPC to be well in excess of one-half, so the multiplier was thought to more than double the effect of investment on output.[11]

When the multiplier is combined with the accelerator, the resulting model is capable of interesting dynamics. This is particularly true if either the consumption func-

[11] As a reality check, it is important to remember that the multiplier only works with this degree of simplicity in a world in which rises in aggregate demand cause producers to raise output one-for-one in response, i.e., when the aggregate-supply curve is horizontal. To the extent that producers raise prices rather than producing more (or if the increase in demand causes a rise in interest rates), much or all of the multiplier effect may be neutralized.

tion or the investment function has a lagged response to changes in income.[12] Suppose that both consumption and investment have a one-period lag in their response to income changes, so that $C_t = (1 - s)Y_{t-1}$ and $I_t = \sigma(Y_{t-1} - Y_{t-2})$. We close the model by assuming that $Y_t = C_t + I_t$, which reflects both the fact that we are ignoring the other components of aggregate demand (government spending and net exports) and the Keynesian assumption that output is determined by demand. In this model, it is often assumed that Y, C, and I represent the deviation of the respective variables from their trend growth paths rather than the total levels of the variables themselves. We will adopt that interpretation here.

To examine the dynamic behavior of real output, we can substitute for C and I to get $Y_t = (1 - s + \sigma)Y_{t-1} - \sigma Y_{t-2}$. This expresses the cyclical component of real output as a second-order difference equation. The dynamic behavior of Y depends on the parameters s and σ. There are four possible cases. Case I: If $\sigma < (1-\sqrt{s})^2$, then the value of Y declines steadily over time. If Y is interpreted as the deviation of output from its trend, then Case I implies that output will return smoothly to its trend growth path. Case II: If $(1-\sqrt{s})^2 < \sigma < 1$, then the system will return to its trend path, but not monotonically. Instead, it will overshoot the trend path, turn itself around, and then overshoot again. The pattern of convergence will be as a damped sine wave, oscillating first above then below the growth path. As long as $\sigma < 1$, the oscillations will become gradually smaller and eventually damp out. Case III: If $1 < \sigma < (1+\sqrt{s})^2$, then the system will again oscillate, but the oscillations will grow ever larger with time instead of gradually damping out. Case IV: If $\sigma > (1+\sqrt{s})^2$, then output explodes monotonically away from its trend path.

Case II appears to hold some promise as an explanation for why an economy might have business cycles. Cases III and IV predict that the economy will explode away from its trend growth path. The explosive cases seem unrealistic, though Hicks built a theory of business cycles around an explosive multiplier-accelerator model coupled with floor and ceiling constraints that bounded output in a range around the growth path.[13]

---

[12] The original mathematical analysis of the multiplier-accelerator model was Samuelson (1939). Samuelson was a true pioneer of the use of mathematics in economic analysis. His *Foundations of Economic Analysis*, originally published in 1947, was the first detailed statement of economic theory in mathematical terms. Samuelson's famous *Economics* text dominated the market for decades. He won the Nobel Prize for economic science in 1970, the second year in which the prize was awarded. The presentation of the multiplier-accelerator model used here is based on Chapter 17 of Allen (1968).

[13] See Hicks (1950). Hicks argued that capacity constraints would prevent the economy from moving too far above its growth path and that the nonnegativity constraint on gross invest-

The multiplier-accelerator model is no longer used much as a theory of business cycles, though the accelerator occasionally crops up in analyses of particular regions or industries. There are several reasons for this. The first is that macroeconomists have become very skeptical of the aggregate-demand-based theory of output determination embodied in the model. As you know, modern approaches to business cycles emphasize the joint determination of output by demand and supply factors, with interest rates and prices playing an important role.

Moreover, the great variation in lengths and severity of business cycles over time and across countries argues against an "endogenous" explanation of the cycle such as that provided by the multiplier-accelerator model. In such a model, each business cycle should be the same length and, depending on the formulation, perhaps of the same magnitude as well. Modern macroeconomic theory usually assumes that the business cycle results from repeated random disturbances to the economy by positive or negative shocks, together with a stable convergence mechanism such as that of Case I (or possibly Case II) above. Depending on the timing and magnitude of shocks, it is possible to have both short and long business cycles and both severe and mild ones in such a model.

D. Understanding Romer's Chapter 9

Because investment is essentially a dynamic problem, Romer's Chapter 9 uses some sophisticated methods of dynamic modeling. Although Romer presents these methods in a fairly understandable form, some additional intuitive interpretation may facilitate your understanding. This section attempts to provide that interpretation.

The firm's profit function

Romer throws us a curve on the first page of Chapter 9 by using a highly abstract functional representation of the firm's profit function. First, think carefully about what Romer means by the function $\pi(K, X_1, X_2, \ldots, X_n)$. Profits are equal to

$$\pi(K, X_1, X_2, \ldots, X_n) - r_K K,$$

where r_K is the nominal rental price the firm pays to use one unit of capital. Thus, the π function must measure profit before subtracting out the cost of capital. To have a convenient title, we shall call this "operating profit."

---

ment would prevent it from moving too far below it. When the economy hit the floor and ceiling that bound the growth path, it would be sent back toward the path.

What would the operating profit function of a typical firm look like? Think about the simplest possible case: a competitive firm that produces using only capital and labor. Such a firm has revenue equal to PQ, where P is the competitive market price of its product and Q is the amount it produces. The firm's production is constrained by its production function, so $Q = F(K, L)$. In addition to capital costs, the firm incurs labor costs equal to wL, where w is the market wage. Thus, the firm's profit excluding capital cost (operating profit, as we are calling it) is $P \cdot F(K, L) - wL = \pi(K, P, w)$. For the competitive firm, the Xs in Romer's profit function are P and w, the prices of output and of the other input, labor.[14]

For a firm with monopoly power, P is not exogenous to the firm but instead is affected by the amount of output the firm wants to sell. For such a firm, P would be replaced in the list of Xs in the π function by the determinants of the position of demand curve for the firm's product, such as consumer incomes and the prices of substitutes and complements.

Maximization of profit involves setting the derivative of $\pi(K, X_1, X_2, \ldots, X_n) - r_K K$ with respect to the amount of capital input equal to zero. Romer's equation (9.1) follows from setting $\partial\pi/\partial K = 0$. For the competitive firm we discussed above,

$$\pi_K(K, X_1, X_2, \ldots, X_n) = \partial[PF(K, L) - wL]/\partial K = PF_K(K, L).$$

Thus, equation (9.1) implies $PF_K(K, L) = r_K$. This equation is nothing more than the standard profit-maximization condition: marginal revenue product equals marginal factor cost. It can be rewritten as $F_K(K, L) = r_K/P$, which says that a competitive firm maximizes profit where the marginal product of capital equals the real rental rate on capital. We can draw the left-hand side as a decreasing function of K since the marginal product of capital is assumed to decrease as more capital is employed (given the level of labor). This means that the MPK curve can be interpreted as the real demand curve for capital, with the firm renting capital up to the level where the marginal product equals the real rental price.

User cost of capital

Firms usually own most of the capital that they use rather than renting it. We can still think of the cost of capital services as a rental price, but we must figure out what a profit-maximizing owner of capital would charge (herself) for using one year's

[14] Notice that L is not included. The amount of labor is a choice variable for the firm at every moment in time; only variables in the environment that are exogenous to the firm are put into the π function. For a given production function we would eliminate L from the profit function by calculating a profit-maximizing labor demand function $L^*(w, P)$ and substituting in for L in the expression above so that $\pi(K, P, w) = P \cdot F[K, L^*(w, P)] - w L^*(w, P)$.

worth of capital services. This price is what we call the ***user cost of capital***, which is (in equilibrium) equal to the ***rental price of capital***.

Dale Jorgenson developed the theory of the user cost of capital in the 1960s.[15] Consider the long-run problem of a firm that produces output in continuous time according to the production function $Q(t) = F(K(t), L(t))$. It sells its output at price $P(t)$ and pays $w(t)$ for labor at time t. The firm buys new capital goods (invests) at rate $\dot{K}(t)$ and pays $p_K(t)$ for each unit of capital. Capital is assumed to depreciate at a constant rate δ.

The present value of this firm's (infinitely long) lifetime flow of profits is

$$V(0) = \int_0^\infty e^{-rt} \left(P(t)F\left[K(t)), L(t)\right] - w(t)L(t) - p_K(t)\dot{K}(t) \right) dt. \qquad (1)$$

The expression in large parentheses is the firm's ***net cash flow*** at time t, which is the difference between its revenues and its current expenditures for labor and new capital goods.[16] The firm's objective is to choose paths for $K(t)$ and $L(t)$ in order to maximize $V(0)$, given the paths of w and P. Since equation (1) contains both the level and the change in K, the maximization problem must be done using the calculus of variations, which is a technique you do not need to learn yet.

Jorgenson was able to demonstrate that a firm that maximizes the present value of its net cash flow $V(0)$ chooses exactly the same path for K and L as a firm that maximizes moment-by-moment the level of profit at each moment in time, provided that the price attached to capital services is the "user cost of capital." In other words, a firm that maximizes the present value of its long-run net cash flow would set MPK $= r_K(t)/P(t)$ at each moment in time, where $r_K(t)$ is given by Romer's equation (9.4). Note that this is the same condition that we derived earlier for a firm that rents its capital. It is reassuring that a firm that owns its capital would choose the same amount of capital as one that rents from someone else in a competitive market.

Intuitively, the expression for the user cost of capital tells us that the implicit rent that the firm charges itself for using its own capital must include compensation for three cost components. The firm must compensate itself for (1) the interest it could earn if it sold the capital and bought bonds with the proceeds, (2) the fraction δ of the capital that wears out during the period, and (3) any decrease in the market price of

[15] Jorgenson is a distinguished ex-Reedie who now teaches at Harvard. The development of the user cost of capital was among his first major contributions to economic research.

[16] Note that we use the term net cash flow rather than profits. The bracketed expression deducts "outlays" at time t on new capital goods rather than the "cost" of the services of the capital goods that are used in producing time t output. We reserve the term "profit" to refer to revenue less the latter concept of cost and use net cash flow to denote the difference between time t revenues and actual cash outlays.

capital that occurs during the period. A decline in the price of capital goods means that the potential resale value of the capital is falling, which makes using capital more costly. If capital-good prices are rising, this cost component is negative and reduces the user cost of capital.

Adjustment-cost model

Jorgenson's dynamic model of investment describes the optimal investment behavior of a firm that operates in an environment where it can adjust its capital stock up or down very quickly in order to stay on its optimal path. This is not a problem as long as the desired capital path is smooth, so that the level of investment $\dot{K}(t)$ does not become too large in either a positive or negative direction. However, if there were a sudden leap in $r_K(t)$, perhaps due to a sudden change in the interest rate, then the firm would want to change its capital stock by some finite amount instantly in order to raise or lower the MPK. A discrete increase in $K(t)$ implies an infinite rate of investment at the moment of the change, $\dot{K}(t) = \infty$.[17]

An infinite rate of investment is obviously implausible, so how can we revise our model in order to preclude such behavior? We do this by introducing costs of rapid adjustment of the capital stock. The more quickly the firm adjusts its capital stock (i.e., the higher the rate of investment) the higher the cost it is assumed to incur. We represent adjustment costs mathematically as $C\left[I(t)\right]$, with $C[0]=0$, $C'[0]=0$, and $C''[I]>0$. The easiest way to think of the adjustment-cost function (and the form of the function that we most often use) is as a parabola opening upward from its vertex at the origin.[18] The firm incurs positive adjustment costs when it changes its capital stock either upward or downward, and those costs rise at an increasing rate as the amount of net investment gets further from zero.[19]

[17] Think about the time path of $K(t)$. The rate of investment is equal to the slope of the time path with respect to time. If $K(t)$ jumps upward at moment t_0, then the slope of the time path is vertical at t_0. Since the slope of a vertical line is infinite, this implies that $\dot{K}(t_0) = \infty$.

[18] This form of the adjustment-cost function has strong implications for the behavior of investment that are not fully consistent with observed behavior. The strict convexity of the adjustment-cost function implies that firms always incur lower adjustment costs if they spread an investment project over more years. If this were the case, then we would expect to see continuous, gradual investment by firms rather than large, discrete factory additions. Since most investment in structures is "lumpy" it may be more appropriate to recognize the presence of fixed adjustment costs in addition to the variable costs modeled here. This is discussed in Romer's Section 9.8 and in the empirical discussion below.

