Chapter 12

AGGREGATE INVESTMENT *

RICARDO J. CABALLERO

MIT and NBER

Contents

Abstract 814
Keywords 814
1. Introduction 815
2. Basic investment theory and findings 816
  2.1. Pre 1990s: Dismay 816
  2.2. "Econometrics": Cost of capital and q matter 818
    2.2.1. Long-run 820
    2.2.2. Short-run 821
3. Lumpy and irreversible investment 822
  3.1. Plant/firm level 822
    3.1.1. Microeconomic adjustment: characterization 823
    3.1.2. "Representative" problem 824
    3.1.3. A detour: q-theory and infrequent investment 828
      3.1.3.1. On the fragility of marginal q 828
      3.1.3.2. When does q-theory work? 830
      3.1.3.3. Taking stock 832
    3.1.4. Another detour: Several misconceptions about irreversible investment 832
    3.1.5. Adjustment hazard 835
  3.2. Aggregation 837
  3.3. Empirical evidence 838
    3.3.1. Microeconomic data 838
    3.3.2. Aggregate data 839
    3.3.3. Pent-up demand 841
  3.4. Equilibrium 842
4. Entry, exit and scrapping 844
  4.1. Competitive entry and irreversibility 844
  4.2. Technological heterogeneity and scrapping 847

* I am grateful to Andrew Abel, Steve Bergantino, Olivier Blanchard, Jason Cummins, Esther Dufflo, Eduardo Engel, Austan Goolsbee, Luigi Guiso, Kevin Hassett, Glenn Hubbard, John Leahy, Kenneth West, and Michael Woodford for many useful comments. I thank the NSF for financial support.

Handbook of Macroeconomics, Volume 1, Edited by J.B. Taylor and M. Woodford

5. Inefficient investment 848
5.1. Informational problems 849
5.2. Specificity and opportunism 851
5.2.1. Generic implications 854
5.2.2. Credit constraints 856
6. Conclusion and outlook 857
References 858

Abstract

The 1990s have witnessed a revival in economists' interest and hope of explaining aggregate and microeconomic investment behavior. New theories, better econometric procedures, and more detailed panel data sets are behind this movement. Much of the progress has occurred at the level of microeconomic theories and evidence; however, progress in aggregation and general equilibrium aspects of the investment problem also has been significant. The concept of sunk costs is at the center of modern theories. The implications of these costs for investment go well beyond the neoclassical response to the irreversible-technological friction they represent, for they can also lead to first-order inefficiencies when interacting with informational and contractual problems.

Keywords

sunk costs, irreversible investment, (S, s) models, adjustment hazard, aggregation, non-linearities, private information, incomplete contracts, pent-up investment

JEL classification: E22, D21, D23, E32

1. Introduction

Aggregate investment is an important topic. Countries and firms are often judged by their performance along this dimension, since investment is viewed as providing hope for future prosperity. It is not surprising, therefore, that much has been written about investment. It is even less surprising that many surveys, and surveys of surveys, already exist[1]. Rather than surveying the surveys of surveys, as one would expect from a handbook chapter, I have chosen to focus most of my discussion on that which is relatively new. The cost of this, of course, is that most of the theories I will discuss have not yet passed the test of time and are often only half the distance toward full development.

Most, but not all, of the subjects I plan to discuss relate directly to investment in equipment and structures. Investing means trading the present for the future; as is the case, for example, when a firm purchases equipment, builds structures, trains its workers, restructures production, spends resources on R&D, hoards labor during a recession; or when a worker leaves a job to search for another one, invests in human capital; or when a country undergoes a structural adjustment, a trade liberalization or a fiscal reform. The more theoretical sections of this survey apply to most of these examples. Further, except for specific empirical results, a large part of the discussion about equipment and structures also applies to other forms of investment.

The style of this review article is mostly empirical in early sections and mostly theoretical in later ones. This ordering is highly correlated with the chronology of research on investment. It follows that I am implicitly advocating for more empirical work on the newest theories.

The layout of the chapter is as follows. Section 2 is rather traditional in content. It describes the basic investment theory and findings, taking the view that the pre 1990s empirical literature was in disarray with respect to finding a role for the cost of capital in investment equations. During the 1990s, however, we have learned from long run relationships and "natural experiments" that the cost of capital does indeed have significant effects on investment, although it is probably not the most important explanatory variable. Neither, I should add, is measured q.

Section 3 describes what has been well known but largely ignored until recently: that microeconomic investment is lumpy and mostly sunk. It turns out that changes in the degree of coordination of lumpy actions play an important role in shaping the dynamic behavior of aggregate investment. The old concept of pent-up demand is back. This section contains a more detailed description of models and techniques than the others. It also attempts to clarify several misconceptions about the implications of these models.

Section 4 is about equilibrium interactions and scrapping. It describes the consequences of free entry and different assumptions about the elasticity of the supply

[1] See, for example, Chirinko (1993), Hassett and Hubbard (1996b), for excellent surveys of traditional investment equations.

of capital for equilibrium investment and scrapping. Vintage and putty-clay models are briefly mentioned as a natural environment in which to address the economic obsolescence issue. This concept is particularly relevant for understanding capital accumulation during episodes of rapid growth and after substantial shocks to the price of intermediate inputs.

Section 5 discusses inefficient investment. The first part of the section deals with informational problems. Discontinuous action due to irreversibility and fixed costs are compounded by the presence of private information and create a powerful drag on investment. Inaction is a natural information trap, small information flows lead naturally to further inaction, and the feedback process goes on. Aggregate investment will appear too sluggish given the ex-post information of an econometrician, and it will probably be too slow in responding to new conditions relative to first and second best scenarios.

The second part of this section describes how the sunk nature of investment, when combined with contractual incompleteness, can lead to underinvestment and, through general equilibrium, to a series of distortions in the scrapping margin and in the response of investment to aggregate shocks. Financial constraints are discussed within this context. The concept of rationing, the effects of underinvestment on complementary factors (and vice versa) and the relation between excessive capital/labor substitution and investment are also part of this section. The issue of property rights and investment also fits very naturally here.

Section 6 concludes.

2. Basic investment theory and findings

2.1. Pre 1990s: Dismay

Since very early on, economists have attempted to explain investment behavior using both scale and relative price variables, and since very early on, the former have been more successful than the latter.

One of the first "theories" of investment was the *accelerator model* [Clark (1917)]. Scarcely a theory, the accelerator model is derived by inverting a simple fixed proportion production function and taking first differences. Unable to account for the serial correlation of investment beyond that of output growth, this model was soon transformed into the *flexible accelerator* model [Clark (1944), Koyck (1954)]:

$$I_t = \sum_{\tau=0}^{n} \beta_\tau \Delta K^*_{t-\tau}, \tag{2.1}$$

where I denotes investment, the β_τ's are distributed lag parameters, and K^* is the desired, as opposed to actual, level of capital. In the simple fixed proportions world, K^* can be written as a linear function of the output level, Y:

$$K^* = \alpha Y.$$

where α is a parameter.

The absence of prices (the cost of capital, in particular) from the right-hand side of the flexible accelerator equation has earned it disrespect despite its empirical success. Jorgenson's (1963) *neoclassical theory of investment* intended to remedy this situation. Starting from the optimization problem of a perfectly competitive firm facing no adjustment costs, myopic expectations, and constant returns Cobb–Douglas technology, Jorgenson obtained the standard static first-order condition:

$$K = \alpha Y/C_k,$$

where C_k stands for the cost of capital and α is now the share of capital in a simple Cobb–Douglas production function. As with the accelerator model, this model was unable to account for the serial correlation of investment, and so gave way to the *flexible neoclassical model* of Hall and Jorgenson (1967), where

$$K^* = \alpha Y/C_k, \tag{2.2}$$

was now used in Equation (2.1).

Soon it was shown, however, that by constraining the coefficient of the cost of capital to be the same as the coefficient of output, this model imposed rather than found a role for the cost of capital in the investment function. Eisner (1969) estimated a modified Hall and Jorgenson model which allowed for different coefficients on output and the cost of capital and found no independent role for the cost of capital.

The cost of capital's rise and fall from grace was not an unknown experience, however. Several decades before, authors such as Tinbergen (1939) and Meyer and Kuh (1957) had pointed out the dominance of liquidity variables over interest rates for short run investment[2].

None of these are full theories of investment, rather they are theories conditional on the level of output[3]. The famous *q-theory* of Tobin (1969) and Brainard and Tobin (1968) went one step further. They argued that investment should be an increasing function of the ratio of the value of the firm to the cost of purchasing the firm's equipment and structures in their respective markets. This ratio, known as *average q*[4], summarizes most information about future actions and shocks that are of relevance for investment[5]. Indeed, average q would later be shown to be a sufficient statistic for investment in a wide variety of scenarios. Thus, the new canonical investment equation became

$$I = \gamma q,$$

where γ is a strictly positive parameter.

[2] See Chirinko (1993) for a more thorough review of the history of the debate over the role of the cost of capital, profits and output in investment decisions.

[3] Things are even worse for the basic frictionless neoclassical model; it is ill defined as a full model because firm level output is not determined under constant returns and perfect competition.

[4] It is also often referred to as Tobin's q.

[5] This includes the optimal path of output.

The elegant theoretical contributions of Abel (1979) and Hayashi (1982) connected the existing theories and partial theories. They showed that the neoclassical model with convex adjustment costs yields a q-model. This q, known as *marginal* q, should be interpreted as the marginal value of an installed unit of capital, which corresponds to the shadow value of a unit of capital in the firm's optimization problem. Further, Hayashi showed that (for price-taking firms) when the production function and adjustment cost function are linearly homogeneous in capital and labor, marginal and average q are equal. This is an important result from an empirical standpoint because marginal q is unobservable to the econometrician whereas average q is, in principle, observable to the econometrician[6].

Soon, however, the q-model, along with expanded and ad-hoc "flexible-q" models (i.e. with additional lags of q on the right-hand side), joined models based on the cost of capital in their lack of empirical success. Scale variables such as cash-flows always seemed to matter more in investment equations than q which, in principle, should have been a sufficient statistic[7].

Figure 2.1, which reproduces figures 1 and 3 of Hassett and Hubbard (1996b), helps us understand the statistical reasons for the problem. The bottom line is clear: In aggregate US data (which is probably representative of many other data sets for this purpose) the unconditional correlation between cost of capital and investment is low, and so is that between average q and investment. On the other hand, cash flows and sales's growth closely track aggregate investment.

The 1980s discontent with respect to investment equations is probably well captured in Blanchard's (1986) discussion of Shapiro's (1986) investment paper at Brookings:

> "... it is well known that to get the user cost to appear at all in the investment equation, one has to display *more than the usual amount of econometric ingenuity,* resorting most of the time to choosing *a specification that simply forces the effect to be there ...* " [my emphasis]

Today, the first emphasized statement still holds, but the second one probably does not. This takes me to the next subsection.

2.2. *"Econometrics": Cost of capital and q matter*

Econometric "ingenuity" eventually pays off, although this often means isolating that part of the relationship which conforms with the theory, rather than explaining a

[6] I have mixed feelings about this equivalence result, however. Not about its theoretical derivation, which is elegant and useful; rather about its abuse in empirical work. Too often, it is used to justify substituting average for marginal q on the right-hand side of investment equations, even though the assumptions required for the equivalence between the two are not nearly satisfied in the industry or firms studied (e.g. Compustat). This does not mean that average q should not be used, but it says that we should not pretend that the foundation for its use is beyond the basic intuition provided by Tobin (1969), and that the additional properties that hold for marginal q are to be expected from average q (e.g. sufficiency).

[7] Fazzari et al. (1988) started a large literature documenting the role of these variables, even after conditioning for average q. I will return to the interpretation of these regressions later in the survey.

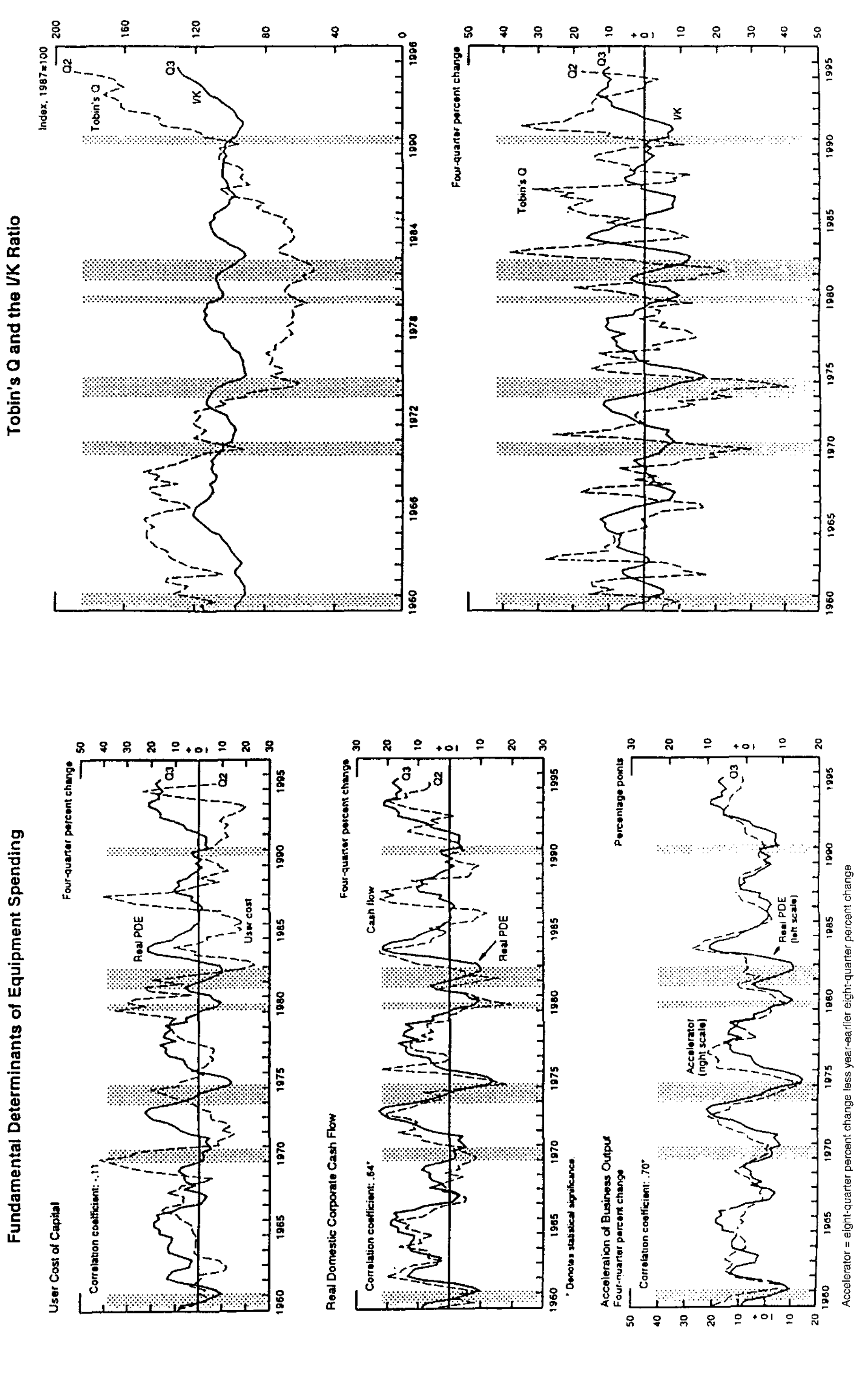

Fig. 2.1. Basic aggregate facts.

substantial fraction of the movements of the left-hand side variable, or even relating a significant fraction of the volatility of the right-hand side variables to that of the left-hand side variable. In my view, this is the type of payoff obtained from the recent incarnations of the "traditional" line. Still, it is progress.