[19] We have assumed that the depreciation rate is zero. Relaxing that assumption makes the adjustment-cost function a little awkward. With positive depreciation, the cost function we

The form of the profit function that Romer describes on page 409 may require some clarification. He posits a model with N identical firms. Each firm has a real operating profit function given by $\pi[K(t)]\ \kappa(t)$, where $\kappa(t)$ is the firm's capital stock and $K(t) = N\ \kappa(t)$ is the aggregate industry-wide capital stock. The π function represents the profit the firm earns per unit of its capital. This is *not* the same π function that Romer used earlier in the chapter. The assumption of constant returns to scale implies that the firm's real profit can be represented in this multiplicative manner. The π function is downward-sloping because the industry faces a downward-sloping demand curve, so that as the industry's capital stock and output increase, profit per unit of output decreases. The function $\pi[K(t)]\kappa(t)$ gives the firm's real operating costs as a function of its own capital stock and the aggregate industry stock.

Romer's equation (9.6) is directly analogous to our equation (1) with four differences. First, Romer has $\pi[K(t)]\kappa(t)$ in place of revenue less labor costs. Second, adjustment costs are included. Third, the expression is in real rather than nominal terms, meaning that we do not need the price of capital in front of the investment variable (and that the $\pi[K(t)]\kappa(t)$ function must implicitly express revenue less labor costs in real rather than nominal terms). Finally, Romer has not substituted for the $I(t)$ variable as $\dot{\kappa}(t)$ in the integral expression, so the maximization has to be done invoking the constraint rather than unconstrained.

You need not worry about the mathematics of maximizing (9.6). Romer does it first in discrete time, then in continuous time. In discrete time, the maximization is done by the method of Lagrange multipliers; in continuous time it is done as a continuous-time Hamiltonian. This is essentially the same method that we used to maximize utility in the Ramsey growth model. The key variable that comes out of this maximization problem is the Lagrange multiplier (in discrete time) or ***costate variable*** (in continuous time) $q(t)$. You may recall from our introduction of the method of Lagrange multipliers (in Chapter 3) that the value of the multiplier can be interpreted as the marginal gain from releasing the constraint by one unit. In the present case, the constraint is that capital growth is accomplished only through (costly) investment. Thus, the marginal gain from releasing the constraint is the increase in profit that would occur if the firm could obtain one additional unit of capital without incurring investment or adjustment costs.

Tobin's q

James Tobin, another Nobel-prize winner, formulated an investment theory based on financial markets. Tobin argued that firms' investment level should depend on the ratio of the present value of installed capital to the replacement cost of capital.

---

have assumed implies that zero adjustment costs occur where the firm exactly replaces its depreciating capital.

This ratio is ***Tobin's q***. The q theory of investment argues that firms will want to increase their capital when $q > 1$ and decrease their capital stock when $q < 1$. If $q > 1$, a firm can buy one dollar's worth of capital (at replacement cost) and earn profits that have present value in excess of one dollar. Under those conditions, firms increase profits by investing in more capital, so we expect investment to be high. If $q < 1$, then the present value of the profits earned by installing new capital are less than the cost of the capital, so more investment lowers profit. We expect investment to be near zero if $q < 1$. When $q < 1$, someone seeking to enter a particular industry can acquire the necessary capital assets more cheaply by buying an existing firm than by building a new one with new capital. This is true because the value of installed capital (i.e., the cost of buying an existing firm) is less than the replacement cost (the cost of building a new firm).

Romer's analysis shows that Tobin's q is exactly the costate variable (or Lagrange multiplier) q. The key to understanding the connection between the costate variable and Tobin's market interpretation of q is Romer's equation (9.24). This equation shows that $q(t)$ is equal to the present value (as of time t) of the stream of real profits per unit of capital that will be earned from time t into the infinite future. Since a prospective buyer of a share in a firm has a claim on this stream of profits, she will be willing to pay exactly this present value of the stream for each unit of capital she implicitly buys when she buys shares in the firm. Because we are normalizing the real cost of new capital at one, q will thus equal the ratio of the market value of a firm's stock ($q\kappa$) to the replacement cost of its capital (κ). If $q > 1$, then firms can sell a share of new stock for more than a dollar, buy a dollar's worth of capital, and pocket the difference as profit. Hence investment will be high when $q > 1$.

When we solve the model for the optimal rate of investment, it turns out to be an increasing function of q, $\dot{K}(t) = f(q(t))$, with $f' > 1$ and $f(1) = 0$. If the adjustment cost function is quadratic, as we suggested earlier, then the f function is linear and investment is a linear function of q.

Average and marginal q

The costate variable $q(t)$ measures the value to the firm of a *marginal* unit of capital relative to its replacement cost. In the real world, it is very difficult to get data on marginal q, but somewhat easier to estimate average q. The reason that average q is easier to measure is that it can be approximated by comparing the market value of the firm's outstanding stock and debt with the estimated replacement cost of its capital stock. The former is easy, the latter somewhat more difficult, but if you measure q in this way you get a q based on average revenue and cost rather than the more useful marginal q. Because of these measurement problems, most empirical work using q has been based on average rather than marginal q.

Dynamic analysis

On pages 415 through 419, Romer develops a phase-diagram analysis of the joint evolution of the capital stock and q over time. The method should be familiar from our analysis of the Ramsey growth model. There are two variables, K and q, both of which have changes (time derivatives) that depend on the level of one or both variables. One variable (K) is a true state variable in that it cannot jump instantaneously. The other (q) is a "control" variable that can change instantaneously. The equilibrium is a saddle point. The control variable q jumps at any instant to the value given by the saddle path at the current value of K. The economy then converges down or up along the saddle path to a steady state in which neither q nor K is changing.

Suppose that everything else in the economy remains constant, but that the capital stock is for some reason below its optimal level. This means that in the initial equilibrium $q > 1$, and the capital stock will expand along the saddle path. As the capital stock expands, q retreats toward one, eventually converging to a steady-state equilibrium with the capital stock at its optimal level and $q = 1$.

After developing the basic convergence properties of the model, Romer then analyzes three examples. First he shows that an increase in output demand will cause the $\dot{q} = 0$ curve to shift to the right. The value of q initially jumps upward to the saddle path, stimulating positive investment and an increase in the capital stock as the economy converges down the saddle path. In the second example, an increase in the interest rate lowers q and reduces the equilibrium capital stock in the long run. The final example shows that an investment tax credit increases the profitability of capital and raises the long-run capital stock.

Note that all of these effects will be different depending on whether the exogenous change is permanent or temporary. A permanent change results in an immediate leap to the new saddle path followed by convergence along the path. A temporary change will not put the economy on the saddle path because the exogenous variables are known to be changing in the future. (The saddle path describes the convergence of the system when the future values of exogenous variables are stable.) Instead, the economy jumps to an unstable saddle path that leads back to the original path at the moment that the policy reverts to its original state. Once the temporary change is reversed, the economy must be on the saddle path associated with the original (and ultimate) equilibrium and converges, since no future changes in exogenous variables are expected.

E. Empirical Studies of Investment

The empirical analysis of aggregate investment spending has been one of the great frustrations of postwar macroeconomics. Unlike most of the other behavioral functions of the basic Keynesian system, the investment function has not fit the aggregate data for most countries well.[20]

All theories based on profit maximization predict that the flow of investment expenditures should be sensitive to the cost of capital. Sometimes this sensitivity is modeled through q, sometimes through the user cost of capital, but in either case, interest rates and tax rates should affect investment. However, the correlation between aggregate investment and real interest rates is extremely low in most samples and is sometimes positive rather than negative. Despite the creative use of lagged effects and refined measures of q and the user cost of capital, no consistently reasonable aggregate empirical specification has emerged.

This lack of empirical support is all the more frustrating because of the overwhelming consensus that the basic theoretical approach based on q or on the user cost of capital is correct. Of course failing to find a significant relationship in aggregate data does not necessarily imply that the theory is wrong. Measurement error, biases associated with reverse causality, and the effects of omitted variables—those ubiquitous nemeses of econometric analysis—may prevent our econometric tests from finding the true underlying relationship among the variables.

One possible explanation for the apparent failure of aggregate investment demand equations is that we do not have good measures of the variables that shift the demand curve. The fundamental factor through which costs affect investment demand is the expected profitability of increments to the capital stock, which depends mainly on the marginal product of capital and the firm's expectation of the future demand for its product. Only the crudest proxies for these latter variables can be observed, so the effects of demand shifts are usually not well represented in the econometric specification. If, as is plausible, the largest changes in investment result from shifts in these unobserved variables rather than from changes in the cost of capital, then it is not surprising that the observed empirical relationship between investment and the cost of capital is weak.

In contrast, the empirical link between aggregate investment and corporate cash-flow measures seems to be quite strong. Firms invest more when they are earning lots of money, almost regardless of the opportunity cost. Is this because cash flow

[20] This section of the chapter relies heavily on work done jointly with Malcolm Spittler under the sponsorship of a Goldhammer Summer Collaborative Research Grant during the summer of 2002.

proxies for variables such as demand expectations that shift the expected marginal revenue product of capital? Is it because capital-market imperfections overturn the assumptions of the Modigliani-Miller theorem and give firms a strong bias toward internally financed investment? Or is it because the whole theoretical framework is wrong and firms follow some other theory, such as a mechanical rule-of-thumb relationship between profits and investment?

The following sections survey the empirical literature on investment models, including traditional neoclassical approaches, several variants of the q model, and modern approaches that incorporate uncertainty, irreversibility, and nonconvexity.

Empirical evidence on the acceleration principle

The accelerator model was introduced in an earlier section. Recall that this model is derived from an extremely simple assumption about the firm's desired amount of capital: it assumes that the firm desires a fixed capital/output ratio. If we denote this ratio by σ, then the optimal level of capital input is

$$K^* = \sigma Y, \tag{2}$$

where Y is the level of output. Assuming that the firm adjusts its capital stock to the desired level immediately (in the current period, not with a one-period lag is in the model discussed above), then its net investment will be[21]

$$I_t^n = \left(K_t^* - K_{t-1}^*\right) = \sigma\left(Y_t - Y_{t-1}\right) \equiv \sigma \Delta Y_t. \tag{3}$$

The variables of this equation are all easily observed from flow national income account data (no measure of the capital stock is required), so equation (3) can be estimated by simple regression techniques. (If a measure of the capital stock is available, then no regression is necessary, since σ in equation (2) can be estimated as the ratio of the capital stock to output.)

Tinbergen (1938) notes that strict interpretation of (2) with complete adjustment implies that

$$\frac{K_t}{K_{t-1}} = \frac{Y_t}{Y_{t-1}},$$

Thus, the elasticity of the proportional change in the capital stock with respect to the proportional change in output should be exactly one.

At the time of Tinbergen's work, there were no national-accounts data to use in estimating aggregate investment functions. Instead, empirical analysis was based on the few industries for which good-quality data on output and capital (or capacity) could be obtained—notably railroads. Tinbergen estimates the elasticity of railroad

[21] Gross investment equals net investment plus replacement investment (depreciation). The latter is usually modeled as proportional to the previous period's capital stock, so a corresponding gross investment equation would add a term δK_{t-1} at the end. Feldstein and Foot (1971) discuss this issue.

capacity (a weighted average of locomotives and railroad cars) to traffic to be closer to 0.5 than to 1.0 for the United States, United Kingdom, and Germany. Reconciling the accelerator principle to these data requires more flexibility than the crude proportionality model.

As a practical matter, firms are unlikely to adjust their capital stocks fully within one quarter or year. While some kinds of capital equipment can be bought off the shelf and put in place quickly, equipment that requires substantial planning, installation, and delivery time as well as nearly all structures require considerable time to build. Such lags in the installation of new capital imply that actual investment spending may lag behind changes in the desired capital stock. Jorgenson (1965) characterized this lag as comprising five distinct steps: (1) initiation of the investment project, (2) appropriation of funds, (3) letting of contracts, (4) issuing of orders, and (5) actual investment. The length of time these processes require obviously varies greatly among firms and among different kinds of capital expenditures.

To allow for a lagged response, a distributed lag can be introduced into equation (3), giving

$$I_t^n = \sum_i \gamma_i \left(K_{t-i}^* - K_{t-i-1} \right). \tag{4}$$

For the accelerator model, this implies

$$I_t^n = \sigma \sum_i \gamma_i \left(\Delta Y_{t-i} \right). \tag{5}$$

Models of the accelerator in which investment responds slowly over time to changes in output are often called "flexible accelerator" models.[22] One of the most prominent distributed lag relationships in econometrics was introduced by Koyck (1954) to estimate the accelerator model of investment. The Koyck lag model assumes that the firm's investment level in each period is a fraction $(1 - \lambda)$ of the gap between its existing level of capital and its desired level. This leads to a set of lag weights γ that decline exponentially as i increases: γ_i is proportional to λ^i. The Koyck lag can be applied either in terms of the absolute level of K and Y, or (as Koyck did) in proportional terms. Koyck's hypothesis was that the capital stock in year $t + 1$ was determined by[23]

$$K_{t+1} = \theta Y_t^{\alpha} Y_{t-1}^{\beta} Y_{t-2}^{\lambda\beta} Y_{t-3}^{\lambda^2\beta} \cdots\cdots Y_{t-i}^{\lambda^{i-1}\beta} \cdots\cdots e^{\delta t}. \tag{6}$$

Denoting the logs of Y and K by lower-case letters (6) becomes

---

[22] This terminology was introduced by Goodwin (1948).

[23] The β term on the first lag of Y does not strictly obey the pattern of exponentially declining weights. Koyck introduces this term to account for the likelihood that exponential decline in the lag weights may not begin at the first lag. If $\beta = \lambda\alpha$, then the equation reduces to a strict exponential lag.

$$k_{t+1} = \ln\theta + \alpha y_t + \beta y_{t-1} + \lambda\beta y_{t-2} + \lambda^2\beta y_{t-3} + \ldots + \lambda^{i-1}\beta y_{t-i} + \ldots + \delta t. \tag{7}$$

Koyck's ingenious insight was to note that the infinite summation of lagged terms on the right-hand side of (7) could be eliminated by lagging (7), multiplying both sides by λ, then subtracting:

$$\begin{array}{lllllll} k_{t+1} & = \ln\theta & + \alpha y_t & + \beta y_{t-1} & + \lambda\beta y_{t-2} & + \lambda^2\beta y_{t-3} + \ldots + \lambda^{i-1}\beta y_{t-i} + \ldots + \delta t & \\ -\lambda k_t & = -\lambda\ln\theta & & -\lambda\alpha y_{t-1} & -\lambda\beta y_{y-2} & -\lambda^2\beta y_{t-3} - \ldots - \lambda^{i-1}\beta y_{t-i} - \ldots - \lambda\delta(t-1) & . \\ \hline k_{t+1} - \lambda k_t & \multicolumn{5}{l}{= (1-\lambda)\ln\theta + \alpha y_t + (\beta - \lambda\alpha) y_{t-1} \qquad\qquad + (1-\lambda)\delta t + \lambda\delta} & \end{array}$$

Subtracting $(1 - \lambda)k_t$ from both sides yields

$$\Delta k_{t+1} = \phi_0 + \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 t + \phi_4 k_t, \tag{8}$$

where

$$\phi_0 = (1-\lambda)\ln\theta + \lambda\delta, \quad \phi_1 = \alpha, \quad \phi_2 = \beta - \alpha\lambda, \quad \phi_3 = \delta(1-\lambda), \quad \phi_4 = \lambda - 1. \tag{9}$$

The left-hand side of equation (8) is the growth rate of the capital stock (the difference in its logs), which is essentially the flow of investment divided by the stock of capital.[24] The short-run, or impact, sensitivity of capital stock growth to output is measured by $\phi_1 = \alpha$. This is the amount that k_{t+1} would change if y_t increased by one unit, but the lagged y values did not change.

However, the long-run sensitivity is potentially much larger as the lagged effects accumulate through ϕ_2 and ϕ_4. To compute the long-run effects, we must think about how k_{t+1} would change in equation (7) if *all* of the y terms on the right-hand side were to increase by one unit. The sum of the coefficients on the y terms is

$$\alpha + \beta + \lambda\beta + \lambda^2\beta + \ldots = \alpha + \sum_{i=0}^{\infty} \lambda^i\beta = \alpha + \frac{\beta}{1-\lambda}. \tag{10}$$

Koyck estimates this equation with data for several industries. We shall focus on his estimates for the railroad industry from 1894 to 1940. He measures the railroad capital stock by the stock of railroad cars and the industry's output by ton-miles of freight carried.

Koyck notes that if there is an error term in (6) and (7), then the error term in the estimating equation (8) will be serially correlated due to the differencing transformation.[25] To account for this, he estimates both by ordinary least squares (ignoring the serial correlation problem) and by a method that corrects for the serial correlation, but notes that the results of the two methods are nearly identical. His OLS estimates of (8) for the railroad sample are

$$\Delta k_{t+1} = \text{constant} + 0.077 y_t + 0.017 y_{t-1} - 0.0033t - 0.110 k_t,$$

[24] More recent empirical investment equations typically use the ratio of investment to the capital stock as the dependent variable.

[25] If the original error term is white noise, then the error term in the transformed equation will be a first-order moving average process.

with R^2 = 0.848.[26] From these estimated coefficients, he calculates the implied parameter values by inverting the formulas in (9):

$$\hat{\alpha} = \hat{\phi}_1 = 0.077,$$
$$\hat{\lambda} = \hat{\phi}_4 + 1 = -0.110 + 1 = 0.890,$$
$$\hat{\beta} = \hat{\phi}_2 + \hat{\alpha}\hat{\lambda} = 0.017 + 0.077 \cdot 0.890 = 0.086,$$
$$\hat{\delta} = \frac{\hat{\phi}_3}{1-\hat{\lambda}} = \frac{-0.0033}{0.110} = -0.030.$$

All of Koyck's coefficient estimates are of the expected sign. Increases in output lead to strong investment in the current year and in the following year (α and β are both positive). As noted above, the strict accelerator model predicts a unitary elasticity of capital with respect to output. The immediate effect of an increase in output is to increase capital stock growth by just 0.077, much less than the unitary effect. However, using formula (10), we can calculate the implied long-run elasticity to be

$$\alpha + \frac{\beta}{1-\lambda} = 0.077 + \frac{0.086}{0.110} = 0.86,$$

which is much closer to one. Koyck's results suggest that capital responds strongly to increases in output, but that the response is very slow—less than ten percent of the eventual increase in capital occurs in the first year and only about 15 percent within two years. This response (which is typical of estimates of accelerator models) seems longer than one would expect to be caused by delivery and construction lags.

Koyck applies a similar model to five other industries in which physical measures of output and capacity are available for a long period: hydroelectric plants, fuel-generated electric plants, Portland cement, steel, and petroleum refining. The results for these industries are quite similar to those reported above for the railroads: the elasticity of investment with respect to changes in output is small, ranging from less than 0.10 to 0.30 in the first two years, and the eventual elasticity is much larger, but with very long estimated lags.

Although it fits the aggregate and firm-level data reasonably well, the simple and flexible accelerator models are rarely used in modern empirical work. Imposing the assumption of a constant capital/output ratio imposes a zero substitution elasticity, which seems unduly restrictive. Moreover, the sensitivity of investment to interest rates plays an important role in many macroeconomic models of aggregate demand. Thus, the focus of empirical work on investment shifted in the 1960s toward models based on the theory of the profit-maximizing firm.

[26] To simplify the calculations in his pre-computer era, Koyck calculates the regression using deviations of all variables from means. This does not affect the calculation of the slope coefficients of (8), but prevents him from estimating the constant term. Standard errors of the coefficients are not reported.

Empirical testing of the neoclassical model

As discussed above, economic theory tells us that investment should depend on two basic factors: (1) the cost of obtaining and using capital and (2) the stream of revenue that firms expect to earn on a marginal addition to capital. Jorgenson (1963) initiated the "neoclassical" theory of investment by showing that the solution to the dynamic maximization problem in equation (1) could be reduced to a sequence of static conditions, in which firms at every moment set the marginal product of labor equal to the real wage and the marginal product of capital equal to the real user cost of capital.[27]

$$\begin{aligned} \frac{\partial F}{\partial L_t}\left(K_t^*, L_t^*\right) &= \frac{W_t}{P_t}, \\ \frac{\partial F}{\partial K_t}\left(K_t^*, L_t^*\right) &= \frac{C_t}{P_t} \equiv \frac{P_t^K}{P_t}\left(i + \delta - \frac{\dot{P}_t^K}{P_t^K}\right), \end{aligned} \tag{11}$$

where C_t is defined as the nominal user cost of capital at time t, i is the (nominal) interest rate used to discount future cash flows, and the last term in the capital equation is the rate of inflation in capital-goods prices. The expressions on the left-hand side of (11) are functions of the levels of capital and labor. Solving these two equations together yields the optimal demands for the two factors at time t, K_t^* and L_t^* as a function of the real wage and the real user cost of capital.

The first generation of empirical studies based on the neoclassical model assumed Cobb-Douglas production function,

$$F\left(K_t, L_t\right) = A K_t^{\alpha} L_t^{1-\alpha}, \tag{12}$$

which implies that the marginal products are

$$\begin{aligned} \frac{\partial F}{\partial K_t} &= A\alpha K_t^{\alpha-1} L_t^{1-\alpha} = A\alpha\left(\frac{K_t}{L_t}\right)^{\alpha-1}, \\ \frac{\partial F}{\partial L_t} &= A\left(1-\alpha\right) K_t^{\alpha} L_t^{-\alpha} = A\left(1-\alpha\right)\left(\frac{K_t}{L_t}\right)^{\alpha}. \end{aligned} \tag{13}$$

It is well known that in the case with constant returns to scale, only the optimal capital/labor ratio can be determined from plugging (13) into (11). The scale of production is arbitrary, since under perfect competition with constant returns to scale, the long-run distribution of output among firms is indeterminate.

Jorgenson and others solved this problem by taking the firm's path of output (Y_t) as fixed. By substituting in the production function (12), equations (13) can be written

[27] The user cost of capital is often called the "rental price of capital." It represents the annual cost of using a unit of capital goods. If a competitive rental market exists for capital goods, this would be the equilibrium rental rate.

$$
\begin{aligned}
\frac{\partial F}{\partial K_t} &= \alpha \frac{Y_t}{K_t}, \\
\frac{\partial F}{\partial L_t} &= (1-\alpha)\frac{Y_t}{L_t}.
\end{aligned} \tag{14}
$$

Substituting (14) into (11) and solving for the optimal levels of capital and labor yields

$$
\begin{aligned}
L_t^* &= (1-\alpha)\frac{P_t Y_t}{W_t}, \\
K_t^* &= \alpha \frac{P_t Y_t}{C_t}.
\end{aligned} \tag{15}
$$

The latter equation in (15) is the basis for neoclassical investment equations. However, empirical implementation requires two refinements: (1) adjustment of the user cost of capital for tax effects, and (2) a specification of the connection between the flow of investment and the desired capital stock. We now proceed to consider these issues.

Based on equation (11), the real user cost of capital in the absence of tax distortions can be written as

$$
c = \frac{C}{P} = \frac{P^K}{P}\left(i + \delta - \frac{\dot{P}^K}{P^K}\right) = p^K(r+\delta) - \dot{p}^K, \tag{16}
$$

where $r \equiv i - \dot{P}/P$ is the real interest rate. If, as is sometimes true in simple aggregated models, capital prices and general output prices are identical, then equation (16) reduces to $c = r + \delta$.

Since depreciation rates are difficult to observe and unlikely to change very much over time, the real interest rate is the centerpiece of the user cost of capital. However, relying on movements in the real interest rate to explain changes in the opposite direction in investment is not likely to be empirical successful, at least for the postwar United States, because interest rates are not strongly and negatively correlated with investment. In contrast to the predicted negative relationship, the correlation is weak and positive. Moreover, real interest rates do not vary greatly in the postwar sample, so even if a relationship is present it may be difficult to detect.

In contrast, changes in the tax treatment of depreciation expenses, interest expenses, and capital gains, and especially the presence or absence of an investment tax credit can have a much larger impact on user cost than the small range of variation in real interest rates. Hall and Jorgenson (1967) apply the neoclassical capital model with careful consideration of how various U.S. taxes affect the user cost of capital. Assuming that real capital goods prices p^K are not expected to change, so the last term in (16) is zero, they show that the user cost of capital in the presence of taxes can be written

$$c' = p^K (r+\delta)\frac{(1-v)(1-uz)}{1-u}, \tag{17}$$

where v is the rate of investment tax credit, u is the rate at which corporate income is taxed (assumed constant over time), and z is the present value of the stream of depreciation-related tax deductions on the purchase of one dollar's worth of capital.