Going from less to more ambitious, there are two generic developments I wish to discuss. First, ignoring high and medium frequency variations, we have come to the realization that the low frequency aspects of the data are not inconsistent with theories that assign an important role to the cost of capital in determining the rate of capital accumulation. Second, there are distinctive episodes during which changes in the cost of capital are sufficiently dramatic that it becomes possible to demonstrate the importance of the cost of capital at higher frequencies as well.

Other recent developments within the traditional line include the use of Euler equation procedures. In my view, and unlike the case in basic finance and consumption applications, these procedures are a form of morphine rather than a remedy: their lack of statistical power allows us to sometimes not see the problem. Since my goal is to discuss progress, I will skip results obtained with these procedures [8].

2.2.1. Long-run

Many of the problems with investment equations have to do with the presence of complex and not well understood dynamic issues (more on this in the next sections). From early on, researchers have found it useful to think about investment in two steps: first, derive some simple expression for a "target" stock of capital, which I have called K^* here; and second, model dynamics as a, possibly complex, function of contemporaneous and lagged changes in K^*. It seems sensible, therefore, to start by asking whether the first step resembles what we expect before going into the difficult issues of timing.

Taking logs on each side of Equation (2.2), disregarding constants, and relaxing the unit elasticity constraint on the cost of capital, yields

$$k^* - y = \gamma c_k, \tag{2.3}$$

where lower case letters denote logarithms and γ is the parameter of interest.

This expression cannot be estimated, of course, because k^* is not observed. There is a simple argument based on cointegration, or a close small sample "cousin," which allows us to get around this observability problem, however. The whole purpose of deriving k^* is to then model k as trying to keep pace with it. Thus, differences between these two variables should only be transitory (up to constants). If k^* and k are

[8] See Oliner, Rudebusch and Sichel (1995) for a damaging evaluation of the statistical properties of these procedures.

sufficiently volatile (ideally with unit roots, in large samples), then we can "ignore" the discrepancy between these two variables in estimating γ. Let

$$k = k^* + \epsilon,$$

with ϵ a stationary residual that captures transitory discrepancies between the two variables due to adjustment costs. Substituting this expression into Equation (2.3) yields an equation that can be estimated:

$$k - y = \gamma c_k + \epsilon. \tag{2.4}$$

Estimating this equation by OLS (the simplest of the cointegration procedures) yields, for aggregate US data, an estimate of γ of –0.4; significantly different from zero [9].

We can do better, however. In any small sample, the cointegration argument will not take its full bite, and the estimates of γ will be affected by the correlation between regressors and ϵ. Caballero (1994a) argues that this is particularly serious and systematic in models with slow adjustment (e.g. due to adjustment costs). The intuition behind this idea is simple. A partial adjustment mechanism implies that, in any finite sample, the variance of K/Y ought to be less than the variance of K^*/Y, which means the left-hand side of Equation (2.4) ought to be less volatile than the right-hand side of Equation (2.3), or γc_k [10]. However, by the normal equations of OLS, the estimated counterparts of γc_k and ϵ on the right-hand side of Equation (2.4) must be orthogonal, so that the variance of $k - y$ is greater than the variance of $\hat{\gamma} c_k$, which is equal to the variance of the estimated $\widehat{k^*} - y$. Since this inequality is in contradiction with what is implied by adjustment cost mechanisms, we conclude that the estimate of γ is biased toward zero.

Using Monte Carlo simulations, I showed in that paper that this bias can be substantial, and then proceeded to correct it using Stock and Watson's (1993) procedure. I obtained an estimate of γ close to minus one, very near the neoclassical benchmark [11].

2.2.2. Short-run

Demonstrating a relationship between capital accumulation and the cost of capital at higher frequencies has required two changes in approach: first, a change in emphasis from aggregate to microeconomic data; and second, the use of natural

[9] This estimate of γ was obtained using US quarterly NIPA data for the period 1957:1–1987:4. Capital corresponds to equipment capital and cost of capital is constructed as in Auerbach and Hassett (1992).

[10] Note that if adjustment costs are non-convex it is possible, at the microeconomic level, and in a sufficiently short sample, to have these relative volatilities reversed. This is not an issue for the aggregate data results discussed here. See the next section for more on non-convex adjustment cost models.

[11] Similar estimates were obtained by Bertola and Caballero (1994) and Caballero, Engel and Haltiwanger (1995) with different data sets.

experiments, such as periods of tax reform, which present the econometrician with more accurate measures of (often substantial) changes in the cost of capital and q. Measures of q, for example, are not only very noisy because of the substitution of average q for marginal q, but also because there may be substantial "non-fundamental" movements in the value of firms, making average q mismeasured as well. However, there are certain episodes (e.g., periods of tax reform) when the movements in q are likely to be large, in a predictable direction, and for the "right reasons". As with cointegration, during those episodes problems with the residual can be more or less disregarded.

The movement from aggregate to microeconomic data, by itself, has not done much to improve affairs. Although microeconomic data has improved precision, coefficients on the cost of capital and q in investment equations have remained embarrassingly small. Combined with the use of natural experiments, however, emphasis on microeconomic data has had much higher payoffs. The work of Cummins, Hassett and Hubbard (1994, 1996a) is salient in this regard. They isolate periods with important tax reforms and find that the coefficient on q is much larger in those episodes. Most recently, using firm level data for 14 developed countries, they find that while using standard instrumental variable procedures yields coefficients on q which range from 0.03 to 0.1, when contemporaneous tax reforms are included among the instruments, the estimates jump to a range between 0.09 to 0.8, with median and mean not very far from 0.5. In the USA, for example, the estimate of the coefficient on q jumps from 0.048 to 0.65 [12].

Although these empirical results represent significant progress, there is still plenty of work needed to retrace the steps back to the aggregate and we must not forget that a substantial component of the variation in aggregate and microeconomic investment remains unexplained. The next sections describe progress on both fronts.

3. Lumpy and irreversible investment

Investment is a flow variable, and as such it is very sensitive to obstacles. Investment is the by-product of the process by which the capital stock catches up with its desired level; but there are many paths leading the former to the latter. In this section I begin discussing some of these obstacles, emphasizing those that have had prominence in the recent literature.

3.1. Plant/firm level

The most basic form of friction occurs at the level of microeconomic units, and goes under the general heading of adjustment costs.

[12] See Caballero (1994b) for a discussion of their results, interpretations and procedures.

3.1.1. Microeconomic adjustment: characterization

There are essentially three basic types of adjustments observed at the establishment level: (a) ongoing frictionless flow (maintenance); (b) gradual adjustments (e.g. refinements and training dependent improvements); (c) major and infrequent adjustments [13].

The structural literature of the 1980s and before, based explicitly or implicitly on convex adjustment cost models (the quadratic adjustment cost model, in particular) dealt with (a) and (b). The implicit "hope" was that the smoothness brought about by aggregation would make disregarding the importance of infrequent adjustments for individual units, unimportant for aggregate phenomena. Instead, the idea was to derive aggregate investment equations as coming from the solution to the optimization problem of a fictitious agent facing adjustment costs which only led to smooth adjustments of type (a) and (b). Many authors disagreed with this strategy [e.g. Rothschild (1971)]; but for most the relative simplicity of the quadratic model was too enticing to resist.

A combination of factors eventually led economists to revisit and reevaluate some of the shortcuts which were in widespread use by the end of the 1980s [14]. First, there was frustration with the disappointing empirical results described above. Second, techniques which could handle models of lumpy investment became part of the modern economist's tool kit. And third, microeconomic data made the obvious even more apparent: microeconomic investment is extremely lumpy, and this lumpiness is unlikely to fully "wash out" at the aggregate level.

The work of Doms and Dunne (1993) was instrumental in stressing the last point. They documented investment patterns of 12 000 plants in US manufacturing over the 17 year period, 1972–1989. Their findings are many, of which I have chosen to emphasize those that are most closely related to the purpose of this survey.

For each establishment, Doms and Dunne constructed a series of the proportion of the total equipment investment of the establishment (over the 17-year period) made in each year. They found that on average the largest investment episode accounts for more than 25 percent of the 17 year investment of an establishment and that more than half of the establishments exhibited capital growth close to 50 percent in a single year. They also note that the second largest investment spike often came next to the largest investment spike (right before or right after) suggesting that both spikes correspond to a single investment episode. [15] Combining the two primary spikes, they find that nearly

[13] Which may, in turn, have a time to build aspect.

[14] Chronologies are never exact, of course. For example, Nickell had already discussed irreversible investment and many of its implications in 1978; but the *mode* did not move until much later.

[15] An investment project may not be fully counted within one year since not all projects start on January 1, and certainly may take more than a few days to implement. One should not confuse "time to build" with the standard convex adjustment costs. Time to build is the optimal scheduling of a *given lumpy project*, while in the standard convex adjustment costs model the firm changes this project continuously and smoothly [see Caballero and Leahy (1996)].

40 percent of the sample investment of the median establishment probably corresponds to a single investment episode [16]. Moreover, this is likely to be a lower bound on the lumpiness of investment since these numbers correspond to establishments that remained in the sample during the entire 17-year period. Adding entry and exit would undoubtedly make the evidence on microeconomic lumpiness even more apparent.

As for evidence on the macroeconomic relevance of microeconomic lumpiness, Doms and Dunne offer several hints. First, using data on about 360 000 establishments for Census years 1977 and 1987, they document that about 18 percent of aggregate investment is accounted for by the top 100 projects. As a metric, only 6 percent of employment is in the top 100 employers, and less than 10 percent of production occurs in the top 100 producers. More importantly, they show that the time series correlation between aggregate investment and a Herfindahl index of microeconomic investments is very high (close to 0.45). They also constructed a series with the number of firms undergoing their primary investment spike during each year. They show that this measure, rather than the average size of these spikes, closely tracked aggregate investment [17].

The subsections that follow describe models which are broadly consistent with these findings, and reviews structural evidence based on these models which lends further support to the view that microeconomic lumpiness is very important for aggregate investment dynamics.

3.1.2. "Representative" problem

There is by now a vast literature (and surveys of it) describing microeconomic models able to capture the essence of the lumpy and discontinuous adjustment highlighted by the evidence described above. Rather than giving a thorough presentation of the canonical model, I refer the interested reader to one of these surveys [18]. Instead, I will only sketch the problem, mostly to characterize the nature of the solution and to develop notation which will prove useful later.

Let actions and realizations of shocks evolve in discrete time, with time intervals, Δt. Having optimized over all inputs but capital during the period, a firm with stock of capital K and facing conditions θ, has a flow of profits net of rental cost of capital:

$$\Pi(K,\theta)\Delta t = \left(K^{\gamma}\theta - rK\right)\Delta t \qquad 0 < \gamma < 1, \tag{3.1}$$

where K is the firm's stock of capital; θ is a profitability index that combines demand, productivity and wage shocks; r is the discount rate; and γ represents the elasticity

[16] Cooper, Haltiwanger and Power (1994) go one step further in characterizing infrequent lumpiness. Using a data set similar to that of Doms and Dunne, they show that the probability of a firm experiencing a major investment spike is increasing in the time since the last major spike. This feature of the data is highly consistent with the implications of the models reviewed later in this section.

[17] Where aggregate investment corresponds to the investment of all the establishments in their sample.

[18] Dixit (1993) provides an excellent discussion of the basic problem and the mathematical techniques needed to solve it.

of gross profits with respect to capital. It is less than one as long as the firm exhibits some degree of decreasing returns or market power, which I assume to be the case. For convenience, capital does not depreciate.

It will also facilitate things to assume that increments in the logarithm of θ are i.i.d., and that time and the sample paths of θ are "almost" continuous (i.e. Δt is small and changes in the value of θ over an interval of time Δt are small). I make these assumptions so I can, informally, use all the convenience of Ito's lemma and Brownian motions. I choose to depart from strict continuous time, on the other hand, because discrete time will allow me to present this section in a more unified manner.