In order to turn a model of the desired capital stock (such as equation (15)) into a theory of investment, we need to describe the dynamic process by which investment moves the actual capital stock over time in response to changes in the desired stock. Hall and Jorgenson (1967) estimate an investment equation in which net investment is a distributed lag on changes in the desired stock. Like Koyck (1954), they employ a model with lag weights that decline exponentially after the period after the change,

$$I_t^n = \gamma_0 \Delta K_t^* + \gamma_1 \Delta K_{t-1}^* + \lambda I_{t-1}^n, \tag{18}$$

with the optimal capital stock given (following (15)) by

$$K_t^* = \alpha \frac{Y_t}{c_t'}. \tag{19}$$

Substituting (19) into (18) yields the estimating equation

$$I_t^n = \alpha\gamma_0 \Delta\left(\frac{Y_t}{c_t'}\right) + \alpha\gamma_1 \Delta\left(\frac{Y_{t-1}}{c_{t-1}'}\right) + \lambda I_{t-1}^n. \tag{20}$$

There are only three coefficients that can be estimated in a regression of (20): $\phi_0 \equiv \alpha\gamma_0$, $\phi_1 \equiv \alpha\gamma_1$, and λ. Thus, without an additional restriction, it is impossible to identify all four parameters. Fortunately, such a restriction is at hand since the long-run impact of a change in desired capital $\alpha\Delta(Y/c')$ should be equal to unity.

In order to find the equation for this restriction we must calculate the long-run effect of a change in Y/c' on I^n. To isolate the effect of a one-time change, suppose that $\alpha\Delta(Y/c') = 1$ in period 1 and zero in all other periods. Net investment in period zero is also zero. Using (20), $I_1^n = \gamma_0$. Substituting this into equation (20) for $t = 2$ yields $I_2^n = \gamma_1 + \lambda\gamma_0$, since $\Delta(Y/c')$ in period two is zero. From period 3 forward, both $\Delta(Y/c')$ terms are zero, so $I_3^n = \lambda I_2^n = \lambda\gamma_1 + \lambda^2\gamma_0$. Adding up all the effects yields

$$\begin{aligned}\text{long-run effect} &= \gamma_0 + \gamma_1 + \lambda\gamma_0 + \lambda\gamma_1 + \lambda^2\gamma_0 + \lambda^2\gamma_1 + \ldots \\ &= (\gamma_0 + \gamma_1)\sum_{i=0}^{\infty} \lambda^i = \frac{(\gamma_0+\gamma_1)}{1-\lambda} = 1.\end{aligned}$$

Thus, if changes in the desired capital stock eventually lead to equal cumulative changes in investment, $\gamma_0 + \gamma_1 = 1 - \lambda$. But $\phi_0 + \phi_1 = \alpha(\gamma_0 + \gamma_1)$, so by substitution, α can be inferred from the three estimated coefficients as $\alpha = (\phi_0 + \phi_1)/(1 - \lambda)$.

Hall and Jorgenson (1967) estimate (20) for four categories of investment: manufacturing equipment, manufacturing structures, non-farm non-manufacturing equipment, and non-farm non-manufacturing structures, using annual U.S. data for 1931–41 and 1950–63. The results are summarized in Table 1.

Table 1. Hall and Jorgenson's estimates for neoclassical model.

	ϕ_0 (= $\alpha\gamma_0$)	ϕ_1 (= $\alpha\gamma_1$)	λ	α = $(\phi_0+\phi_1)/(1-\lambda)$	Mean lag	R^2
Manufacturing Equipment	0.0142 (0.0037)	0.0124 (0.0044)	0.615 (0.100)	0.0069 (0.0156)	2.1	0.72
Manufacturing Structures	0.0040 (0.0013)	0.0053 (0.0015)	0.766 (0.079)	0.039 (0.013)	3.8	0.85
Non-farm, non-mfg equipment	0.0245 (0.0084)	0.0146 (0.0104)	0.469 (0.134)	0.074 (0.014)	1.3	0.69
Non-farm, non-mfg equipment	0.0130 (0.0020)	0.0023 (0.0022)	0.880 (0.032)	0.127 (0.025)	7.5	0.98

Source: Hall and Jorgenson (1967), Table 2.
Standard errors are reported in parentheses below coefficients. Mean lag is the weighted average number of years between a change in desired capital and the corresponding changes in investment.

Some aspects of the estimates in Table 1 support the neoclassical model. All of the coefficients are positive and the effect of the current year's $\Delta(Y/c')$ is statistically significant for each kind of investment, as predicted by the model. The goodness-of-fit statistics are substantial.

However, there are some anomalies in Hall and Jorgenson's results. First, the inferred values of α, the elasticity of output with respect to capital, are smaller than those suggested by other evidence. (Moreover, the coefficient for manufacturing equipment is not statistically different from zero.) In the Cobb-Douglas production function, the share of national income going to capital should be equal to α. Capital's share has been relatively stable over time at approximately one-third. The combined (sum) estimated share of equipment and structures is 0.05 for manufacturing and 0.20 for the non-farm non-manufacturing sector, which is far below one-third. [28]

A second puzzle in Table 1 is that (as in Koyck's study of the accelerator model) the estimated lags are very long. Decision-making, delivery, and installation lags are unlikely to explain an *average* lag of over 2 years for manufacturing equipment or over 7 for non-manufacturing structures.[29]

[28] The structure of the national-income accounts makes it impossible to identify the capital and labor components in such categories as "proprietors' income." It is conventional to lump this category together with profits, interest, and rent as capital income. Doing this leads to a value for α of 0.36 in 1950, near the middle of Hall and Jorgenson's sample.

[29] For some more recent evidence on lags in the installation of structures, see Montgomery (1995). One rationale for long lags in the adjustment of the capital stock to changes in prices is that it may be more difficult (or impossible) to change the technological characteristics of

Hall and Jorgenson interpreted the statistically significant positive estimate of ϕ_0 and (sometimes) ϕ_1 to reflect a strong effect of the user cost of capital on investment. However, the simple specification they used combines two effects in a single term: (1) the effect of changes in output demand (Y) and (2) the effects of changes in the user cost of capital. The collapsing of these two effects into one is an idiosyncrasy of the Cobb-Douglas production function that does not hold for more general specifications.

Eisner (1969) uses a specification that separates the effects of output from those of user cost. In this specification, he finds that output has a very strong effect on investment, but that the effects of user cost are much smaller and less significant. Based on this evidence, Eisner argues that the Cobb-Douglas production function is an inappropriate specification and that Hall and Jorgenson's measured effects were mostly due to output changes rather than to the sensitivity of investment to the user cost of capital.[30]

Empirical specification of adjustment costs and q models

The neoclassical model provides a benchmark by solving for the optimal path of capital when there is no uncertainty or adjustment costs. These assumptions are reasonable for a steady state where demand and prices are stable. The neoclassical model is less suited to describing the dynamics of investment in an environment where these variables fluctuate. As we saw in the previous section, empirical work based on the neoclassical model added on an *ad-hoc* distributed lag such as (18) in order to try to capture the dynamics of the adjustment of the actual capital stock to the steady-state desired level based on current values of demand and prices.

Among the objections to the neoclassical model are the ad hoc nature of the lag structure between investment and the revenue/user cost variable. This lag presumably reflects costs of adjusting the capital stock immediately to its optimal level, since if no such costs existed there would be no reason why K would not exactly equal K^*.[31]

Beginning with the seminal work of Lucas (1967), Gould (1968), and Treadway (1969), economists began to incorporate adjustment costs explicitly into firms' dynamic profit-maximization problem. These costs are typically represented as a func-

---

capital goods once they have been installed. This means that adjustment of factor proportions occurs only in *new* capital, which could slow down aggregate adjustment considerably. See Bischoff (1971) for an analysis of a "putty-clay" model incorporating these effects.

[30] To read more of the debate, see Hall and Jorgenson (1969) for a reply and Eisner (1970) for a further comment.

[31] Early studies of adjustment costs include Eisner and Strotz (1963), Rothschild (1971), and Nerlove (1972).

tion of the flow of investment and the existing stock of capital, $\Psi(I, K)$.[32] If adjusting the capital stock rapidly is costly, then it is no longer possible to simplify the capital decision in the way that Jorgenson did, making the desired capital stock at time t depend *only* on prices and demand at that moment. Adjustment costs means that it will be costly for a firm to "undo" high or low levels of today's investment if they turn out to be wrong for tomorrow's prices and demand levels. Thus, firms facing adjustment costs will take expected future demand and prices into account in making today's investment decisions.

Most of the early theoretical work on adjustment cost models was done in continuous time. This is the q model discussed in Romer's Chapter 9. However, because empirical data are available only over discrete time intervals, models for empirical implementation have tended to be in discrete time instead. The discrete-time model described here follows the exposition of Chirinko (1993).

To incorporate adjustment costs, we consider a maximand such as

$$V_0 = E_0\left\{\sum_{t=0}^{\infty}(1+r)^{-t}\left[F\left(K_t, L_t; \varepsilon_t\right) - \Psi\left(I_t, K_t; \varepsilon_t\right) - w_t L_t - p_t^K I_t\right]\right\}, \tag{21}$$

with $w_t \equiv W_t/P_t$ being the real wage and ε_t being a random shock that can affect both productivity of production and the cost of adjustment.[33] The firm's objective, as before, is to choose paths for K and L that maximize (21) subject to the dynamic evolution of the capital stock.

Maximizing using dynamic methods leads to the following first-order conditions for investment and labor input at time t:

$$\begin{aligned} &E_t\left[F_L\left(K_t, L_t; \varepsilon_t\right) - w_t\right] = 0, \\ &E_t\left[\Lambda_t - p_t^K - \Psi_I\left(I_t, K_t; \varepsilon_t\right)\right] = 0, \end{aligned} \tag{22}$$

where

$$\Lambda_t \equiv \sum_{s=0}^{\infty}\left(\frac{1-\delta}{1+r}\right)^s\left[F_K\left(K_{t+s}, L_{t+s}; \varepsilon_{t+s}\right) - \Psi_K\left(I_{t+s}, K_{t+s}; \varepsilon_{t+s}\right)\right] \tag{23}$$

and subscripts on the F and Ψ functions denote partial derivatives. Λ_t is the present value of the real marginal benefit the firm receives from an additional unit of capital installed at time t. The bracketed term in (23) shows the two components of this benefit: (1) the marginal product of capital F_K, which measures the additional output the firm gets from the added capital, and (2) the marginal reduction in adjustment costs

[32] Specifications of adjustment costs as quadratic in the level of investment or in the ratio of investment to capital have been popular, although recent evidence discussed in a later section suggests that nonconvex specifications incorporating fixed costs may be more realistic.

[33] The real interest rate is used to discount equation (21) rather than the nominal rate because the cash flow expression in (21) is in real terms.

in future periods $-\Psi_K$, which arises because having more capital in the future may lower future adjustment costs. These terms are discounted using both the depreciation rate (because a decreasing share of currently installed capital will remain as time passes) and the real interest rate (for the usual present value reasons).

With Λ_t interpreted as the marginal benefit from adding a unit of capital, the equations of (22) have a ready marginal-benefit-equals-marginal-cost interpretation. The first is simply asserts that the firm will hire labor up to the point where the expected marginal product of labor equals the real wage. The second says that it will invest in capital up to the point where the expected marginal value of an additional unit of capital equals the sum of the price of the new capital good and the marginal adjustment cost incurred by investing now.

The adjustment-cost function can be specified in a variety of ways. If adjustment costs are to explain why firms do not adjust the actual capital stock *rapidly* in response to changes in the desired stock, then these costs must levy a disproportionate penalty on *large* changes in the capital stock. In words, adjustment costs must increase more than proportionally with investment. In mathematical terms, the adjustment-cost function must be convex in investment—the second derivative Ψ_{II} must be positive.[34] A common specification in empirical applications is

$$\Psi\left(I_t, K_t; \varepsilon_t\right) = \frac{\alpha}{2}\left(\frac{I_t}{K_t} - \varepsilon_t\right)^2 K_t, \tag{24}$$

where α is a coefficient measuring the marginal cost of adjustment.

With adjustment costs specified as in (24), the optimal investment equation reduces to the simple form of

$$\frac{I_t}{K_t} = \frac{1}{\alpha}\left[E_t\left(\Lambda_t\right) - p_t^K\right] + \varepsilon_t. \tag{25}$$

Equation (25) is of limited value for empirical applications because $E_t(\Lambda_t)$ depends on the expected marginal product of capital at all future dates. However, if markets function efficiently, then prices of equities (shares of stock in firms) should be related to expectations of future profits and productivity. Under some restrictive assumptions, measurable data on share prices can proxy for the bracketed term in (25), making it a useful empirical specification.

Economists at least as far back as Keynes have recognized the potential importance of share prices as a determinant of investment. Modern interest in this linkage was rekindled in a famous paper by James Tobin, in which he criticized the sim-

[34] Convex adjustment costs imply that firms will spread the investment associated with a change in desired capital over multiple periods. However, recent evidence has suggested that non-convexities in adjustment costs may be important in explaining other aspects of investment behavior at the micro level. We examine this literature in a later section.

plistic role of asset markets in the basic IS/LM model. Tobin (1969) argued that "the rate of investment—the speed at which investors wish to increase the capital stock—should be related … to *q,* the value of capital relative to its replacement cost." High share prices make it cheaper for firms to finance new investment by issuing new shares. Conversely, low share prices reduce the proceeds from share issue and make it cheaper for firms looking to expand to buy up additional capacity by taking over (undervalued) existing firms rather than investing in new capital.

Empirically, the q theory of investment is attractive because Tobin's q can be measured on a firm-by-firm or aggregate basis. A firm's q is the ratio of the stock-market value of the firm (share price times number of shares outstanding) to the replacement value of its installed capital stock. Difficult issues arise with the treatment of various categories of preferred stock, options, and convertible bonds, and with the measurement of the replacement cost of old capital goods. Despite these obstacles, empirical economists quickly began using simple measures of q in investment equations.

Theoretical work quickly developed a connection between Tobin's q and the right-hand side of equation (25).[35] However, a fundamental difference soon became apparent. The expression $E_t(\Lambda_t)$ is the present value of the *marginal* product of an *additional* unit of capital. If markets function efficiently, then a firm investing in one unit of capital should see its market value increasing by this amount.

In contrast, stock market valuations reflect the *average* product of capital. Each share of a firm's stock is an ownership claim on the firm and its prospective profits. If a firm issues additional shares, the owners have claim not to the additional profit that might result but to a proportional share of all profits. Thus, share value—even for newly issued shares—is based on the expected *average* rather than *marginal* product of capital.

Mathematically, if we denote the nominal value of a firm's outstanding shares at time t by Ω_t, then

$$\text{Average } q = \frac{\Omega_t / p_t^K}{K_t},$$

$$\text{Marginal } q = \frac{\partial\left(\Omega_t / p_t^K\right)}{\partial K_t} = \frac{E_t \Lambda_t}{p_t^K}.$$

The last expression for marginal q is closely related to the investment equation (25). When it is larger than one, then the bracketed expression is positive and the firm will desire positive investment. Empirical investment equations are often estimated using $q - 1$ as the regressor.

[35] Mussa (1977) and Abel (1980) are major contributions.

But what is the relationship between marginal q, which is *the* important determinant of investment in the adjustment-cost model, and average q, which we can observe from stock-market data? Hayashi (1982) showed that marginal and average q are equal under quite restrictive conditions, including perfect competition and constant returns to scale. Based on this result, many economists have estimated investment functions relating I/K to average q. We now examine some of these results.

Empirical results using average q

According to equation (25), (marginal) q should be a "sufficient statistic" for describing investment behavior: once the effects of the current value of q are accounted for, no other variables should have independent effects on investment. By this criterion, the empirical performance of the q model has been unsatisfactory. Three basic problems are common in empirical studies:

- In the adjustment cost model, all relevant determinants of investment work through their effects on q. Empirical studies nearly always find that variables such as output, capacity utilization, profits, cash flow, and government investment policies have independent effects on investment apart from any effects they have on q. (As in the neoclassical model, taxes and credits on investment and capital income lead to a more complex definition of q; these issues are discussed below.)
- Although current (average) q has a significant positive effect on investment, the estimated coefficient on q is quite small. According to (25), the reciprocal of the coefficient on $q - 1$ can be interpreted as the marginal cost of adjustment. Estimated coefficients in many studies would imply unreasonably high adjustment costs.
- The effect of q on investment seems to involve substantial lags. The estimated lags in the q-investment relationship are often longer than can plausibly be explained by delivery and construction lags. This problem is familiar from our discussion above of the empirical studies using the neoclassical model.
- Even with lagged values of q and other variables, the error in aggregate investment equations is usually highly serially correlated. This is often attributed to the effects of omitted variables that are highly persistent.

A prominent early empirical application of the q model to aggregate investment data was von Furstenberg (1977). He estimates a regression of I/K for nonfinancial firms on lagged values of q and other variables using quarterly aggregate data for the United States from 1952 to 1976. Von Furstenberg corrects q for the investment tax credit and includes regressors that measure changes in other aspects of capital taxation.[36] To test whether q effectively summarizes all the information relevant to firms' investment decisions, he adds a measure of capacity utilization (CU).

Von Furstenberg's estimates the following regression:

$$\frac{I_t}{K_t} = \cdots + \underset{(0.17)}{0.047} \sum_{i=1}^{7} w_i q_{t-i} + \underset{(0.023)}{0.060} \sum_{i=1}^{4} v_i CU_{t-i}, \tag{26}$$

where the constant term and terms relating to tax variables and equipment share are not reported.[37] The lag weights w and v in equation (26) sum to one and are constrained to lie along a second-order polynomial. Because the lag weights sum to one, the coefficient in front of the summation reflects the long-run effect of a change in the respective right-hand variable on the investment rate.

While the coefficient on q is positive and statistically significant, capacity utilization also has a significant positive effect. Thus, utilization must be proxying for some aspect of prospective profitability that is not being adequately captured by measured q. Moreover, the coefficient on the sum of the q variables (0.047) implies that the adjustment-cost parameter α is $1/0.047 \approx 21$. The marginal adjustment cost of an additional dollar's worth of investment is α (I/K). For a typical value of I/K of 0.15, this implies that the last dollar of firms' investment typically cost $21 \times 0.15 = 3.15$ dollars in adjustment costs!

Similar results are reported by Blanchard and Wyplosz (1981), who estimate the following equation over a quarterly U.S. sample from 1953 to 1978.[38]

$$\frac{I_t}{K_t} = -0.064 - \underset{(0.011)}{0.009} q_t + \underset{(0.007)}{0.019} q_{t-1} + \underset{(0.006)}{0.013} q_{t-2} + \underset{(0.005)}{0.012} q_{t-3} + \underset{(0.042)}{0.162} \frac{Y_t}{K_t}. \tag{27}$$

Equation (27) is estimated under the assumption of a first-order autocorrelated error with estimated coefficient $\hat{\rho} = 0.97$. Both the severe autocorrelation and the significant estimated coefficient on current output suggest that q is not capturing all of the incentives for firms to invest. Adding up the four coefficients on current and lagged q gives 0.035, which corresponds to $\alpha = 1/0.035 \approx 29$—an even larger estimate for adjustment costs than von Furstenberg's.

Von Furstenberg incorporates tax effects by including tax variables explicitly as regressors. However, as with the neoclassical theory, the simple version of the q theo-

---

[36] He also includes a variable measuring the share of equipment in the total capital stock, which has been increasing steadily throughout the postwar period. His dependent variable covers both equipment and structures. The increasing relative importance of equipment, which has a shorter life span than structures, would imply that his dependent variable should increase over time. Including this variable attempts to correct for the increase in the dependent variable due solely to this effect.

[37] von Furstenberg (1977), Table 4, equation 4.5.

[38] See Blanchard and Wyplosz (1981), Table 3, equation (I.4). The version of q used in this study does not correct for taxes.

ry of investment can be augmented to derive a tax-corrected measure that takes full account of taxes and credits. Summers (1981), in a paper discussed in Romer's Section 9.6, develops a tax-adjusted q that has been used in many investment studies, incorporating the effects of investment tax credits, taxes on corporate profits, capital gains, and dividends, and the tax treatment of depreciation expenses. The tax-adjusted version of $q - 1$ is usually denoted by Q. Because one is subtracted in the calculation of Q, a value of $Q = 0$ corresponds to $q = 1$ and zero desired net investment.

Summers compares regressions of I/K on an uncorrected $q - 1$ with those on Q for a 1948–78 annual sample.[39] He finds that Q explains fluctuations in investment slightly better than $q - 1$ and that when both are included together in a regression, the estimated coefficient on $q - 1$ becomes slightly negative and insignificantly different from zero while the coefficient on Q remains significantly positive and essentially unchanged by the addition of $q - 1$ to the regression. However, the magnitude of the estimated coefficients on Q are even smaller than those for $q - 1$. When he includes a lagged value of Q as well as the current value, the sum of the coefficients is 0.025, corresponding to $\alpha = 40$.

To summarize, the empirical support for the q model of investment is quite thin. Although q is usually found to have a significant positive effect on investment, the coefficients are so small that the implied lag lengths and adjustment costs are unrealistic. Shapiro (1986) argues that large estimates of adjustment costs are inevitable given the amount of variation in stock prices relative to investment: "The stock market is much more variable than investment. The q theory … would have investment respond to all of these changes except for adjustment costs. Therefore, estimated adjustment costs must be very high to rationalize the relatively small response of investment to changes in the stock market."

Alternative strategies for estimating the adjustment-cost model

How much of the empirical disappointment of the q model is due to mismeasurement of marginal q in the estimation of investment equations? There are several reasons for thinking that this may be a problem. One is the above-discussed issue of average vs. marginal q. A second is questions about whether fluctuations in stock prices accurately reflect expectations about future firm profitability. Economists have tried a couple of alternative methods of estimating the adjustment-cost model, and in some cases have generated more plausible results. However, the empirical case in support of the adjustment-cost model is still tenuous.

[39] See Summers (1981), Table 5. Equation 5-6 is described above. Summers includes no variables other than $q - 1$ or Q.

Abel and Blanchard (1986) address the mismeasurement issue by constructing a measure of marginal q based on forecasts of future marginal profitability. They follow the constant-returns and perfect competition assumptions of Hayashi (1982), which imply that marginal and average profit from an additional unit of capital are equal. They then construct variables for marginal/average profit and the discount rate and use vector autoregression methods to model the evolution of these variables over time in response to their own past movements and the movements of other related variables.

Their calculated marginal q series picks up the effects on the present value of expected future profitability of changes in the marginal product of capital, the cost of capital, and the discount rate. Of these, they find that fluctuations in the cost of capital account for the largest measured share of variation in q.

Using aggregate data, Abel and Blanchard (1986) obtain results that are quite similar to those described above. The coefficient on their q variable (adjusted for taxes) is so small that it implies unrealistically large marginal adjustment costs. They also find that other variables (output and cash flow) add significant predictive power to the investment equation, so their q fails to capture all of the relevant effects on investment. The error term of their regressions is also strongly serially correlated, pointing to the possibility of additional omitted variables.

Gilchrist and Himmelberg (1995) find somewhat more positive results for the adjustment-cost model using firm-level data with a forecast-based marginal q variable. Although it has a relatively large standard error, their estimated coefficient on q is 0.183 when they use marginal q, compared with 0.050 with the more traditional average q based on stock prices. The resulting estimate of $\alpha = 1/0.183 \approx 5.5$ still implies very large marginal adjustment costs, but is considerably smaller than the estimates considered so far.

Abel (1980) pioneered another approach to estimating the adjustment-cost model. Rather than solving the first-order conditions (22) implicitly for an investment function, he worked exclusively with the second of these conditions, the so-called Euler equation. Although the Euler equation itself includes unobserved expectations of future variables, Abel showed that a clever differencing operation similar to that used by Koyck (1954) for the accelerator model could reduce the infinite summations of the Euler equation to involve only current variables and those one period ahead.