As in the previous section, we can find an expression for the static optimum of the stock of capital, or "desired" capital:

$$K^* = \operatorname{argmax}_K \Pi(K, \theta) = \left(\frac{\gamma\theta}{r}\right)^{\frac{1}{1-\gamma}}. \tag{3.2}$$

It is apparent from this expression that K^* inherits the stochastic properties of θ, so it also follows a geometric random walk. Moreover, the measure of capital "imbalance":

$$Z \equiv \frac{K}{K^*},$$

also inherits the geometric random walk process, for any given K. Substituting this expression into Equation (3.1) and using Equation (3.2) to solve for θ yields

$$\Pi(Z, K^*) = \frac{r}{\gamma}\left(Z^\gamma - \gamma Z\right) K^* \qquad 0 < \gamma < 1. \tag{3.3}$$

In order to generate infrequent actions, the cost of adjusting the stock of capital must increase sharply around the point of no adjustment. A cost proportional to the size of adjustment is enough to do so. Lumpiness requires a little more, for there must be an advantage in bunching adjustment; increasing returns in the adjustment technology is the standard recipe, of which a fixed cost is the simplest. Let $C(\eta, K^*)$ denote the cost of adjusting the capital stock by $K^*\eta$:

$$C(\eta, K^*) = \{\Delta K \neq 0\} K^* \begin{cases} c_f + c_p^+\eta & \text{if} \quad \eta > 0, \\ c_f - c_p^-\eta & \text{if} \quad \eta < 0, \end{cases} \tag{3.4}$$

where the K^* term ensures that the relative importance of adjustment costs remains unchanged over time [19]. Figure 3.1 illustrates an example of $C(.,.)/K^*$.

[19] This goal would also be accomplished by K, but K^* yields slightly simpler mathematical expressions at a low cost in terms of substantive issues.

Also, I have allowed proportional costs to differ with respect to upward and downward adjustments in order to talk later about the irreversible investment case; for this purpose, I could have done it equally well through asymmetric fixed costs. Allowing for both forms of asymmetries simultaneously is a trivial but uninteresting extension.

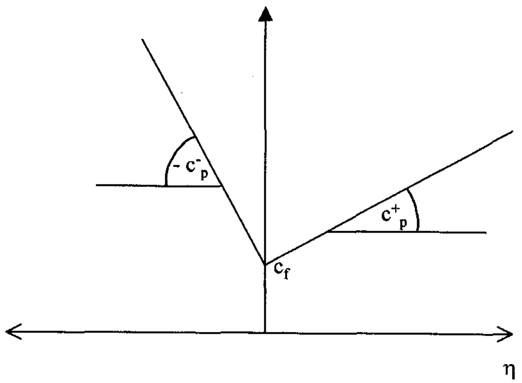

Fig. 3.1. Adjustment costs.

The problem of the firm can be characterized in terms of two functions of Z and K^*: $V(Z,K^*)$ and $\tilde{V}(Z,K^*)$. The function $V(Z,K^*)$ represents the value of a firm with imbalance Z and desired capital K^* if it does not adjust in this period, and $\tilde{V}(Z,K^*)$ is the value of the firm which can choose whether or not to adjust. Thus,

$$V(Z_t,K_t^*) = \Pi(Z_t,K_t^*)\Delta t + (1 - r\Delta t)\,\mathrm{E}_t\left[\tilde{V}(Z_{t+\Delta t},K_{t+\Delta t}^*)\right], \tag{3.5}$$

and:

$$\tilde{V}(Z_t,K_t^*) = \max\left\{ V(Z_t,K_t^*), \max_{\eta}\left\{V(Z_t+\eta,K_t^*) - C(\eta,K_t^*)\right\}\right\}. \tag{3.6}$$

The nature of the solution of this problem is now intuitive. Given the function $V(Z,K^*)$, Equation (3.6) provides most of what is needed to characterize the solution. First, since C is positive even for small adjustments, it is apparent that when Z is near that value for which $V(Z,K^*)$ is maximized, the first term on the right-hand side of Equation (3.6) is larger than the second term; that is, there is a range of inaction. Second, since both adjustment costs and the profit function are homogeneous of degree one with respect to K^*, so are V and $\tilde{V}$. Thus, it is possible to fully characterize the solution in the space of imbalances, Z. Among other things, this implies that the range of inaction described before, is fixed in the space of Z. Let L denote the minimum value of Z for which there is no investment, and U the maximum value for which there is no disinvestment; thus the range of inaction is (L,U). Third, conditional on adjustment, changes must not only be large enough to justify incurring the fixed cost, but also the (invariant) target points must satisfy

$$V_Z(l) = c_p^+ \tag{3.7}$$

and

$$V_Z(u) = -c_p^-, \tag{3.8}$$

where V_Z is the derivative of V with respect to Z, while l and u denote the target points from the left and right of the inaction range, respectively. These first-order

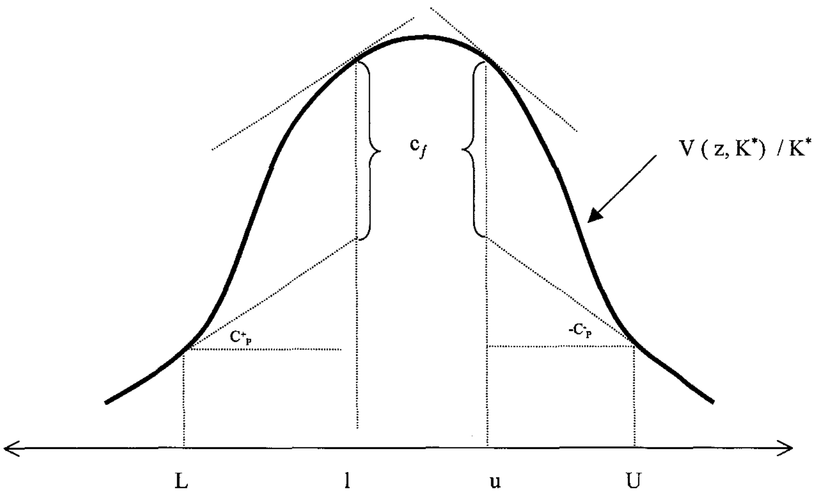

Fig. 3.2. Value function.

conditions are known as "smooth pasting conditions," and simply say that, conditional on adjustment taking place, it must cease when the value of an extra unit of investment (or disinvestment) is equal to the additional cost incurred by that action.

There are two additional smooth pasting conditions:

$$V_Z(L) = c_p^+ \tag{3.9}$$

and

$$V_Z(U) = -c_p^-, \tag{3.10}$$

which ensure no expected advantage from delaying or advancing adjustment by one Δt around the trigger points.

These smooth pasting conditions are enough to find the optimal (L, l, u, U) rule, given the value function. In order to find the latter, however, we need to go back to Equation (3.5). Standard steps reduce this equation, in the interior of the inaction range, to a second-order differential equation. The two boundary conditions required to find V are obtained from equalizing the two terms on the right-hand side of Equation (3.6):

$$V(L, K^*) = V(l, K^*) - \left(c_f + c_p^+(l - L)\right) K^*, \tag{3.11}$$

$$V(U, K^*) = V(u, K^*) - \left(c_f + c_p^-(U - u)\right) K^*, \tag{3.12}$$

which simply say that since the investment rule (optimal or not!) dictates that once a trigger point is reached, adjustment must occur at once, the only difference in the value of being at trigger and target points must be the adjustment cost of moving from the former to the latter.

Figure 3.2 illustrates the value function. Smooth pasting says that the tangents at L and l have slope c_p^+, while those at U and u have slope $-c_p^-$. Value matching says that the value function evaluated at the target minus the value function evaluated at the trigger point is equal to the variable cost paid at adjustment plus the fixed cost (all these normalized by K^* in the figure).

There are a few particular cases which are worth highlighting because they appear often in the literature:

(1) If there is no variable cost of investment, once the adjustment decision has been taken, adjustment from both sides is complete since the marginal cost of adjustment is zero. Thus, the (L, l, u, U) rule reduces to an (L, c, U) rule, where c is the common target for investment as well as disinvestment, and is that value which maximizes $V(Z, K^*)$ for any K^* [20].

(2) If there are variable costs but no fixed cost, there is no reason for adjustment to be lumpy, for there are no increasing returns in the adjustment technology [21]. Once the boundaries of the inaction range, L and U, are reached, the firm adjusts just enough to avoid crossing outside the inaction range; that is L and U become reflecting barriers.

(3) If there is a large (not necessarily infinite) cost to disinvestment, then investment becomes irreversible. In the absence of investment costs, the investment rule reduces to a single reflecting barrier L, which is to the left of one (reluctance to invest). This is the standard irreversible investment case.

3.1.3. A detour: q-theory and infrequent investment

One of the main manifestations of the empirical failure of previous investment theories, has been the difficulty in finding either a significant and sizable role for q, or evidence that it is a sufficient statistic for investment.

Do the theories studied in this section help explain these empirical failures? I see two reasons to believe so. The first one is rather negative. q-theory is no longer robust in our setting, so there are many scenarios where we should not expect it to work. The second one is more positive. In the subclass of models where it does work, the functional form relating q and investment is likely to be highly non-linear, thus quite different from the standard linear regressions leading to the rejection of q-theory.

3.1.3.1. On the fragility of marginal q. It is apparent from the lack of global concavity of the value function in Figure 3.2, that traditional q-theory is not likely to work in the presence of jumps. Caballero and Leahy (1996) develop the argument in detail, which I summarize below.

[20] Note that in general $c \neq 1$. That is, the optimal dynamic target is generally different from the static one.

[21] And we have already assumed that shocks are "small" in any given Δt.

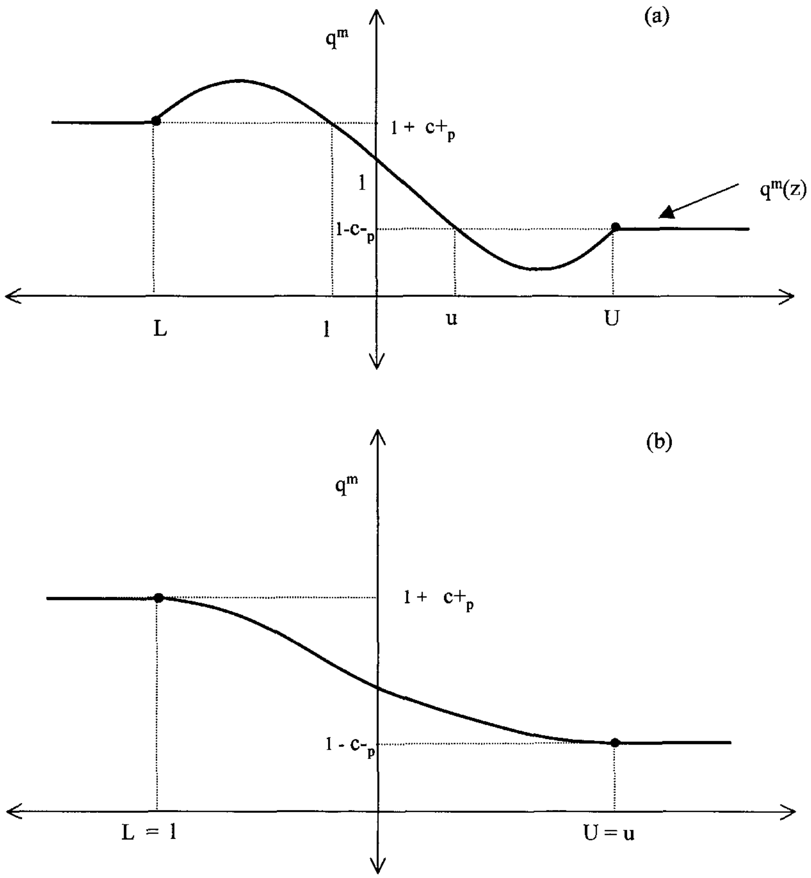

Fig. 3.3. Marginal q.

The value of the firm is equal to $K + \tilde{V}$, thus marginal q is[22]

$$q^M(Z) = 1 + \tilde{V}_K = 1 + \frac{\tilde{V}_Z}{K^*}. \tag{3.13}$$

Figure 3.3a plots q^M against the imbalance measure Z[23]. Smooth pasting implies that q^M must be the same at trigger and target points (because $\tilde{V}_Z$ must be the same at trigger and target points); if there are jumps, these are points very far apart in state

[22] Recall that $\tilde{V}$ was defined as the present value of profits net of adjustment costs *and* interest payments on capital.

[23] See, for example, Dixit (1993) for a characterization of the (L, l, u, U) solution in terms of a similar diagram.

space. Two points with the same value of q^M lead to very different levels of investment (zero and large). Moreover, since the value function becomes linear outside the inaction range, all points outside the inaction range (on the same side) have the same q^M, and all of them lead to different levels of investment. It is apparent, therefore, that the function mapping q^M into investment no longer exists. Worse, in between trigger and target points, the relation between q^M and Z is not even monotonic.

What is happening? Marginal q is the expected present value of the marginal profitability of capital. Far from an adjustment point, it behaves as usual with respect to the state of the firm: if conditions improve, future marginal profitability of capital rises, and so does q^M. Close to the investment point, on the other hand, the effect of a change in the state of the firm over the probability of a large amount of investment in the near future dominates. An abrupt increase in the stock of capital brings about an abrupt decline in the marginal profitability of capital as long as the profit function is concave with respect to capital[24]. Thus, an improvement in the state of the firm makes it very likely that it invests in the near future, reducing the expected marginal profitability of capital in the near future, thus lowering the value of an extra unit of installed capital.

Caballero and Leahy (1996) show that adding a convex adjustment cost to the problem does not change the basic intuition of the mechanism described above. They also show, somewhat paradoxically, that average q, which is often thought of as a convenient albeit inappropriate proxy for marginal q, turns out to be a good predictor of investment even in the presence of fixed costs, although it is no longer a sufficient statistic, except for very special assumptions about the stochastic nature of driving forces.

3.1.3.2. When does q-theory work?. The failure of q-theory described above is rooted in the presence of increasing returns in the adjustment cost function (3.4). This feature of the adjustment technology is responsible for the loss of global concavity of the value function, which is behind the non-monotonicity of marginal q[25].

Monotonicity of q^M inside the inaction range is recovered by dropping the fixed cost from Equation (3.4), as was done in cases 2 and 3 in Section 3.1.2. Figure 3.3b portrays this scenario. Adjustment at the trigger points no longer involves large projects, thus proximity to these triggers no longer signal the sharp changes in future marginal profitability of capital which were responsible for the "anomalous" behavior of q^M[26]. There is still the issue that in the (very) rare event that a firm finds itself outside the inaction range it will adjust immediately to the trigger, at a constant marginal cost, so

[24] Which I take to be the standard case.

[25] Indeed, value functions for (S,s) models are often only K-concave.