Abel's investment function relates current investment to future investment, the current marginal product of capital, the ratio of the future to current user cost of capital, and a term reflecting the effect of new information in the next period on the expected marginal product of capital discounted into the future. Abel estimates this investment function using a variety of econometric techniques, which yield a wide range of elasticities of investment with respect to q. His preferred regressions place the elasticity in the range of 0.5 to 1.1 (Abel (1980)), which is considerably larger than the estimates from q models based on stock market values.

Most subsequent work using Euler equations has taken advantage of the subsequently developed generalized method of moments (GMM) estimators.[40] These estimators are particularly well suited to the estimation of dynamic Euler equation models with rational expectations.

Pindyck and Rotemberg (1983) use GMM methods to estimate an adjustment-cost model. Pindyck and Rotemberg's model is not formally a q model. Rather than allowing the level of output to be determined by the firm in order to maximize profits, they assume that the firm's output is given exogenously and examine the firm's choice of inputs given the exogenous path of output. Demand changes and factor price changes are treated asymmetrically, so investment depends on output and the cost of capital independently rather than jointly through q.

Pindyck and Rotemberg estimate dynamic factor demand equations for blue-collar and white-collar labor, equipment, and structures using annual U.S. manufacturing data for 1949–76. They find marginal adjustment costs for equipment and structures to be 23 cents and 34 cents per dollar of investment, respectively. These values are much more plausible than those of the market-based q models discussed above.

The long-run elasticity of equipment demand with respect to its cost is estimated by Pindyck and Rotemberg to be –0.52; the long-run cost elasticity of structures is –0.16. Elasticities of equipment and structures demand with respect to output are 1.1 and 0.5, respectively. These elasticities suggest a fairly strong effect of both output and capital cost on the demand for capital.

Matthew Shapiro, in a pair of papers, uses GMM and similar methods to estimate the Euler equations of an aggregate production model with adjustment costs using quarterly U.S. manufacturing data from 1955 to 1980. Shapiro (1986a) includes adjustment costs for capital, labor employment, and hours worked. Shapiro (1986b) also incorporates variable rates of capital utilization.[41] His estimates of the marginal adjustment cost are even smaller than Pindyck and Rotemberg's. The marginal cost of a dollar of investment in both studies is less than 10 cents. In Shapiro (1986), he reports a long-run elasticity of the capital stock with respect to the cost of

[40] GMM estimation was introduced by Hansen (1982). Among the prominent studies employing GMM methods to estimate the investment Euler equation are Pindyck and Rotemberg (1983), Shapiro (1986), Hubbard and Kashyap (1992), Whited (1992), and Bond and Meghir (1994).

[41] In this paper, Shapiro uses non-linear three-stage least squares, which is similar to GMM under appropriate assumptions.

capital as –0.31, within the range of Pindyck and Rotemberg's estimates for equipment and structures.[42]

Using tax reforms as natural experiments

Although many of the empirical studies discussed above use a variant of q that is corrected in one way or another for tax effects, several studies have focused specifically on tax reforms. Unlike other variables affecting q, we can easily observe the timing and magnitude of changes in taxes and more confidently assume that tax reforms are exogenous to the investment decision. Thus changes in federal tax laws may provide useful "natural experiments" for identifying the cost sensitivity of investment spending.

Cummins, Hassett, and Hubbard (1994) use yearly cross-sectional samples of individual firms to estimate the effects of tax-adjusted Q on investment in individual years. They estimate a separate cross-sectional regression for each year from 1964 to 1988. Because tax laws apply differently to structures and equipment and to capital goods of different kinds within those classes, there is substantial cross-sectional variation in the tax component of Q. To explicitly highlight the effects of changes in q arising from tax reforms and to avoid problems of endogeneity of other components of q such as stock prices and capital goods prices, their Q variable is constructed using *current-year* tax values and *two-year-lagged* values of the market components of q.

Table 2. Estimated effects of Q in tax-reform and other years.

	Number of years	Average of Q coefficients	Range of Q coefficients	Average $\lvert t \rvert$ on Q coefficients
No tax reform	13	0.056	–0.119 to 0.138	0.77
Minor tax reform	9	0.555	0.446 to 0.742	5.14
Major tax reform	4	0.639	0.470 to 0.874	5.33

Source: Cummins, Hassett, and Hubbard (1994), Table 5.

As shown in Table 2, their estimated coefficients show much stronger effects of Q on investment in tax-reform years than in years with no tax changes. In every tax reform year, Q has a statistically significant effect on the investment rate, with coefficients ranging from 0.446 to 0.874. In non-reform years, Q is never significant and the largest reported coefficient is 0.138.

[42] He also reports a larger elasticity of –0.97 of capital utilization with respect to capital cost. The utilization rate responds immediately, whereas adjustment costs imply a gradual adjustment of the capital stock.

Recall that traditional q regressions typically find coefficients smaller than 0.05, which is near the middle of the range of results for non-reform years but much smaller than the estimated coefficients when explicit jumps in Q as a result of tax reforms are used. If we take the tax-reform-year coefficients in Table 2 as an estimate of $1/\alpha$ in equation (25), the implied marginal adjustment costs are between 5 and 12 cents per dollar of investment.

Imperfect capital markets: Cash flow and related variables

Early investment studies often examined the role of corporate cash flow, profits, or retained earnings. For example, Tinbergen (1938) finds that changes in investment are more closely associated with changes in profits than with changes in output (as predicted by the acceleration principle) for the industries he examines.

However, the famous theorem of Modigliani and Miller (1958) shows that when capital markets are "perfect" and there are no tax distortions, a firm's financial status would not affect its investment decisions. Both the q theory and the neoclassical theory assume that firms have access to unlimited borrowing at the given interest rate. Under these conditions, even a firm that has low or negative current cash flow and little cash reserve can invest if capital is expected to have a high future marginal profitability.

While assumptions such as perfect capital markets help us develop theories based on a simplified, stark environment, deviations from such assumptions can be empirically important. The general dissatisfaction with the empirical performance of the basic theories has fueled the development of a large literature examining the effects of limited access to credit—so-called financing or liquidity constraints—on firms' investment spending.

Financial constraints are usually attributed to asymmetry of information about firms' investment prospects and to moral-hazard problems associated with monitoring the behavior of indebted firms. The firm's owners and managers base their desired investment decision on their assessment of the profit potential of added capital. However, it may be very difficult for an outsider (for example, a banker or a bond-rating agency) to evaluate the firm's assessment. Thus, outside lenders are likely to demand a risk premium inversely related to the degree of confidence they have in the firm's prospects and directly related to the likelihood of default should outcomes prove disappointing.

Firms with large stocks of tangible assets that could be used as collateral, those with a long track record of profitable investment decisions, and those investing in industries and technologies that are well established may have relatively little difficulty attracting lenders on reasonable terms. However, small, new firms in new industries, with assets that are largely intangible are likely to pay a substantial risk premium if they can borrow at all.

Because of these information problems, firms may face a "hierarchy" of financing options, with internal finance being the least costly, debt financing next, and financing by issuing new equity the most expensive.[43] While all firms may find internal finance to be cheapest, the difference between the cost of external and internal finance is likely to be largest for small, new, and low-tangible-asset firms. Thus, such firms are more likely than better-established firms to be constrained to internal financing of investment.

The empirical literature on financial constraints and investment has attempted to classify firms as constrained or unconstrained based on non-investment criteria, then to compare the investment behavior of the two groups of firms. For constrained firms, liquidity variables such as cash flow, profits, or stocks of liquid assets should affect investment strongly while q and the cost of capital may play a lesser role. Unconstrained firms should invest when q or revenue-cost conditions are favorable, regardless of their liquidity situation.

In a seminal study, Fazzari, Hubbard, and Petersen (1988) (FHP) group firms by the share of their profits that have been paid out in dividends. Firms that have ready access to borrowed capital would not need to retain earnings in order to invest and therefore could issue dividends more liberally. In contrast, each dollar of dividends issued by a constrained firm reduces its potential investment funds. Therefore, FHP considered low-dividend firms to be potentially liquidity constrained and high-dividend firms to be unconstrained. Later studies have used a variety of criteria to categorize firms' access to credit, including firm size and age, financial affiliations with other firms, bond ratings, and liquid assets.

Based on a 1970–84 sample, FHP grouped their 422 manufacturing firms into three categories: Class 1 firms paid out less than ten percent of their income in dividends in at least ten of the fifteen sample years; class 2 firms paid out ten to twenty percent; and class 3 firms paid out dividends of more than twenty percent of income. Based on the dividend-income ratio, class 1 firms were argued to be the most likely to be liquidity constrained and class 3 firms least likely.

The reported characteristics of the three classes of firms correspond closely to the criteria discussed above. In comparison to the other groups, the 49 firms in class 1

[43] Two factors might lead debt financing to be cheaper than equities. First, bank financing (in the United States) is strictly executed through debt. Compared with the general financial market, bankers may have better information about small firms, so bank loans are likely to be more readily (and cheaply) available to small firms than market debt or equity. Second, the double taxation of dividends raises the cost of equity finance relative to debt finance. See Fazzari, Hubbard, and Petersen (1988) for a brief exposition of the "hierarchy of finance" model.

tend to be small and to have rapidly growing sales, high investment rates, higher measured q values, and higher debt.[44]

FHP test for the effects of liquidity constraints on investment by running a regression of the form

$$\frac{I_{i,t}}{K_{i,t-1}} = \beta_0 + \gamma_i + \theta_t + \beta_1 Q_{i,t} + \beta_2 \frac{CF_{i,t}}{K_{i,t-1}} + u_{i,t}, \tag{28}$$

where Q is tax-adjusted $q - 1$ and CF is the firm's net cash flow. The terms γ_i and θ_t are firm-specific and year-specific effects that are included to compensate for possible missing variables that vary strictly across firms or over time.

Table 3 shows FHP's estimated coefficients on Q and CF/K for the full fifteen-year sample period. Two results are strongly consistent with the presence of liquidity constraints and with the hierarchy of finance model in general. First, the cash-flow coefficient is significantly positive for all firms, but it is largest for the firms that are most likely to be constrained. Second, the class 1 firms that FHP regard as being most likely to be financially constrained have the smallest investment sensitivity to Q. This is consistent with a situation in which even when a high Q value suggests highly profitable investment opportunities, these firms may be unable to take advantage of them.[45]

Table 3. Regression coefficients on Q and cash flow, 1970–84 sample.

	Class 1	Class 2	Class 3
Q	0.0008 (0.0004)	0.0046 (0.0009)	0.0020 (0.0003)
CF/K	0.461 (0.027)	0.363 (0.039)	0.230 (0.010)
$\bar{R}^2$	0.46	0.28	0.19

Source: Fazzari, Hubbard, and Petersen (1988), Table 4.

FHP's evidence in favor of financing constraints depends on two crucial assumptions, both of which have been challenged. First, sensitivity of investment to cash flow must result from financing constraints rather than from some other cause, such as a possible effect of current cash flow on expected future demand. Second, firms' dividend ratios must be a good measure of their access to borrowed capital.

[44] See Fazzari, Hubbard, and Petersen (1988), Tables 2 and 3.

[45] Results for sub-periods of the sample confirm this pattern of results with one additional striking result: The differences between classes are stronger for the first few years of the sample than for the sample as a whole. The class 1 firms may be "growing up" by the end of the sample period and beginning to behave more like class 2 and 3 firms.

Gilchrist and Himmelberg (1995) address the first assumption by looking at the relationship between current cash flow and investment opportunities. They argue that Tobin's *Q* might be a particularly poor measure of future profitability for precisely those firms that FHP's criteria classify as constrained: "Since the firms identified *a priori* as financially constrained are typically newer, smaller, and faster growing than the other firms in the sample, the stock market is less likely to have accumulated the usual stock of knowledge that arises through detailed evaluation and monitoring of firms over time. Thus, Tobin's *Q* might contain less information about investment opportunities for these relatively 'unseasoned' firms than it does for firms that have been identified as unconstrained" (Gilchrist and Himmelberg (1995)).

To test whether the cash-flow effect is merely proxying for profit opportunities that are poorly measure by *Q*, Gilchrist and Himmelberg use vector autoregression forecasts to construct a "fundamental *Q*" measure of the discounted stream of marginal profits. Since cash flow is one of the variables used in the VAR to predict future profitability, any effects of cash-flow on investment that arise indirectly through the cash-flow-future-profit channel will be picked up in fundamental *Q*. If financing constraints are unimportant and this is the only source of FHP's cash-flow effect, then cash flow should have now marginal power to explain investment once the effects of fundamental *Q* have been accounted for.