[26] Of course, once at the trigger, large projects may result from the accumulated and – more or less – continuous response to a sequence of shocks with the same sign. But this does not give rise to a sharp change in profitability since investment occurs only in response and to offset new, as opposed to accumulated, changes in profitability.

different levels of investment are consistent with the same value of q^M. This is easily remedied by adding a convex component to the adjustment cost function[27]:

$$C(\eta, K^*) = \{\Delta K \neq 0\} K^* \left\{ c_p |\eta| + c_q |\eta|^{\beta} \right\} \qquad \beta > 1. \tag{3.14}$$

This is essentially what Abel and Eberly (1994) do[28]. Absent the advantage of lumping adjustment brought about by the presence of fixed costs, standard q-theory is recovered whenever the firm invests. Provided adjustment takes place, the firm equalizes the marginal benefit of adjustment and the marginal cost of investing, which is now an increasing function of adjustment:

$$q^M = 1 + \operatorname{sgn}(\eta) \left(c_p + \beta c_q |\eta|^{\beta - 1} \right),$$

for $\eta \neq 0$. By setting η to zero, we can obtain the boundaries of inaction in q^M-space. Indeed, investment will not occur if

$$1 - c_p < q^M < 1 + c_p.$$

Abel and Eberly (1994) go further, and show that their insight is robust to the presence of *flow*-fixed costs. That is, fixed costs which are multiplied by Δt; if adjustment occurred instantaneously, the firm effectively would pay no fixed cost. Because of the convex adjustment component, the firm *chooses* not to adjust instantaneously and pays the fixed costs instead. In a sense, the endogenous adjustment decisions and the fact that the fixed cost goes to zero as adjustment speeds up, ensures that the fixed cost remains relatively "small," and so do investment projects[29]. It is important to realize that their paper "unifies" q-theory with irreversible investment and regulation (i.e. infrequent but infinitesimal adjustments) problems, but it does not unify it with the standard (S, s) literature on lumpy adjustment, which is, unfortunately, the way many have interpreted their results.

Barnett and Sakellaris (1995) study a panel of US firms searching for evidence on a reduced sensitivity of investment to changes in q when the latter is close to one (the "inaction" range). They find the opposite; in their panel, a firm's investment seems to be more rather than less responsive to q when q is close to one. Abel and Eberly (1996a), however, show that allowing for unobserved heterogeneity in the inaction

[27] Which, at the same time, makes transitions outside the inaction range less rare.

[28] Needless to say, it is trivial to add asymmetries to the adjustment cost function. But that is beside the point of this section.

[29] Alternatively, if one assumes perfect competition and constant returns to scale, the profit function becomes linear with respect to capital (if the other factors of production can be adjusted at will), so changes in investment do not feed back into q. In this extreme case, the modified (i.e. with an inaction range) q-theory works well even in the presence of traditional fixed costs.

range relevant for different types of investments could explain the negative Barnett–Sakellaris finding.

3.1.3.3. Taking stock. One may be inclined to conclude from this section that before going ahead with q-theory one should check whether investment literally exhibits jumps or not. This is not the lesson I draw, however.

For once, this is not right. It is not difficult to add a time to build mechanism such that a lumpy project is decomposed into a fairly smooth flow, without altering the argument of why marginal q fails in the presence of fixed costs. But more importantly, I suspect the main lesson is one of modesty. I doubt that researchers will often find the required data and/or patience to determine whether one scenario or the other holds. In this case, we might as well acknowledge that the relationship between marginal q and investment is not robust, and that average q is unlikely to be a sufficient statistic for investment. Of course it is important to include variables that capture knowledge of the future on the right-hand side of investment equations, but we should avoid reading "too much" from these regressions.

3.1.4. Another detour: Several misconceptions about irreversible investment

As I mentioned before, when describing the special case of irreversible investment, the regulation barrier, L, is to the left of one. That is, investment occurs only when the stock of capital is substantially below the frictionless stock of capital. Alternatively, investment occurs when the marginal profitability of capital is substantially above the cost of capital. This is the famous "reluctance to invest" result.

There are several misconceptions about the implications of this "reluctance" result. I will mention three of them. It is often said that, (a) reluctance implies that, in the presence of irreversibility, the firm accumulates less capital; (b) since reluctance rises with uncertainty (the regulation point moves further to the left), more uncertainty implies less capital; and (c) standard present value techniques are inappropriate because reluctance reflects the value of the "option to wait" for more information before irreversibly sinking resources and this is not taken into account by the standard formulae.

In order to show the fallacious nature of the first statement, it is useful to go back to our canonical problem and simulate the path of the (log of) stock of capital of a firm facing no irreversibility constraint. Panel (a) in Figure 3.4 does so for a random realization of the path of θ. Panel (b) in the figure shows the corresponding path of the marginal profitability of capital, which is equal to the constant – frictionless – cost of capital, r[30].

Imagine now imposing an irreversibility constraint on the firm, but assume that the firm does not modify its "frictionless" investment rule whenever it can invest. This is

[30] These figures are from Caballero (1993a).

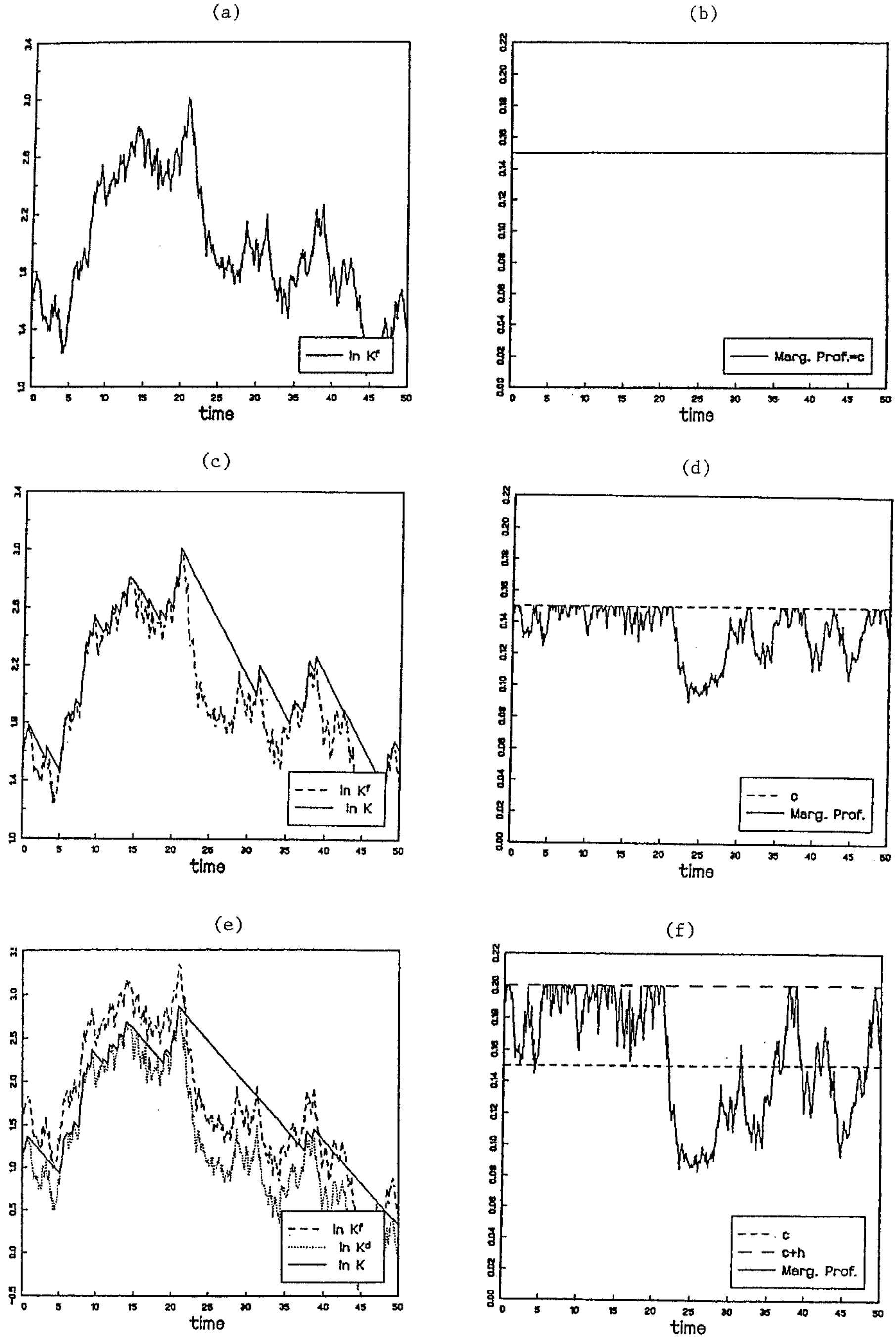

Fig. 3.4. Reluctance and its counterpart.

portrayed in panel (c) of the figure. The solid and dashed lines represent the actual and frictionless stocks of capital, respectively. It is apparent that the firm would, on average, have *too much* capital, for it would have the same stock of capital in good times, but too much in bad times. The counterpart of this is in panel (d), which shows that on average the marginal profitability of capital is below the cost of capital.

Reluctance to invest in good times is an optimal response attempting to offset the natural tendency to *over*-accumulate capital induced by the irreversibility constraint. Panel (e) illustrates this point. The solid and dashed lines represent the same variables as in panel (c), while the dotted line illustrates the target stock of capital when the firm behaves optimally. The counterpart of the negative value of $ln(K^d/K^f)$ is a positive constant h in the marginal profitability of capital required for investment to take place (panel f). It is apparent that whether the stock of capital is on average higher or lower than without the irreversibility constraint is unclear; the firm has too little capital during good times but too much during very bad times. A precise answer depends on things about which we know little, and which may turn out to yield only second-order effects[31].

It is now easy to see the fallacious nature of the second statement. More uncertainty raises reluctance precisely because it raises the need to reduce the extent of excessive capital during the now deeper recessions. Without raising reluctance, an increase in uncertainty would raise the average stock of capital in the presence of irreversibility constraints. This occurs because there would now be greater capital accumulation during extremely good times which would not be offset by large disinvestment during extremely bad times[32].

The third misunderstanding is of a different nature. In my view, it is the result of insightful but, unfortunately, abused language. First, what is right: there is nothing mysterious about irreversibility constraints as a mathematical problem. Dynamic programing works, in the same way it does with other, more traditional, adjustment frictions. This means that present value formulae, using the *correct* calculation of future marginal profitability of capital also work. Of course such calculations must be performed along the optimal investment path, constraints included! What is wrong: the standard analysis must be modified to consider the value of the "option to wait".

[31] See Bertola (1992), Caballero (1993a), Bertola and Caballero (1994) for early discussions of this issue and of the related uncertainty-investment misconception. More recently, Abel and Eberly (1996b) have formalized these claims and made them more precise.

[32] This does not mean that one cannot construct scenarios where an increase in uncertainty reduces investment. For example, if there is an increase in perceived future uncertainty, the investment threshold may jump today – i.e., before the variance of shocks does – resulting in an unambiguous decline in investment.

Also, one should not confuse changes in uncertainty with changes in the probability of a bad event. The latter links increases in uncertainty to a reduction in expected value, an entirely different and more straightforward effect on investment. One can find traces of this confusion in the (informal) credibility literature.

As we have seen, there is no need to do so. However, one may choose to follow an alternative path, in which one starts by evaluating the future marginal profitability of capital without considering the effect of future optimal investment decisions on marginal profitability. This "mistake" can then be "corrected" with a term that has an option representation. This alternative way of doing things is akin to the arbitrage approach in finance, and it was nicely portrayed in Pindyck (1988). The confusion arises, in my view, from mixing the language in the two approaches[33].

A related claim exists for a once and for all project (as opposed to incremental investment). It is said that the simple positive net present value rule used in business schools to decide whether a project should be implemented does not hold because it does not consider the option to wait and decide tomorrow, when more information is available. Since I have never taught at a business school I cannot argue directly against that claim. However, if the issue is whether to invest today or tomorrow, the right criterion has never been invest if NPV is positive – at least that is what we teach economics undergraduates. This is a case of mutually exclusive projects, thus the right criterion has always been to compare their net present value and take that with the highest NPV, provided it is positive. If investment is irreversible, the project *invest tomorrow* has a lower bound at zero (because investment will not occur if NPV looks negative tomorrow), which the project *invest today* does not. Thus, other things equal, irreversibility necessarily makes investing tomorrow more attractive than investing today.

3.1.5. Adjustment hazard

At a qualitative level, the (L, l, u, U) models described above capture well the nonlinear nature of microeconomic adjustment. Maintenance expenditures aside, investment is mostly sporadic and often lumpy; scarcely reacting to small changes in the environment but abruptly undoing accumulated imbalances when they become sufficiently large, and with possibly significant asymmetries between investment and disinvestment.

At an empirical level, however, these characterizations are too stark. For reasons, some of which we understand and most of which we do not, firms respond differently to similar imbalances over time and across firms. Caballero and Engel (1999) propose a probabilistic instead of a deterministic adjustment rule. Rather than having a clear demarcation between regions of adjustment and inaction, they model a situation where large imbalances are more likely to trigger adjustment than small ones[34].

[33] See Bertola (1988) for one of the first discussions of this issue in the economics literature. There is also a related discussion in applied mathematics; see, for example, El Karoui and Karatzas (1991). Abel et al. (1996) have recently revisited and expanded the discussion on the relation between the two approaches.

[34] Another advantage of this approach is that it nests linear models as the probability of adjustment becomes independent of Z.

There are many formal motivations for such an assumption. A particularly simple one, pursued by Caballero and Engel (1999), is to assume that c_f in the adjustment cost function (3.4) is an i.i.d. random variable, both across firms and time. Although technically more complex, the nature of the problem is not too different from that of the simpler (L, l, u, U) model. Let ω denote the random fixed cost, and $G(\omega)$ its time invariant distribution. It is possible to characterize the problem of the firm in terms of two functions similar to those used before: $V(Z, K^*)$ and $\tilde{V}(Z, K^*, \omega)$, the value of a firm with imbalance Z, desired capital K^*, and realization of fixed adjustment cost ω. In particular, $V(Z, K^*)$ is the value of the firm provided it does not adjust, while $\tilde{V}(Z, K^*, \omega)$ is the value of the firm when it is left free to choose whether or not to adjust. Thus,

$$V(Z_t, K_t^*) = \Pi(Z_t, K_t^*)\Delta t + (1 - r\Delta t)\mathrm{E}_t\left[\tilde{V}(Z_{t+\Delta t}, K_{t+\Delta t}^*, \omega_{t+\Delta t},)\right], \tag{3.15}$$

$$\tilde{V}(Z_t, K_t^*, \omega) = \max\left\{V(Z_t, K_t^*), \max_{\eta}\left\{V(Z_t + \eta, K_t^*) - C(\eta, K_t^*, \omega_t)\right\}\right\}. \tag{3.16}$$

Not surprisingly, the nature of the solution is not too different from that of the (L, l, u, U) case. Indeed, conditional on ω it is an (L, l, u, U) rule, although there are additional intertemporal considerations, since the firm weighs the likelihood of drawing higher or lower adjustment costs in the future. Without conditioning on ω, it is a probabilistic rule in the space of imbalances.

In order to simplify the exposition, I will suppress the proportional costs. Thus, conditional on adjustment, the target point is the same regardless of whether the firm is adding or subtracting to its stock of capital (i.e. $l = u = c$). Moreover, let me define a new imbalance index centered around zero:

$$x \equiv \ln(Z/c).$$

The probability of adjustment rises with the absolute value of x because there are more realizations of adjustment costs which justify adjustment. This is the sense in which the (S, s) nature of the simpler models is preserved. Let $\Lambda(x)$ denote the function describing the probability of adjustment given x, and call it the adjustment hazard function [see Caballero and Engel (1999)].

Given an imbalance x, it is no longer possible to say with certainty whether or not the firm will adjust, but the expected investment by the firm is given by

$$\mathrm{E}\left[\frac{I_{it}}{K_{it}}\middle| x\right] = (\mathrm{e}^{-x} - 1)\Lambda(x) \approx -x\Lambda(x), \tag{3.17}$$

which is simply the product of the adjustment if it occurs, and the probability that adjustment occurs[35]. Aggregation is now only a step away.

[35] Caballero and Engel (1999) refer to $\Lambda(x)$ as the "effective hazard" to capture the idea that, through a normalization, it also captures scenarios where adjustment, if it occurs, is only a fraction of the imbalance x.

3.2. Aggregation

Unlike microeconomic data, aggregate investment series look fairly smooth. Large microeconomic adjustments are far from being perfectly synchronized. The question arises, and this was the maintained hypothesis during the 1980s, as to whether aggregation eliminates all traces of lumpy microeconomic adjustment. The answer is a clear no. Doms and Dunne's evidence on the role of synchronization of primary spikes in accounting for aggregate investment, and on the high time series correlation between aggregate investment and a Herfindahl index of microeconomic investments, as well as the more structural empirical evidence reviewed in the next section, support this conclusion.

With the setup at hand, aggregation proceeds in two easy steps. To simplify things further, I will define the aggregate as the behavior of the average, rather than the weighted average[36].

Both steps rely on having a large number of establishments, so that laws of large numbers can be applied. In the first step, one takes as the average investment rate (i.e. the ratio of investment to capital) of establishments with more or less the same imbalance of capital, x, the conditional expectation of this ratio given in Equation (3.17):

$$\left(\frac{I_t}{K_t}\right)^x = -x\Lambda(x), \tag{3.18}$$

where the superscript x denotes the aggregate for plants with imbalance x.

The second step just requires averaging across all x. Let $f(x, t)$ denote the cross sectional density of establishments' capital imbalances just before investment takes place at time t. Then the aggregate investment rate at time t, $(I_t/K_t)^A$, is

$$\left(\frac{I_t}{K_t}\right)^A = -\int x\Lambda(x) f(x, t)\,\mathrm{d}x. \tag{3.19}$$

This is an interesting equation, with macroeconomic data on the left and microeconomic data on the right-hand side. An example serves to illustrate this aspect of the investment equation: If the adjustment hazard is quadratic,

$$\Lambda(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2,$$

Equation (3.19) reduces to

$$\left(\frac{I_t}{K_t}\right)^A = -\lambda_0 X_t^{(1)} - \lambda_1 X_t^{(2)} - \lambda_2 X_t^{(3)}, \tag{3.20}$$

where $X_t^{(1)}$, $X_t^{(2)}$ and $X_t^{(3)}$ denote, respectively, the first, second and third moments of the distribution of establishments' imbalances.

[36] Using microeconomic data, Caballero, Engel and Haltiwanger (1995) show that in US manufacturing this approximation is. See Caballero and Engel (1999) for a detailed discussion of the issue.

If $\lambda_1 = \lambda_2 = 0$, the model only has aggregate variables, both on the right and left-hand side. Indeed this case corresponds to the celebrated partial adjustment model, and it also coincides with the equation obtained from a quadratic adjustment cost model with a representative agent [e.g. Rotemberg (1987) and Caballero and Engel (1999)]. If either λ_1 or λ_2 is different from zero, however, information about the cross sectional distribution of imbalances is needed on the right-hand side. All the microeconomic models discussed in this section yield situations where higher moments of the cross sectional distribution play a role.

3.3. Empirical evidence

There are two polar empirical strategies used to estimate Equation (3.19), with a continuum of possibilities in between. At one extreme, one can use microeconomic data to construct all the elements on the right-hand side; in particular one can construct the path of the cross sectional distribution and estimate the adjustment hazard as an accounting identity, or estimate a parametric version of it. At the other extreme, one can attempt to learn about the adjustment hazard from aggregate data only, by putting enough structure on the stochastic processes faced by firms and by starting with a guess on the initial cross sectional distribution. Both avenues have been explored, with similar results along dimensions they can be compared.

3.3.1. Microeconomic data

Caballero, Engel and Haltiwanger (1995) use information on approximately seven thousand US manufacturing plants from 1972 to 1988 to empirically recreate the steps described in the previous section [37]. The figures below were constructed with data from that paper [38].

The procedure used by Caballero, Engel and Haltiwanger is essentially accounting, except for the first step, which requires estimating a series of frictionless capital for each establishment, and, from this, a measure of x_{it} (an index of the capital imbalance of firm i at date t).

The series of frictionless capital were constructed using a procedure similar to that described in Section 2, but cointegration regressions were run at the individual establishment level [39]. The average estimate of the long run elasticity of capital with

[37] As in Doms and Dunne (1993), we used data from the Longitudinal Research Datafile (LRD). The LRD was created by longitudinally linking the establishment-level data from the Annual Survey of Manufacturing. The data used in that paper is a subset of the LRD, representing all large, continuously operating plants over the sample. The data sets include information on both investment and retirement of equipment (i.e. the gross value of assets sold, retired, scrapped, etc.

[38] Warning: x in that paper corresponds to $-x$ in this survey.

[39] The results reported there constrained the coefficient on the elasticity of capital with respect to its cost to be equal across two-digit sectors, but all principal results were robust to different constraints and specifications.

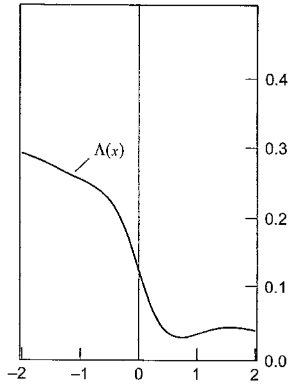

Fig. 3.5. Adjustment hazard.

respect to its cost was close to minus one, with substantial heterogeneity across sectors. The measures of x_{it}, up to a constant, correspond to the difference between actual and estimated frictionless capital[40].

There are two results from that paper which seem particularly relevant for this section of the survey. One on the shape of the adjustment hazard, and the other on the consequences of this shape for aggregate dynamics. I discuss the former here and the latter after the next subsection. Figure 3.5 reports the average adjustment hazard constructed from simply averaging the investment rates of establishments in a small neighborhood of each x, divided by minus the corresponding x. The hazard is clearly increasing for positive adjustment (i.e. expected investment rises more than proportionally with the shortage of capital), as one would expect from the nonlinearities implied by (L, l, u, U) type models, and unlike the linear models which imply a constant hazard. The estimated hazard is also very low for negative changes, suggesting irreversibility[41].

Following a similar procedure, Goolsbee and Gross (1997) have studied very detailed and high quality microeconomic data on capital stock decisions in the US airline industry. They found clear evidence of behavior consistent with non-convex adjustment costs.

3.3.2. Aggregate data

If only aggregate data are available, one needs to make some inference about the path of the cross sectional distribution of capital imbalances, $f(x, t)$, from these data. This is possible if enough structure is placed on the stochastic processes faced by firms.

[40] The establishment specific constants were estimated as the average gap between their respective k_{it} and k^*_{it} for the five points with investment closest to their median (broadly interpreted as maintenance investment).

[41] Retirements include assets sold, scrapped or retired. It is possible that observations are very noisy on this side. The right-hand side of the figure should therefore be viewed with some caution.

The basic operations affecting the evolution of $f(x,t)$ are quite simple. Given the density, or histogram, at time $t-1$, there are three basic operations in its transformation into $f(x,t)$. First, aggregate shocks and common depreciation shift everybody's x in the same direction; second, given the adjustment hazard, the density at each x is split into those that stay there and those that adjust and move to some other position in the state space (in the simplest case, they move to $x = 0$, but this is not necessary); and third, idiosyncratic shocks hit, which amounts to a convolution of the density resulting from the second step and that of idiosyncratic shocks. Making distributional assumptions about idiosyncratic shocks and the initial cross sectional distribution, is enough, therefore, to keep track of the evolution of the cross sectional density, conditional on aggregate shocks and for a given adjustment hazard.

In continuous time, and assuming Brownian motions for aggregate and idiosyncratic shocks, Bertola and Caballero (1994) estimated the irreversible investment model, and Caballero (1993b) did so for the (L,l,u,U) model.[42]

In discrete time but continuous state space, Caballero and Engel (1999), estimated the more general adjustment hazard model described in the previous sections, assuming that both idiosyncratic and aggregate shocks were generated by log-normal processes. We did so for US manufacturing investment in equipment and structures (separately) for the 1947–1992 period[43]. The results were largely consistent with those found with microeconomic data by Caballero, Engel and Haltiwanger (1995). There is clear evidence of an increasing hazard model; that is, the expected adjustment of a firm grows more than proportionally with its imbalance[44]. An important point to note is that since only aggregate data were used, these microeconomic nonlinearities must matter at the aggregate level, for otherwise they would not be identified. The improvement in the likelihood function from estimating this non-linear model rather than a simple linear model (including the quadratic adjustment cost model) was highly significant, and so was the improvement in the out-of-sample forecasting accuracy[45].

[42] See Bertola and Caballero (1990) for a discrete time and space model and estimation procedure.

[43] Another important difference between this and the previous papers is that estimation was done by a single step maximum likelihood procedure, which did not require estimating frictionless capital separately.

[44] We did not allow for asymmetries between ups and downs but this turned out not to matter much because given the strong drift induced by depreciation and the small value we found for the hazard in an interval around zero, the model effectively behaves as if investment is irreversible (i.e. It is very asymmetric around the median value of x and with a very small hazard for values of x much higher than that.).

[45] For within sample criteria, we ran Vuong's [Rivers and Vuong (1991)] test for non-nested models, and we rejected strongly the hypothesis that both models (linear and non-linear) are equally close to the true model against the hypothesis that the structural (non-linear) model is better. For out-of-sample criteria, we dropped the last ten percent of the observations and evaluated the Mean Squared Error of the one step ahead forecasts for these observations [see Caballero and Engel (1999)].

3.3.3. Pent-up demand

What is the aspect of the data that makes these models better than linear ones at explaining aggregate investment dynamics?

The simplest answer comes from an example. Suppose that a history of mostly positive aggregate shocks displaces the cross sectional distribution of imbalances toward the high part of the hazard. Such a sequence of events will not only lead to more investment along the path but also to more pent-up investment demand; indeed, the cross sectional distribution represents unfulfilled investment plans. But as unfulfilled demand "climbs" the hazard, more units are involved in responding to new shocks; incremental investment demand is more easily boosted by further positive aggregate shocks, or depressed by a turnabout of events. This time-varying/history-dependent aggregate elasticity plays a very important role for aggregate investment dynamics. It captures the aggregate impact of changes in the degree of synchronization of large adjustments; already an important explanatory variable in Doms and Dunne's less structural study. In particular, their observation that the Herfindahl of investment rises during episodes of large aggregate investment matches well this mechanism.

Using the path of cross sectional distributions and hazards described at the beginning of this subsection, Caballero, Engel and Haltiwanger (1995) found an important role for the mechanism described above. Figure 3.6 depicts the relative contribution of the time-varying aggregate elasticity for aggregate investment dynamics. A positive value reflects an amplification effect (micro-nonlinearities exacerbate the economy's response to aggregate shocks), while a negative value reflects an offsetting effect. The impact of the time-varying elasticity appears to be especially large after the tax-reform of 1986 (when tax-incentives for investment were removed). The decline in investment was 20 percent greater than it would have been under a linear model.

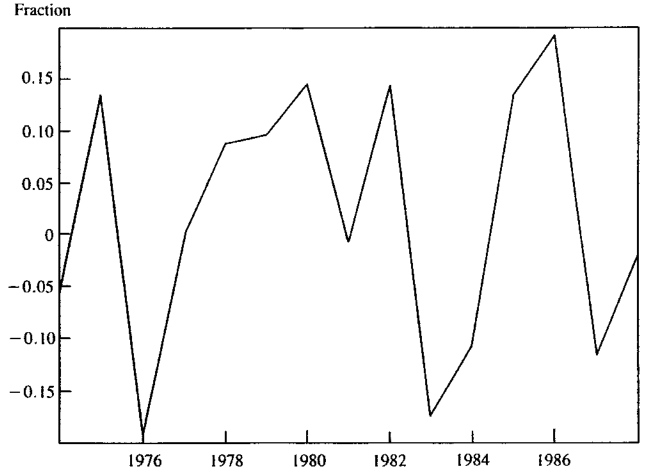

Fig. 3.6. Relative contribution of time-varying marginal response, 1974–1988.