Gilchrist and Himmelberg divide their sample of firms into constrained and unconstrained based on a variety of alternative criteria: dividend ratios, size, whether the firm has a bond or commercial paper rating. For the constrained firms, cash flow has a strong positive effect on investment (as in FHP) and the coefficient of fundamental *Q* is larger and more significant than a corresponding coefficient on the traditionally measured Tobin's *Q*. This suggests that, for these firms, Tobin's *Q* may indeed be a poor measure of prospective profitability. However, this mismeasurement does not account for the significance of cash flow, since that result is present even when fundamental *Q* is used. For the unconstrained firms, there is little evidence of a significant cash-flow effect with the fundamental *Q* measure. The authors interpret this result as consistent with the original FHP conclusion: some firms are liquidity constrained and rely on cash flow to finance investment.

Whited (1992) develops a formal model of investment by a firm subject to financing constraints. She estimates the Euler equations of this model using GMM methods for two groups of firms classified depending on whether or not they have bond ratings. Within each group, the "shadow cost of external finance" is modeled as depending on the firm's debt ratio and the magnitude of its interest expense. For the group with bond ratings, there is some marginal statistical evidence of financial constraint, but the evidence for the non-rated group is overwhelming. In both cases, increases in debt and interest expenses seem to increase the degree to which the firm is financially constrained.

In addition to Gilchrist and Himmelberg (1995) and Whited (1992), a substantial body of literature has emerged using alternative criteria to classify firms as constrained or unconstrained. Much of this literature has found results in support of those of FHP. However, we begin by discussing one that disagrees.

Kaplan and Zingales (1997) look specifically at the financial condition of FHP's Class 1 firms.[46] By detailed examination of the financial reports of these firms, they find that more than half of the firm-year observations were almost surely not financially constrained. Moreover, they find that firms within that group that faced more extreme financial constraints (for example, firms that were in arrears in payments on existing debt) had a *weaker* rather than stronger effect of cash flow on investment.[47] Based on this evidence, Kaplan and Zingales question the importance of financing constraints for investment.

Devereux and Schiantarelli (1990) examine the effects of cash flow and other variables using firm-level British data. They use size of firm as their main classification variable and find that cash flow has a significant effect for all sizes of firms, but (counter to FHP) it has the greatest effect for the largest firms.

Schaller (1993) uses firm age, the degree of concentration of ownership, and the availability of collateral as indicators of the likelihood of financial constraints. He argues that information problems will be less severe for older firms and for firms with concentrated ownership, where principal/agent problems between owners and managers may be less severe. Having more "standardized assets," such as those in the manufacturing industry, can imply the availability of more collateral, which should also mitigate financing constraints. Schaller uses a methodology similar to FHP and finds support for the existence of financing constraints based on all three criteria.

Several studies have examined the effects of inter-firm affiliations on investment. Hoshi, Kashyap, and Scharfstein (1991) look at Japanese firms and classify them as members of keiretsu conglomerates and nonmembers. Since keiretsu groups usually include a large bank that serves as a source of funds for members, individual affiliates are unlikely to be financially constrained. They find that cash flow has a strong effect on investment for non-keiretsu firms, but a small and statistically insignificant effect for those that are affiliated with a keiretsu. Galeotti, Schiantarelli, and Jaramillo (1994) and Schiantarelli and Sembenelli (2000) find similar results for Italian firms, classifying firms as constrained or unconstrained by size or by presence of connections to national or international conglomerates.

---

[46] This paper is discussed in Romer's Section 9.10.

[47] Fazzari, Hubbard, and Petersen (2000) respond that the most financially constrained firms most likely had to apply any positive cash flow to paying off existing debt, hence were unable to use even current cash proceeds for new investment. See also the rejoinder by Kaplan and Zingales (2000).

Lamont (1997) uses the dramatic 1986 decline in world oil prices as a natural experiment to attempt to identify financially constrained firms. He argues that if financial constraints are important, then non-oil-related companies affiliated with oil-industry firms should have invested less in 1986 than other firms in their respective industries. He finds that oil-related affiliates did indeed invest less and that the decline in investment in 1986 was strongest among firms that were subsidized in 1985 by then-cash-rich oil affiliates.

Cash windfalls provide another opportunity to observe the possibility of firms changing behavior in response to a change in financial constraints. Blanchard, Lopez de Silanes, and Shleifer (1994) look at the behavior of thirteen firms who have recently received large windfalls from settlements of lawsuits. They find that these firms spent very little of their windfalls on investment, but they attribute this to the relatively distressed condition of the firms.[48]

Effects of irreversibility, uncertainty, and non-convex adjustment costs

Beginning in the late 1980s, major advances occurred in the literature on investment. These breakthroughs extended the standard neoclassical and q models to relax questionable assumptions. Three of the important strands of this literature have dealt with (1) the implications for investment of irreversibility—once a factory is built it cannot by "unbuilt," (2) a modification of the adjustment-cost theory to allow fixed costs or other nonconvex aspects of adjustment costs, and (3) the effects of uncertainty on investment, which are important in their own right, but especially in combination with irreversibility and nonconvex adjustment costs. Empirical analysis of these issues is still fragmentary, but a flow of studies has begun to emerge. Some of these issues are introduced by Romer in Section 9.8.

Irreversibility is important for investment decisions because it leads to an asymmetry between having too little capital (which can be ameliorated through investment) and having too much (which can only be eliminated slowly through depreciation). Under certainty, irreversibility would affect a firm that knew with certainty that its desired capital stock would fall in the future, as for example when a firm is experiencing a surge in demand that is known to be temporary.

However, irreversibility becomes even more important when combined with uncertainty: firms that cannot be sure that current high levels of desired capital are permanent will be reluctant to undertake irreversible investment. In a book summarizing the seminal literature, Dixit and Pindyck (1994) describe the effects of irreversibility under uncertainty in great detail. They show that by investing in capital a firm forfeits an "option" to have a lower future capital stock. This option can be valued using standard techniques from the analysis of financial options. Because the firm

[48] All of the firms in their sample had values of q well below 1 and most sold off significant amounts of assets after the settlement.

forgoes this option when it invests, its value should be added to the cost of capital. Calibration studies have shown that the additional cost due to irreversibility can be substantial. McDonald and Siegel (1986) show that irreversibility may as much as double the required return on capital. Majd and Pindyck (1987) show that the presence of substantial lags in delivery and installation of new capital may increase the required return even more.

The effects of irreversibility are difficult to observe even in firm-level investment data. One testable implication of irreversibility is that increased uncertainty about the future marginal product of capital should raise the option-cost of investment, given the value of the traditional user cost of capital. This implies a negative effect of increased uncertainty on investment.[49] A second implication is that the faster is the trend growth in a firm's demand, the less important irreversibility will be. If demand is growing rapidly, a new capital project that turns out to be undesirable in the short run will soon be justified. In a slow-growing industry, it may take many years before demand catches up to a bulge of overinvestment.

Ferderer (1993) constructs a measure of the risk premium on long-term bonds to measure uncertainty about the prospective marginal product of capital.[50] He finds that aggregate U.S. investment is negatively affected by his uncertainty measure in either a model based on Q or in a neoclassical user-cost model.

Leahy and Whited (1996) study investment at the firm level, measuring uncertainty by VAR-based forecasts of the day-to-day variance in the firm's stock price. Using a sample of 600 U.S. manufacturing firms from 1982 to 1987, they find that their uncertainty measure is negatively associated with firm-level investment, but that uncertainty has no statistically significant effect once (average) q is included in the equation. They conclude that, in contradiction to the predictions of the irreversibility model, any effects of uncertainty on investment work through effects on stock prices that are measured in q.

Another approach to estimating the effects of uncertainty on investment is to attempt to measure firms' "hurdle rate," the expected rate of return required to induce them to undertake new investment. For a given level of the traditional cost of capital, irreversibility implies that increases in uncertainty should raise this hurdle rate by increasing the option-cost component.

To use this approach, one must be able to observe, at least approximately, firms' hurdle rates of return. Although no direct data exist on hurdle rates, Pindyck and

[49] Note, however, that there are other theories that predict a negative relation between uncertainty and investment, notably simple risk aversion on the part of firms.

[50] The risk premium is defined as the difference between the expected yield on a long-term bond over a given holding period and the expected return on a short-term bond (or a series of such bonds) over the same period. Ferderer uses surveys of expectations of future bond yields to construct the expected returns over a six-month holding period.

Solimano (1993) and Caballero and Pindyck (1996) attempt to extract information about hurdle rates from estimates of the ex-post marginal profitability of capital for each period and industry or country.[51] In theory, the marginal profitability of capital should never exceed the hurdle rate, since if it did then investment would increase until they were equal. Thus, for any cross-sectional unit, the upper tail of the observed distribution of ex-post marginal profitability may provide information about the hurdle rate. These studies use several alternative measures based on observed marginal profitability (maximum value, average of top decile values, average of top quintile values) to approximate the hurdle rate.[52]

Both studies find that high uncertainty (measured by the standard deviation of marginal profitability of capital) is associated with industries or countries with high hurdle rates by their measure. This is consistent with the predictions of the irreversibility model. Pindyck and Solimano (1993) go further and look at the effect of their uncertainty measure on investment. Uncertainty seems to have a strong negative effect on investment in their sample of developing countries, but among OECD countries, those with higher uncertainty do not seem to have significantly lower investment rates.

A second issue in modern empirical analysis of investment is the possibility that adjustment costs may be nonconvex—the adjustment cost per unit of investment may fall with the size of the investment project rather than rising as is assumed in the *q* model. This would lead to fewer, larger investment projects rather than to smooth adjustment.

Is investment smooth or lumpy? The answer probably depends on the level of time and cross-sectional aggregation at which one views the data. The rationale for convex-adjustment-cost models, which predict that firms will smooth investment so as to avoid large changes, is that new factories and the equipment installed in them typically require many quarters (several years) to plan and build. The flow of investment spending will occur smoothly over this period rather than being concentrated in one quarter. This smoothed, gradual response of investment to changes in the cost or marginal product of capital is apparent in the studies we have examined using aggregate investment data.

However, looked at over longer periods or with finer cross-sectional detail, investment seems very lumpy, especially investment in structures. Individual firms have many years in which their investment in structures is zero. When they do undertake a new investment project (which may last several years), it is typically quite

[51] Caballero and Pindyck (1996) use U.S. manufacturing industries as their cross-sectional unit; Pindyck and Solimano (1993) use thirty countries.

[52] Of course, marginal profitability is not observed directly, but must be calculated for each set of output and input observations using estimated production function coefficients.

large. This evidence suggests that the smoothing outcome predicted by the convex-adjustment-cost model may not be telling the whole story.

An alternative model, described in Abel and Eberly (1994), augments the usual convex adjustment costs with a fixed cost that the firm incurs whenever its investment rate is positive. Such adjustment costs can explain why firms would keep investment at zero for many periods (to avoid the fixed cost) then invest at a positive, but not extremely rapid, rate (to avoid high marginal adjustment costs) when new capacity is needed.[53]

Abel and Eberly's model predicts three possible investment regimes for the firm depending on the level of q. For very low levels of q, the firm may desire negative investment. In this range, a higher q means less desired disinvestment, giving the conventional positive relationship between q and I/K. For an intermediate range of q values, desired investment will be zero. Within this range, changes in q have no effect on investment. For high levels of q, desired investment is positive and the usual positive investment/q relationship will hold.

Barnett and Sakellaris (1998) examine the investment behavior of individual firms using an econometric specification that allows for three regimes as discussed above. Their findings suggest support for, but not strict conformation to, the Abel and Eberly model. They find strong evidence for three distinct regimes for different regions of q values. However, the pattern of investment-q sensitivities does not match the model's predictions. First, there are very few zero or negative investment rates observed in their sample. This may be due to the aggregation of investment to the firm level and to the aggregation together of many kinds of investment goods. They find that investment is sensitive to q in all three regimes with the sensitivity increasing with q for very low values, then decreasing in q as q gets larger. This S-shaped relationship between I/K and q suggests that investment is *most* (rather than least) sensitive to q at intermediate values.