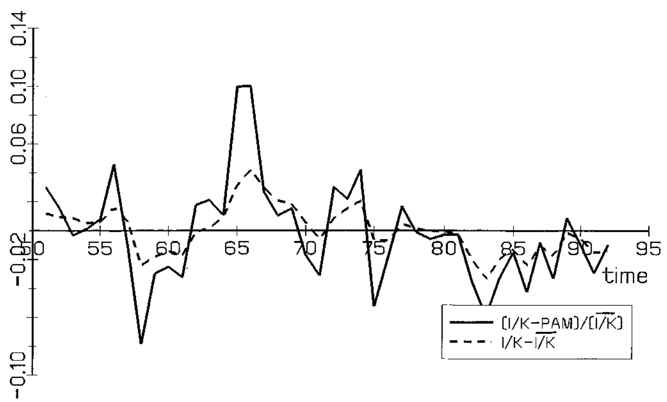

Fig. 3.7. Equipment-mean-difference.

The importance of the time-varying elasticity is confirmed by Caballero and Engel (1999), this time using only aggregate data. As before, it is the flexible cyclical elasticity of the increasing hazard model which allows it to better capture the high skewness and kurtosis imprinted on aggregate data by brisk investment recoveries[46]. The solid line in Figure 3.7 plots the difference between the path of the US manufacturing equipment investment-capital ratio and the predictions of a linear model (partial adjustment) fed with the shocks estimated for the increasing-hazard model; the dashed line portrays the path of the aggregate investment-capital ratio around its mean. It is apparent from these figures that the linear model makes its largest errors at times of large investment changes.

3.4. Equilibrium

The literature described in the previous section only considers exogenous aggregate shocks. What the econometric procedures identified as aggregate shocks are in all likelihood a combination of "deep" aggregate shocks and the feedback and constraints brought about by factor markets, goods markets, and intertemporal preferences, among other things. Bottlenecks may certainly limit the extent of synchronized investment.

Equilibrium constraints not only affect the response of aggregate investment to deep aggregate shocks, but also affect the nature of the stochastic processes faced

[46] Note that just allowing for skewness and kurtosis in shocks, although it improves the performance of linear models, is not nearly enough to make the linear model as good as the non-linear one. In Caballero and Engel (1999) we compared the structural model with normal shocks (to the rate of growth of desired capital) with a linear model which flexibly combined normal and log-normal shocks (which allows for skewness and kurtosis). We found that Vuong's test still favored the non-linear model very clearly. Moreover, in Caballero, Engel and Haltiwanger (1995) we found no evidence that would allow us to reject the hypothesis that shocks have a normal distribution.

by firms and the dimension of the state space. It is this last observation which has inhibited progress in constructing general equilibrium versions of these models. In principle, the entire cross sectional distribution is needed to forecast future prices faced by any particular firm, which means that actions today, and therefore equilibrium determination, depend on these complex forecasts, and so on.

We are, however, beginning to see progress along this dimension. Much of this has occurred in models with active extensive margins, and will be discussed in the next sections, together with the reasons why the presence of an extensive margin (entry and/or exit) may facilitate rather than complicate the solution of the model.

However, there has also been recent progress along the lines of the intensive margin models discussed up to now. Krieger (1997) embeds the heterogeneous agents irreversible investment model of Bertola and Caballero (1994) into a more or less standard Real Business Cycle model. He deals with the curse of dimensionality by arguing that, except for very high frequency aspects of the data, expectations can be well approximated by keeping track of a finite (and not too large) number of statistics of the Fourier representation of the cross sectional distribution. I suspect that the quality of this approximation is facilitated by the fact that, in Krieger's model, aggregate shocks occur only infrequently. Nonetheless, I view his as an important step forward.

At this stage, the primary effect of general equilibrium is not surprising. It brings important sources of aggregate convexity into the problem, smoothing further the response of aggregate investment to aggregate shocks. How important are aggregate sources of convexity? I suspect that, together with time to build considerations, they are among the *main* sources of convexity in the short run. On one hand, we have already presented substantial evidence on microeconomic lumpiness, which is largely inconsistent with a dominant role for generalized convexity at the microeconomic level. On the other, not only is it well known that estimated partial adjustment coefficients grow with the degree of disaggregation of the data, but we also have direct evidence on the importance of bottlenecks. Goolsbee (1995a) provides interesting evidence on the latter. He exploits the variation across time and assets (capital) in investment tax incentives, as instruments for short-run investment demand. He shows that the price of assets is highly responsive to ITCs: A 10 percent increase in ITCs leads to an average increase in the price of capital goods of about 6 percent. This price effect slowly vanishes over the following three years[47].

Equilibrium considerations will play a central role in the sections that follow. In particular, the issue of the elasticity of the supply of capital, generally interpreted, as well as that of other bottlenecks will be revisited often.

[47] In further work, Goolsbee (1997) concludes that an important fraction of the increase in short run marginal cost is due to an increase in the wages of workers who produce capital goods. In the last part of Section 5 I will discuss the connection between sunk investment and payments to complementary factors.

Questioning the robustness of Goolsbee's (1995a) findings, Hassett and Hubbard (1996a), find evidence of a positive effect of tax credits on prices of capital goods before 1975 but not after that.

4. Entry, exit and scrapping

Changes in the aggregate stock of capital are not only due to the expansion of existing establishments and projects, but also result from the entry (creation) decisions of existing and new entrepreneurs, the exit decisions of some incumbents, and the restructuring of possibly outdated forms of production. There is a very extensive and interesting industrial organization literature on these issues which I will not discuss here. Instead, I will focus on issues that directly relate to our current discussion: the impact of sunk costs on aggregate investment and the feedback of equilibrium considerations into individual decisions about lumpy actions.

This section contains three main messages: First, by truncating the distribution of perceived future returns, free entry acts *as if* each competitive investor internalized the negative effect of its entry decision on expected future industry prices. Second, equilibrium scrapping and creation are closely connected: if industry wide creation costs are linear, scrapping will be less responsive to aggregate shocks than if these costs are convex (i.e. if there is an upward sloping short-run supply of (newly) installed capital). Among other things, this is important for capital accumulation and the patterns of its mismeasurement. And third, in equilibrium, shocks to the scrapping margin can lead to investment booms, and to double-counting problems in the measurement of capital[48].

4.1. Competitive entry and irreversibility

Dixit (1989), Leahy (1993), and Caballero and Pindyck (1996), among others, have provided simple models of competitive equilibrium investment in which the only meaningful investment decision of firms is whether or not to enter into and, in some cases, exit from the industry[49]. Below, I sketch a representative model of this type.

Investment is sunk upon entry in the sense that selling the firm's capital does not change its productivity. The flow accruing to a firm i at time t is summarized by the product of an idiosyncratic productivity level, $S_{it} > 0$ and the industry price, P_t. The idiosyncratic productivity level is such that industry output, Y_t, is

$$Y_t = \int_0^{N_t} S_{it}\,\mathrm{d}i = N_t, \tag{4.1}$$

where N_t is the measure of firms at time t. Given N_t, the industry price is determined from the demand equation:

$$P_t = V_t Y_t^{-1/\eta} = V_t N_t^{-1/\eta}, \tag{4.2}$$

[48] See Greenspan and Cohen (1996), for a discussion of the importance of considering endogenous scrappage to forecast sales of new motor vehicles in the USA.

[49] See Hopenhayn (1992) for an elegant characterization of the steady state properties of a competitive equilibrium model of entry and exit.

where V_t is an aggregate demand shock that follows a geometric Brownian motion with drift $\mu > 0$ and standard deviation σ, and η is the elasticity of demand with respect to price [50].

Let there be an infinite supply of potential entrants, whose initial productivity upon entry is drawn from the distribution of productivities of existing firms. There is an entry cost F and no depreciation or higher productivity alternative use (issues of exit will be discussed in the next subsection). Free entry implies:

$$F \geqslant \mathrm{E}_i \left\{ \mathrm{E}_t \left[\int_t^{\infty} P_s S_{is} \mathrm{e}^{-r(s-t)} \, \mathrm{d}s \right] \right\}. \tag{4.3}$$

Using Fubini's Theorem (i.e. moving the expectation with respect to the idiosyncratic shocks inside the integral) allows us to remove the idiosyncratic component from Equation (4.3), yielding

$$F \geqslant \mathrm{E}_t \left[\int_t^{\infty} P_s \mathrm{e}^{-r(s-t)} \, \mathrm{d}s \right]. \tag{4.4}$$

Given N_t, the industry price is exclusively driven by the aggregate demand shock. Thus, absent entry, the right-hand side of Equation (4.4) is an increasing function of P_t, call it $f_0(P)$. Entry, however, cannot always be absent, for that would occasionally violate the free entry condition. Indeed, as soon as $f_0(P) > F$, there would be infinite entry which, in turn, would lower the equilibrium price instantly. There is only one price, call it $\bar{P}_0$, such that the free entry condition holds with equality. Once this price is reached, enough entry will occur to ensure that the price does not cross this upper bound; but, to be justified, entry must not occur below that bound either. Entry, therefore, changes the stochastic process of the equilibrium price from a Brownian Motion to a *regulated* Brownian Motion. This change in the price process, however, means that f_0 is no longer the right description of the expression on the right-hand side of Equation (4.4). There is a new function, $f_1(P)$, which is still monotonic in the price, but which satisfies $f_1(P) < f_0(P)$ for all P because of the role of entry in preventing the realization of high prices. This, in turn, implies a new reservation/entry price $\bar{P}_1 > \bar{P}_0$, which leads to a new function $f_2(P)$, such that $f_0 > f_2 > f_1$, which leads to a new regulation point in between the previous ones, and so on until convergence to some equilibrium, $(f(P), \bar{P})$ [51].

Thus, through competitive equilibrium, we have arrived at a solution like that of the irreversible investment problem at the individual level, but now for the industry as a whole. Periods of inaction are followed by regulated investment (through entry) during favorable times. The constructive argument used to illustrate the solution isolates

[50] Adding an aggregate productivity shock is straightforward. The Brownian Motion assumption is not needed, but it simplifies the calculations.

[51] Needless to say, this iterative procedure is not needed to obtain the solution of this problem.

the feedback of equilibrium on individual decisions. Potential entrants (investors) know that if market conditions worsen they will have to absorb losses (this is where irreversibility kicks in), while if market conditions improve, entry will occur, limiting the potential gains (since the price will never be higher than $\bar{P}$). As a result, they delay entry because the expected value of future market prices is necessarily lower than the current/entry price.

There is a methodological angle in this literature. Entry (and exit) is a very powerful mechanism. With the "appropriate" assumptions about potential entrants, entry often simplifies the computation of equilibrium in models with heterogeneity and sunk costs. Essentially, the methodological "trick" is that the degree of complexity of the computational problem in cases where both extensive and intensive margins are present is often largely determined by the nature of the distribution of potential entrants, which can be made much simpler than the endogenous evolution of the cross sectional distributions discussed in the previous section. Of course, in reality there is substantial inbreeding, so the distribution of potential entrants is in all likelihood related to that of incumbents. Nonetheless, the current set of models are convenient machines that allow us to cut the chain of endogeneity before it gets too forbidding, but after the first stage, where there are no endogenous interactions.

This methodological advantage has allowed researchers to explore some of the equilibrium issues left open in Section 3. Caballero and Hammour (1994) have explored in more detail the consequences of different assumptions on the supply of capital for the pattern of aggregate investment (job creation) and scrapping (job destruction). The latter is a very important, and often disregarded, aspect of the timing of capital accumulation. I will return to the scrapping issue in the next sections, but for now I just want to interpret it as an incumbent's decision (as opposed to a potential entrants' decision). The issue at hand is how does the entry pattern affect the response of incumbents to aggregate shocks.

A scrapping margin can easily be added to the entry model discussed above by, for example, allowing S_i to take negative values (e.g. due to the increase in the price of an intermediate input). Imagine, however, that the drift in the aggregate shock (and/or the failure rate of incumbents) is strong enough so there is continuous entry. Since the supply of capital faced by the industry is fully elastic (the entry cost is constant), continuous entry implies that the industry price is constant and equal to $\bar{P}$ (corrected for the exit possibility). That is, aggregate shocks are accommodated by the flow of investment by new entrants; fully *insulating* insiders from aggregate shocks. Insiders go about their scrapping decisions only considering their idiosyncratic shocks; adding a standard intensive margin does not change the basic insight [see Campbell and Fisher (1996)]. Caballero and Hammour (1994) refer to this result as *perfect insulation*.

From a technical point of view, the simplicity of the computation of equilibrium in the perfect insulation case carries through to situations where the cost of investment fluctuates exogenously, although in that case perfect insulation breaks down. If the industry faces an upward sloping supply of capital, a sensible assumption at least in the short run (remember Goolsbee's evidence), we return to a scenario in which the "curse of dimensionality" appears. Caballero and Hammour (1994, 1996a) have dealt with this case in scenarios where aggregate shocks follow deterministic cycles [52]. Besides the specific issues addressed in those papers, the main implication for the purpose of this survey is that investment by potential entrants becomes less responsive to aggregate shocks, which also means a break down of perfect insulation and therefore a more volatile response of the scrapping and intensive margins.

Krieger (1997) also discusses equilibrium interactions between creation and destruction margins, although he obtains positive rather than negative comovement between investment and scrapping. In his model, a permanent technology shock leads to a short term increase in interest rates which squeezes low productivity units relative to high productivity ones. The ensuing increase in scrapping frees resources for new higher-productivity investment. Similarly, Campbell (1997) studies the equilibrium response of entry and exit to technology shocks embodied in new production units. He argues that the increase in exit generated by positive technological shocks is an important source of resources for the creation of new production sites.

4.2. Technological heterogeneity and scrapping

Scrapping is an important aspect of the process of capital accumulation. Understanding it is essential for constructing informative measures of the quantity and quality of capital at each point in time. Nonetheless, the scrapping margin is seldom emphasized, I suspect, mostly because of the difficulties associated with obtaining reliable data [53]. As a result, many time series comparisons of capital accumulation and productivity growth (especially across countries) are polluted by inadequate accounting of scrapping. Effective capital depreciation must surely be higher in countries undergoing rapid modernization processes.