More direct evidence on possible nonconvexities has come from a series of studies looking at the Longitudinal Research Database (LRD) created by the U. S. Bureau of the Census. The LRD contains panel data on production, investment, and other variables at the level of the *establishment* rather than the firm or industry.[54] Doms and Dunne (1998) show that zero investment is much more common at the establishment level than Barnett and Sakellaris (1998) found it to be at the firm level. Their typical investment episode at the plant level is highly concentrated in a single

[53] Abel and Eberly (1994) also incorporate irreversibility by allowing the price at which the firm can sell its installed capital goods to be lower (even after accounting for depreciation) than the price at which it purchases new capital goods.

[54] An establishment is a place at which production occurs, such as a factory or complex. Each firm may have several or many producing establishments, so the observations of the LRD are at a much finer level of detail than other data sets,

year, with some spillover into adjacent years but no strong tendency toward smoothness.

Caballero, Engel, and Haltiwanger (1995) use the LRD to examine the relationship between investment and the cost of capital. They define "mandated investment" to be the amount of additional capital the firm would like to have if adjustment were costless. Mandated investment is assumed to depend on the neoclassical cost of capital. The focus of the study is on the response of actual investment to mandated investment. They find that actual investment is much lumpier than mandated investment, which is consistent with a threshold model in which mandated investment "builds up" until it reaches a threshold at which the need for new capacity overcomes the fixed costs of adjustment, at which time the firm undertakes a large plant expansion.

The fine level of disaggregation and the explicit treatment of dynamics also allow Caballero, Engel, and Haltiwanger (1995) to estimate the sensitivity of (mandated) investment to the cost of capital without some of the potential distortions that arise in more aggregated data. Their estimates suggest that mandated investment is highly sensitive to the cost of capital, which contrasts with the weak and delayed responses typical of aggregate studies.

F. Works Cited in Text

Abel, Andrew B. 1980. Empirical Investment Functions: An Integrative Framework. *Carnegie Rochester Conference Series on Public Policy* 12:39-91.

Abel, Andrew B., and Olivier Blanchard. 1986. The Present Value of Profits and Cyclical Movements in Investment. *Econometrica* 54 (2):240-273.

Abel, Andrew B., and Janice C. Eberly. 1994. A Unified Model of Investment under Uncertainty. *American Economic Review* 84 (5):1369-84.

Allen, R.G.D. 1968. *Macro-Economic Theory*. London: St. Martin's Press.

Barnett, Steven A., and Plutarchos Sakellaris. 1998. Nonlinear Response of Firm Investment to Q: Testing a Model of Convex and Non-Convex Adjustment Costs. *Journal of Monetary Economics* 42 (2):261-288.

Bernanke, Ben S., and Mark Gertler. 1995. Inside the Black Box: The Credit Channel of Monetary Policy Transmission. *Journal of Economic Perspectives* 9 (4):27-48.

Bischoff, Charles. 1971. The Effect of Alternative Lag Distributions. In *Tax Incentives and Capital Spending*, edited by G. Fromm. Washington: Brookings Institution.

Blanchard, Olivier Jean, Florencio Lopez de Silanes, and Andrei Shleifer. 1994. What Do Firms Do with Cash Windfalls? *Journal of Financial Economics* 36 (3):337-360.

Blanchard, Olivier, and Charles Wyplosz. 1981. An Empirical Structural Model of Aggregate Demand. *Journal of Monetary Economics* 7 (7):1-28.

Bond, Stephen, and Costas Meghir. 1994. Dynamic Investment Models and the Firm's Financial Policy. *Review of Economic Studies* 61 (2):197-222.

Caballero, Ricardo J., Eduardo M. R. A. Engel, and John C. Haltiwanger. 1995. Plant-Level Adjustment and Aggregate Investment Dynamics. *Brookings Papers on Economic Activity* (2):1-39.

Caballero, Ricardo J., and Robert S. Pindyck. 1996. Uncertainty, Investment, and Industry Evolution. *International Economic Review* 37 (3):641-662.

Carver, T. N. 1903. A Suggestion for a Theory of Industrial Depressions. *Quarterly Journal of Economics*.

Chirinko, Robert S. 1993. Business Fixed Investment Spending: Modeling Strategies, Empirical Results, and Policy Implications. *Journal of Economic Literature* 31 (4):1875-1911.

Clark, J. Maurice. 1917. Business Acceleration and the Law of Demand: A Technical Factor in Economic Cycles. *Journal of Political Economy* 25 (3):217-235.

Cummins, Jason G., Kevin A. Hassett, and R. Glenn Hubbard. 1994. A Reconsideration of Investment Behavior Using Tax Reforms as Natural Experiments. *Brookings Papers on Economic Activity* 1994 (2):1-59.

Devereux, Michael, and Fabio Schiantarelli. 1990. Investment, Financial Factors, and Cash Flow: Evidence from U.K. Panel Data. In *Asymmetric Information, Corporate Finance, and Investment*, edited by R. G. Hubbard. Chicago: University of Chicago Press.

Dixit, Avinash K., and Robert S. Pindyck. 1994. *Investment under Uncertainty*. Princeton, N.J.: Princeton University Press.

Doms, Mark, and Timothy Dunne. 1998. Capital Adjustment Patterns in Manufacturing Plants. *Review of Economic Dynamics* 1 (2):409-429.

Eisner, Robert. 1969. Tax Policy and Investment Behavior: Comment. *American Economic Review* 59 (3):379-388.

———. 1970. Tax Policy and Investment Behavior: Further Comment. *American Economic Review* 60 (4):746-752.

Eisner, Robert, and Robert H. Strotz. 1963. Determinants of Investment Behavior. In *Impacts of Monetary Policy*. Englewood Cliffs, N. J.: Prentice-Hall.

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen. 1988. Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity* 1988 (1):141-195.

———. 2000. Investment-Cash Flow Sensitivities are Useful: A Comment. *Quarterly Journal of Economics* 115 (2):695-705.

Feldstein, Martin, and David K. Foot. 1971. The Other Half of Gross Investment: Replacement and Modernization Expenditures. *Review of Economcs and Statistics* 53 (1):49-58.

Ferderer, J. Peter. 1993. The Impact of Uncertainty on Aggregate Investment Spending: An Empirical Analysis. *Journal of Money, Credit and Banking* 25 (1):30-48.

Galeotti, Marzio, Fabio Schiantarelli, and Fidel Jaramillo. 1994. Investment Decisions and the Role of Debt, Liquid Assets and Cash Flow: Evidence from Panel Data. *Applied Financial Economics* 4 (2):121-132.

Gilchrist, Simon, and Charles Himmelberg. 1995. Evidence on the Role of Cash Flow in Reduced-Form Investment Equations. *Journal of Monetary Economics* 36 (3):541-572.
Goodwin, Richard M. 1948. Secular and Cyclical Aspects of the Multiplier and the Accelerator. In *Income, Employment and Public Policy: Essays in Honor of Alvin H. Hansen*. New York: W. W. Norton & Co.
Gould, J. P. 1968. Adjustment Costs in the Theory of Investment of the Firm. *Review of Economic Studies* 35 (1):47-55.
Hall, Robert E., and Dale W. Jorgenson. 1967. Tax Policy and Investment Behavior. *American Economic Review* 57 (3):391-414.
———. 1969. Tax Policy and Investment Behavior: Reply and Further Results. *American Economic Review* 59 (3):388-401.
Hansen, Lars Peter. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50 (4):1029-54.
Hayashi, Fumio. 1982. Tobin's Marginal *q* and Average *q*: A Neoclassical Interpretation. *Econometrica* 50 (1):213-24.
Hicks, J.R. 1950. *A Contribution to the Theory of the Trade Cycle*. Oxford: Clarendon Press.
Hoshi, Takeo, Anil Kashyap, and David Scharfstein. 1991. Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups. *Quarterly Journal of Economics* 106 (1):33-60.
Hubbard, R. Glenn, and Anil K. Kashyap. 1992. Internal Net Worth and the Investment Process: An Application to U.S. Agriculture. *Journal of Political Economy* 100 (3):506-534.
Jorgenson, Dale W. 1963. Capital Theory and Investment Behavior. *American Economic Review* 53 (2):247-259.
———. 1965. Anticipations and Investment Behavior. In *The Brookings Quarterly Econometric Model of the United States*, edited by J. S. Duesenberry, G. Fromm, L. R. Klein and E. Kuh. Chicago: Rand McNally.
Kaplan, Steven N., and Luigi Zingales. 1997. Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints. *Quarterly Journal of Economics* 112 (1):169-215.
———. 2000. Investment-Cash Flow Sensitivities are Not Valid Measures of Financing Constraints. *Quarterly Journal of Economics* 115 (2):707-12.
Koyck, L. M. 1954. *Distributed Lags and Investment Analysis*. Amsterdam: North-Holland.
Lamont, Owen. 1997. Cash Flow and Investment: Evidence from Internal Capital Markets. *Journal of Finance* 52 (1):83-109.
Leahy, John V., and Toni M. Whited. 1996. The Effect of Uncertainty on Investment: Some Stylized Facts. *Journal of Money, Credit and Banking* 28 (1):64-83.
Lucas, Robert E., Jr. 1967. Adjustment Costs and the Theory of Supply. *Journal of Political Economy* 75 (4):321-334.
Majd, Saman, and Robert S. Pindyck. 1987. Time to Build, Option Value, and Investment Decisions. *Journal of Financial Economics* 18 (1):7-27.
McDonald, Robert, and Daniel Siegel. 1986. The Value of Waiting to Invest. *Quarterly Journal of Economics* 101 (4):707-727.

Modigliani, Franco, and Merton H. Miller. 1958. The Cost of Capital, Corporation Finance and the Theory of Investment. *American Economic Review* 48 (3):261-297.
———. 1961. Dividend Policy, Growth, and the Valuation of Shares. *Journal of Business* 34:235-264.
Montgomery, Michael R. 1995. 'Time-to-Build' Completion Patterns for Nonresidential Structures, 1961-1991. *Economics Letters* 48 (2):155-163.
Mussa, Michael. 1977. External and Internal Adjustment Costs and the Theory of Aggregate and Firm Investment. *Economica* 44 (174):163-178.
Nerlove, Marc. 1972. Lags in Economic Behavior. *Econometrica* 40 (2):221-251.
Pindyck, Robert S., and Julio J. Rotemberg. 1983. Dynamic Factor Demands under Rational Expectations. *Scandinavian Journal of Economics* 85:223-238.
Pindyck, Robert S., and Andres Solimano. 1993. Economic Instability and Aggregate Investment. In *NBER Macroeconomics Annual 1993.*, edited by O. J. Blanchard and S. Fischer. Cambridge and London: MIT Press.
Rothschild, Michael. 1971. On the Cost of Adjustment. *Quarterly Journal of Economics* 85 (4):605-622.
Samuelson, Paul A. 1939. Interactions between the Multiplier Analysis and the Principle of Acceleration. *Review of Economcs and Statistics* 21 (2):75-78.
Schaller, Huntley. 1993. Asymmetric Information, Liquidity Constraints and Canadian Investment. *Canadian Journal of Economics* 26 (3):552-74.
Schiantarelli, Fabio, and Alessandro Sembenelli. 2000. Form of Ownership and Financial Constraints: Panel Data Evidence from Flow of Funds and Investment Equations. *Empirica* 27 (2):175-192.
Shapiro, Matthew D. 1986. Capital Utilization and Capital Accumulation: Theory and Evidence. *Journal of Applied Econometrics* 1 (3):211-234.
———. 1986. The Dynamic Demand for Capital and Labor. *Quarterly Journal of Economics* 101 (3):513-547.
Summers, Lawrence H. 1981. Capital Taxation and Accumulation in a Life Cycle Growth Model. *American Economic Review* 71 (4):533-44.
———. 1981. Taxation and Corporate Investment: A q-Theory Approach. *Brookings Papers on Economic Activity* 1981 (1):67-127.
Tinbergen, Jan. 1938. Statistical Evidence on the Acceleration Principle. *Economica* 5 (18):164-176.
Tobin, James. 1969. A General Equilibrium Approach to Monetary Theory. *Journal of Money, Credit and Banking* 1 (1):15-29.
Treadway, A. B. 1969. On Rational Entrepreneurial Behaviour and the Demand for Investment. *Review of Economic Studies* 36 (2):227-239.
von Furstenberg, George M. 1977. Corporate Investment: Does Market Valuation Matter in the Aggregate? *Brookings Papers on Economic Activity* (2):347-397.
Whited, Toni M. 1992. Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data. *Journal of Finance* 47 (4):1425-60.