Partly to address these issues, vintage capital and putty-clay models have regained popularity lately. Benhabib and Rustichini (1993), for example, describe the investment cycles that follow scrapping cycles in a vintage capital model. While Atkeson and Kehoe (1997) argue that putty-clay models outperform standard putty-putty models with adjustment costs in describing the cross sectional response of investment and output to energy shocks. Gilchrist and Williams (1996), on the other hand, embody the putty-clay model in an otherwise standard RBC model and document a substantial gain over the standard RBC model in accounting for the forecastable comovements of economic aggregates. And Cooley et al. (1997) describe the medium/low frequency

[52] In work in progress [Caballero and Hammour (1997b)], we have obtained an approximate solution for the stochastic case, in a context where the sources of convexity are malfunctioning labor and credit markets.

[53] See Greenspan and Cohen (1996) for sources of scrapping data for US motor vehicles.

aspects of a multisectoral vintage capital economy, and show how tax policy can have significant effects on the age distribution of the capital stock[54].

The technological embodiment aspect of these models captures well the creative-destruction component of capital accumulation and technological progress[55]. Salter's (1960) careful documentation of the technological status of narrowly defined US and UK industries is very revealing with respect to the simultaneous use of different techniques of production and the negative correlation between productivity ranking and the technological age of the plant[56]. For example, his table 5 shows the evolution of methods in use in the US blast furnace industry from 1911 to 1926. At the beginning of the sample, the "best practice" plants produced 0.32 gross tons of pig-iron per man-hour, while the industry average was 0.14. By the end of the sample, best practice plants productivity was 0.57 while the industry average was 0.30. While at the beginning of the sample about half of the plants used hand-charged methods of production, only six percent did at the end of the sample.

As mentioned above, obsolescence and scrapping are not only driven by slowly moving technological trends, but also by sudden changes in the economic environment. Goolsbee (1995b) documents the large impact of oil shocks on the scrapping of old and fuel-inefficient planes. For example, he estimates that the probability of retirement of a Boeing 707 (relatively inefficient in terms of fuel) more than doubled after the second oil shock. This increase was more pronounced among older planes. Once more, the endogenous nature of the scrapping dimension must be an important omitted factor in our accounting of capital accumulation and microeconomic as well as macroeconomic performance.

The sunk nature of technological embodiment is a source of lumpy and discontinuous actions at the microeconomic level. The (S, s) apparatus, with its implications for aggregates, is well suited for studying many aspects of vintage and putty-clay models. In particular, episodes of large investment which leave their technological fingerprints, and remain in the economy, reverberating over time.

5. Inefficient investment

Fixed costs, irreversibilities and their implied pattern of action/inaction, have microeconomic and aggregate implications beyond the mostly technological (and neoclassical) ones emphasized above. Indeed, they seed the ground for powerful inefficiencies. This section describes new research on the consequences of two of

[54] Jovanovic (1997) studies the equilibrium interaction of the cross sectional heterogeneity implied by vintage capital and putty-clay models with heterogeneity in labor skills.

[55] Besides obsolescence and scrapping, these models are also useful for studying the issues of "mothballing" and capital utilization.

[56] This correlation is less clear in modern data; perhaps because retooling occurs within given structures.

the most important sources of inefficiency in aggregate investment: informational and contractual problems.

5.1. Informational problems

Information seldom arrives uniformly and comprehensively to every potential investor. Each investor probably holds part of a truth which would be more easily seen if all investors could (or would) pool their information. Actions by others are a partial substitute for information pooling, for they reveal, perhaps noisily, the information of those that have taken actions.

If, however, investment is irreversible, it may pay to wait for others to act and reveal their information before investing. Moreover, if lumpiness leads to periods of no or little action, information may remain trapped for extended periods of time, and when agents finally act, an avalanche may occur because accumulated private information is suddenly aggregated. These issues form the crux of a very interesting new literature, summarized in Gale (1995) under the heading of "social learning."

There are two themes emerging from this literature which are of particular importance for this survey. The first is the existence of episodes of gradualism, during which industry investment can occur at an excessively slow pace, or even collapse altogether. The second is an exacerbation of the aggregate nonlinearities implied by the presence of fixed costs; aggregation of information coincides with the synchronization of actions, further synchronizing actions.

Caplin and Leahy (1993, 1994) cleanly isolate the issues I have chosen to stress here. Caplin and Leahy (1993) describe a model very similar to the free entry model reviewed in Section 4.1, except that their model has neither aggregate nor idiosyncratic shocks. Instead there is a flow marginal cost of producing which is only known to industry insiders. Insiders have the option to produce one unit of output or none and they will produce if price is above marginal cost. This generates an information externality. If all incumbents are producing, potential investors know that marginal cost is below the current equilibrium price; if not, the industry's marginal cost is revealed to be equal to the current price. Whenever a new establishment is created, equilibrium price either declines or stays constant, improving the precision of potential investors' assessment of the industry's marginal cost.

In a second best solution, investment occurs very quickly up to a point at which, even if marginal cost has not yet been reached, no further investment takes place because it is very unlikely that the present value of future social surpluses is enough to cover the investment costs.

The industry equilibrium outcome has the same critical point at which investment stops, but unlike the second best outcome, it yields a much slower pace of industry investment. A potential entrant must weigh the value of coming early into the industry (expected profits are higher than they will be later), not only against the cost of capital (as in the second best solution) but also against the probability of learning in the next second from the investment decisions of others that it was not worth entering the industry. Caplin and Leahy show that the price process $x(t)$ obeys the following differential equation:

$$\tfrac{1}{2}x(t) = rF - F\frac{\dot{x}(t)}{x(t)}, \tag{5.1}$$

where F is the fixed entry cost paid by the firm and r is the real interest rate. This equation has a natural interpretation which captures the idea that competitive firms are indifferent between entry today and entry tomorrow. The left-hand side represents the loss in current revenue incurred by a firm which delays entry for a brief instant beyond t[57]. The right-hand side captures the expected gain from this delay. The term rF reflects the gain due to the postponement of the entry cost, while the last term represents the saving due to the possibility that delay will reveal the true industry's marginal cost, aborting a wasteful investment[58].

In equilibrium, entry is delayed and price declines slowly; "gradualism" maintains prices high enough for sufficiently long so as to offset (in expectation) the risk incurred by investors who act early rather than wait and free-ride off of others' actions[59].

Caplin and Leahy (1994) characterize the opposite extreme, one of delayed exit. The key connection with the previous sections is that the problem of information revelation arises from the fact that, as we have seen, fixed costs of actions make it optimal not to act most of the time. Thus, information that could be revealed by actions remains trapped.

Their model is one of time-to-build. Many identical firms simultaneously start projects which have an uncertain common return several periods later (e.g. a real estate boom). Along the investment path, firms must continue their investment and receive private signals on the expected return. The nature of technology is such that required investment is always the same if the firm chooses to continue in the project. The firm has the option to continue investing ("business as usual"), to terminate the project, or to suspend momentarily, but the cost of restarting the project after a suspension is very large. Project suspension reveals (to others) negative idiosyncratic information; if nobody suspends, it is good news. However, the costly nature of suspension delays it, and therefore information revelation is also delayed. Bad news may be accumulating but nobody suspends, because everybody is waiting for a confirmation of their bad signals by the suspension of other people. Eventually, some firms will receive enough bad signals to suspend in spite of the potential cost of doing so (i.e., if they are wrong

[57] At the time when the industry starts, potential investors' priors are that the price is distributed uniformly on $[0, 1]$. As entry occurs and the price declines, the priors are updated. If convergence has not happened at time t, marginal cost is assumed uniformly distributed on $[0, x(t)]$. The expected cost of waiting is, therefore, equal to the price minus the expected marginal cost, $\frac{1}{2}x(t)$.

[58] Here $\frac{\dot{x}(t)}{x(t)}\mathrm{d}t$ is the probability that price hits marginal cost during the next $\mathrm{d}t$ units of time.

[59] Even though entrants make zero profits in expectation, ex-post, early entrants earn positive profits, while late entrants lose money.

in their negative assessment of market conditions). Since the number of firms in their model is large, the number of firms that suspend for the first time fully reveals future demand: if demand is low, everybody exits; if it is high, all those that suspended restart.

If it were not for the interplay between inaction (investment as usual) and private information, the fate of the market would be decided correctly after the first round of signals. Information aggregation does not take place until much later, however. Thus, substantial investment may turn out to be wasted because the discrete nature of actions inhibits information transmission. The title of their paper beautifully captures the ex-post feeling: Wisdom after the fact.

The "classic" paper from the literature on information and investment is due to Chamley and Gale (1994). In their model all (private) information arrives at time zero; the multiple agent game that ensues may yield many different aggregate investment paths, including suboptimal investment collapses. In reviewing the literature, Gale (1995) illustrates the robustness of the possibility of an inefficient investment collapse (or substantial slowdown and delay). He notices that in order for there to be any value to waiting to see what others do before taking an action (investing for example) it must be the case that the actions of others are meaningful. That is, the action taken in the second period by somebody who chose to wait in the first period must depend in a non trivial way on the actions of others at the first date. If a firm chooses to wait this period, possibly despite having a positive signal, it will only invest next period if enough other firms invest this period. It must therefore be possible for every firm to decide not to invest next period because no one has invested this period, even though each firm may have received a positive signal this period, in which case, investment collapses.

This is a very interesting area of research for those concerned with investment issues and is wanting for empirical developments.

5.2. *Specificity and opportunism*

The quintessential problem of investment is that it is almost always sunk, possibly along many dimensions. That is, the number of possible uses of resources is reduced dramatically once they have been committed or tailored to a specific project or use. Every model I discussed in the previous sections, at some stage hinges in a fundamental way on this feature of investment.

To invest, often means opening a vulnerable flank. Funds which were ex-ante protected against certain realizations of firm or industry specific shocks, for example, are no longer so. In equilibrium, investment must also allow the investor to exploit opportunities which would not be available without the investment. If the project is well conceived, the weight of good and bad scenarios is such that the expected return is reasonable. Indeed, this is precisely the way I characterized the standard irreversible investment problem early on.

The problem is far more serious, and more harmful for investment, when the probability of occurrence of the bad events along the exposed flanks are largely controlled by economic agents with the will and freedom to behave opportunistically. In a sense, this is a property rights problem, and as such it must have a first-order effect in explaining the amount and type of capital accumulation and, especially, differences in these variables across countries.

Thus, the window for opportunism arises when part of the investment is specific to an economic relationship, in the sense that if the relationship breaks up, the potential rewards to that investment are irreversibly lost. Further, such opportunism is almost unavoidable when this "fundamental transformation" from uncommitted to specialized capital is not fully protected by contract [Williamson (1979, 1985)] [60].

Specificity, that is, the fact that factors of production and assets may be worth more inside a specific relationship than outside of it, may have a technological or an institutional origin. Transactions in labor, capital and goods markets are frequently characterized by some degree of specificity. The creation of a job often involves specific investment by the firm and the worker. Institutional factors, such as labor regulations or unionization also build specificities.

There is a very extensive and interesting microeconomic literature on the impact of unprotected specificity on the design of institutions, organizations and control rights. Hart (1995) reviews many of the arguments and insights. For the purpose of this survey, however, the fundamental insight is in Simons (1944), who clearly understood that hold-up problems lead to underinvestment:

> ... the bias against new investment inherent in labor organizations is important .... Investors now face ... the prospect that labor organizations will appropriate most or all of the earnings .... Indeed, every new, long-term commitment of capital is now a matter of giving hostages to organized sellers of complementary services.

More recently, Grout (1984) formalized and generalized Simons' insight, and Caballero and Hammour (1998a) studied, at a general level, the aggregate consequences of opportunism [61]. Here, I borrow the basic model and arguments from that paper to discuss those aspects of the problem which are most relevant for aggregate investment.

Everything happens in a single period [62]. There is one consumption good, used as a numeraire, and two active factors of production, 1 and 2 [63]. Ownership of factors 1 and 2 is specialized in the sense that nobody owns more than one type of factor.

[60] This is known as the hold-up problem.

[61] For specific applications which relate to investment see Kiyotaki and Moore (1997) [credit constraints]; Caballero and Hammour (1996a, 1998b) and Ramey and Watson (1996) [turnover and unemployment]; Caballero and Hammour (1996b), Blanchard and Kremer (1996) [transition economies and structural adjustments]; Caballero and Hammour (1997b) [interactions between labor market and credit market opportunism]; Acemoglu (1996) [human capital investment].

[62] Many of the insights discussed here can and have been made in dynamic, but more specialized contexts. I am confident, therefore, that this section's discussion is fairly robust to generalizations along this dimension.

[63] Also, there is a passive third factor which earns the rents of decreasing returns sectors.

There are two modes of production. The first is *joint production*, which requires, in fixed proportions, x_1 and x_2 units of factors 1 and 2, respectively, to produce y units of output. Let E denote the number of joint production units, so $E_i = x_i E$ represents employment of factor i in joint production. The other form of production is *autarky* where each factor produces separately, with decreasing returns technologies $F_i(U_i)$, and where U_i denotes the employment of factor i in autarky, such that $E_i + U_i = 1$. The autarky sectors are competitive, with factor payments, p_i:

$$p_i = F_i'(U_i). \tag{5.2}$$

For now, there are no existing units. At the beginning of the period there is mass one of each factor of production. There are no matching frictions so that, in the efficient/complete contracts economy, units move into joint production (assuming corners away) until

$$y = p_1^* x_1 + p_2^* x_2, \tag{5.3}$$

where asterisks are used to denote efficient quantities and prices.

Specificity is captured by assuming that a fraction ϕ_i of each factor of production cannot be retrieved from a relationship once they have agreed to work together. If the relationship breaks up, $(1 - \phi_i)x_i$ units of factor i can return to autarky, where it produces for the period, while $\phi_i x_i$ is irreversibly wasted. In the simple deterministic-single-period model discussed here, specificity plays no role in the efficient economy, where there are no separations.

Contracts are needed because investment occurs before actual production and factor participation. There are myriad reasons why contracts are seldom complete. An extreme assumption which takes us to the main issues most directly, is the assumption that there are no enforceable contracts. It turns out that, in equilibrium, the incomplete contracts economy has no separations either; but unlike the efficient economy, the mere possibility of separations alters equilibrium in many ways.

Generically, equilibrium rewards in joint production will have ex-post opportunity cost and rent-sharing components. For simplicity, let us assume that factors split their joint surplus 50/50. Thus, the total payment to the x_i units of factor i in a unit of joint production is[64]

$$w_i x_i = (1 - \phi_i) x_i p_i + \tfrac{1}{2} s, \tag{5.4}$$

where s denotes the (ex-post) quasi-rents of a production unit:

$$s \equiv y - (1 - \phi_1) p_1 x_1 - (1 - \phi_2) p_2 x_2. \tag{5.5}$$

64 Factors bargain as coalitions within the production unit.

For a factor of production to willingly participate in joint production it must hold that

$$w_i x_i \geqslant p_i x_i. \tag{5.6}$$

Substituting Equations (5.4) and (5.5) into Equation (5.6), transforms factor i's participation condition into

$$y \geqslant p_1 x_1 + p_2 x_2 + \Delta_i, \tag{5.7}$$

with

$$\Delta_i \equiv \phi_i p_i x_i - \phi_j p_j x_j, \tag{5.8}$$

which measures the net sunk component of the relationship for factor i. In other words, it is a measure of the "exposure" of factor i to factor j. When Δ_i is positive, part of factor i's contribution to production is being *appropriated* by factor j[65].

5.2.1. Generic implications

Figure 5.1 characterizes equilibrium in both efficient and incomplete contract economies. The two dashed curves represent the right-hand side of condition (5.7) for factors 1 and 2. They are increasing in the number of production units because the opportunity cost of factors of production (the p_is) rise as resources are attracted away from autarky. The thick dashed curve corresponds to that factor of production (here factor 1) whose return in autarky is less responsive to quantity changes[66]. If one thinks of capital and labor, arguably capital is this factor; which is a maintained assumption through most of this section. The horizontal solid line is a constant equal to y, which corresponds to the left-hand side of condition (5.7). Equilibrium in the incomplete contracts economy corresponds to the intersection of this line with the highest (at the point of intersection) of the two dashed lines. In the figure, the binding constraint is that of capital.

An efficient equilibrium, on the other hand, corresponds to the intersection of the horizontal solid line with the solid line labeled Eff. The latter is just the sum of the *ex-ante* opportunity costs of factors of production [the right-hand side of Equation (5.3)]. This equilibrium coincides with that of the incomplete contracts economy only when both dashed lines intersect; that is, when net appropriation is zero ($\Delta_i = -\Delta_j = 0$).

There are several features of equilibrium which are important for investment (or capital accumulation). First, there is underinvestment; equilibrium point A is to the left of the efficient point A^*. Because it is being appropriated, capital withdraws into autarky (e.g. consumption, investment abroad, or investment in less socially-valuable

[65] It should be apparent that $\Delta_i = -\Delta_j$.

[66] That is, autarky exhibits relatively less decreasing returns for this factor.

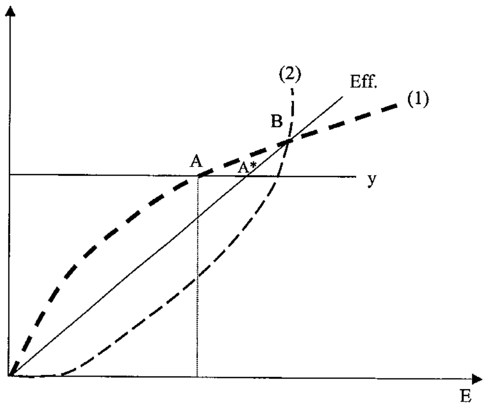

Fig. 5.1. Opportunism in general equilibrium.

activities)[67]. Second, the withdrawal of capital constrains the availability of jobs and segments the labor market[68]. In equilibrium, not only are there fewer joint production units, but also the right-hand side of condition (5.7) for labor is less than y, reflecting the net appropriation of capital; outside labor cannot arbitrage away this gap because its promises are not enforceable. Third, investment is more volatile than it would be in the efficient economy[69]. Changes in y translate into changes in the number of joint production units through capital's entry condition (thick dashes), which is clearly more elastic (at their respective equilibria) than the efficient entry condition ("Eff" line).

If profitability in joint production is high enough, equilibrium is to the right of the balanced specificity point, B. In that region, it is the labor entry condition which binds. In principle, problems are more easily solved in this region through contracts and bonding. If not solved completely, however, there are a few additional conclusions of interest for an investment survey. First, there is underinvestment since the complementary factor, labor, withdraws (relative to the first best outcome) from joint production. Second, capital is now rationed, so privately profitable investment projects do not materialize. Third, investment is now less volatile than in the efficient economy. Changes in y translate into changes in the number of joint production units through labor's entry condition (thin dashes), which is clearly less elastic than the efficient entry condition ("Eff" line).

[67] See Fallick and Hassett (1996) for evidence on the negative effect of union certification on firm level investment.

[68] This holds even in the extreme case where capital and labor are perfect substitutes in production. See Caballero and Hammour (1998a).

[69] In a dynamic model, this translates into a statement about net capital accumulation rather than, necessarily, investment. The reason for the distinction is that the excessive response of the scrapping margins and intertemporal substitution effects on the creation side may end up dampening actual investment. See Caballero and Hammour (1996a).

The equilibrium implications of incomplete contracts also affect the scrapping decisions of firms. The easiest way to see this is to examine an existing production unit and ask how low its profitability would have to be for it to scrap itself and seek other opportunities. Moreover, assume that neither factor suffers from specificity in this production unit, so that the efficient rule is scrap whenever profitability is less than y. Two, apparently contradictory, features characterize the incomplete contracts economy. First, because the opportunity cost of factors of production is depressed by the excessive allocation to autarky, there is sclerosis; that is, there are units with profitability below y which are not scrapped because the opportunities in autarky are depressed. Second, *given* the depressed level of investment, there is excessive destruction. Since the appropriating factor earns rents in joint production, some of them leave socially valuable production units in order to improve their chances of earning these excess returns.

Caballero and Hammour (1998a,b) argue that, over the long run, capital/labor substitution takes place. If capital is being appropriated, it will seek to exclude labor from joint production by choosing a capital intensive technology. This effect goes beyond purely neoclassical substitution, as it also seeks to reduce the appropriability problem [70].

At a general level, of course, unenforceability of contracts results from the absence of well defined property rights. There is plenty of evidence on the deleterious consequences of such problems for investment. Two recent examples in the literature are Besley (1995) and Hall and Jones (1996). The former provides a careful description of land rights in different regions of Ghana. He documents that an "extra right" over a piece of land increases investment in that land by up to 9 percent in Anloga and up to 28 percent in Wassa [71]. Hall and Jones (1996) use a large cross section of countries to show, among other things, that capital/labor ratios are strongly negatively related to "divertment activities."

5.2.2. Credit constraints

There is by now a large body of evidence supporting the view that credit constraints have substantial effects on firm level investment. Although there are a number of qualifications to specific papers in the literature, the cumulative evidence seems overwhelmingly in favor of the claim that investment is more easily financed with internal than external funds [72]. I will not review this important literature here because there are already several good surveys [73].

[70] We argue that this is a plausible factor behind the large increase in capital/labor ratios in Europe relative to the USA.

[71] Rights to sell, to rent, to bequeath, to pledge, to mortgage, etc.

[72] For a dissenting view, see e.g. Kaplan and Zingales (1997) and Cummins, Hassett and Oliner (1996b).

[73] See e.g. Bernanke et al. (1996, 1999) and Hubbard (1995) for recent ones.

While there are extensive empirical and theoretical microeconomic literatures, the macroeconomics literature on credit constraints is less developed. Notable exceptions are: Bernanke and Gertler (1989, 1990), Kiyotaki and Moore (1997) and Greenwald and Stiglitz (1993)[74]. Although the exact mechanisms are not always the same, many of the aggregate insights of this literature can be described in terms of the results in the preceding subsections.

Changing slightly the interpretation of factor 2, from labor to entrepreneurs, allows us to use Figure 5.1 to characterize credit constraints. Rationing in the labor market becomes rationing of credit available to projects. To the left of point B, which is the region analyzed in the literature, net investment is too responsive to shocks; there is more credit rationing as the state of the economy declines; and there is underinvestment in general.

Internal funds and collateralizable assets reduce the extent of the appropriability problem by playing the role of a bond, and introduce heterogeneity and therefore ranking of entrepreneurs. Since the value of collateral is likely to decline during a recession, there is an additional amplification effect due to the decline in the feasibility of remedial "bonding"[75].

6. Conclusion and outlook

This survey started by arguing that the long run relationship between aggregate capital, output and the cost of capital is not very far from what is implied by the basic neoclassical model: in the US, the elasticity of the capital-output ratio with respect to permanent changes in the cost of capital is close to minus one.

In the short run things are more complex. Natural-experiments have shown that, in the cross section, the elasticity of investment with respect to changes in investment tax credits is much larger than we once suspected.

How to go from these microeconomic estimates to aggregates, and to the response of investment to other types of shocks is not fully resolved. We do know, however, that these estimates represent expected values of what seems to be a very skewed distribution of adjustments. A substantial fraction of a firm's investment is bunched into infrequent and lumpy episodes. Aggregate investment is heavily influenced by the degree of synchronization of microeconomic investment spikes. For US manufacturing, the short run (annual) elasticity of investment with respect to changes in the cost of capital is less than one tenth the long run response when the economy has had a depressed immediate history, while this elasticity can rise by over 50 percent when the economy is undergoing a sustained expansion.

[74] Also see Gross (1994) for empirical evidence and a model integrating financial constraints and irreversibility.

[75] See e.g. Kiyotaki and Moore (1997).

Still, the mapping from microeconomics to aggregate investment dynamics – especially *equilibrium* aggregate investment dynamics – is probably more complex than just the direct aggregation of very non-linear investment patterns. Informational problems lead to a series of strategic delays which feed into and feed off of the natural inaction of lumpy adjustment models. This process has the potential to exacerbate significantly the time varying nature of the elasticity of aggregate investment with respect to aggregate shocks.

Moreover, sunk costs provide fertile ground for opportunistic behavior. In the absence of complete contracts, aggregate net investment is likely to become excessively volatile. The lack of response of equilibrium payments to complementary – and otherwise inelastic – factors (e.g. workers), exacerbates the effects of shocks experienced by firms. Also, the withdrawal of financiers' support during recessions further reduces investment. Thus, capital investment seems to be hurt at both ends: workers that do not share fairly during downturns, and financiers that want to limit their exposure to potential appropriations from entrepreneurs which cannot credibly commit not to do so during the recovery.

The last two themes, equilibrium outcomes with informational problems and opportunism, are wanting for empirical work. I therefore suspect that we will see plenty of research filling this void in the near future.

References

Abel, A.B. (1979), Investment and the Value of Capital (Garland, New York).

Abel, A.B., and J.C. Eberly (1994), "A unified model of investment under uncertainty", American Economic Review 84(December):1369–1384.

Abel, A.B., and J.C. Eberly (1996a), "Investment and q with fixed costs: an empirical analysis", mimeograph (Wharton School, University of Pennsylvania, January).

Abel, A.B., and J.C. Eberly (1996b), "The effects of irreversibility and uncertainty on capital accumulation", mimeograph (Wharton School, University of Pennsylvania, May).

Abel, A.B., A.K. Dixit, J.C. Eberly and R.S. Pindyck (1996), "Options, the value of capital, and investment", Quarterly Journal of Economics 111(3, August):753–777.

Acemoglu, D. (1996), "A microfoundation for social increasing returns in human capital accumulation", Quarterly Journal of Economics 111(3, August):779–804.

Atkeson, A., and P.J. Kehoe (1997), "Models of energy use: putty-putty vs. Putty-clay", Federal Reserve Bank of Minneapolis, Research Department, Staff Report 230 (March).

Auerbach, A.J., and K.A. Hassett (1992), "Tax policy and business fixed investment in the United States", Journal of Public Economics 47(2):141–170.

Barnett, S., and P. Sakellaris (1995), "Non-linear response of firm investment to Q: testing a model of convex and non-convex adjustment costs", mimeograph (University of Maryland, August).

Benhabib, J., and A. Rustichini (1993), "A vintage model of investment and growth: theory and evidence", in: R. Becker, ed., General Equilibrium, Growth and Trade, vol. II (Academic Press, New York).

Bernanke, B.S., and M. Gertler (1989), "Agency costs, net worth, and business fluctuations", American Economic Review 95:14–31.

Bernanke, B.S., and M. Gertler (1990), "Financial fragility and economic performance", Quarterly Journal of Economics 105:87–114.

Bernanke, B.S., M. Gertler and S. Gilchrist (1996), "The financial accelerator and the flight to quality", Review of Economics and Statistics 78:1–15.
Bernanke, B.S., M. Gertler and S. Gilchrist (1999), "The financial accelerator in a quantitative business cycle framework", ch. 21, this Handbook.
Bertola, G. (1988), "Adjustment costs and dynamic factor demands: investment and employment under uncertainty", Ph.D. Dissertation (MIT, Cambridge, MA).
Bertola, G. (1992), "Labor turnover costs and average labor demand", Journal of Labor Economics 10(4):389–411.
Bertola, G., and R.J. Caballero (1990), "Kinked adjustment costs and aggregate dynamics", in: O.J. Blanchard and S. Fischer, eds., NBER Macroeconomics Annual (MIT Press, Cambridge, MA) 237–288.
Bertola, G., and R.J. Caballero (1994), "Irreversibility and aggregate investment", Review of Economic Studies 61(2, April):223–246.
Besley, T. (1995), "Property rights and investment incentives: theory and evidence from Ghana", Journal of Political Economy 103(5, October):903–937.
Blanchard, O.J. (1986), "Investment, output, and the cost of capital: a comment", Brookings Papers on Economic Activity 1986(1):153–158.
Blanchard, O.J., and M. Kremer (1996), "Disorganization", mimeograph (MIT, October).
Brainard, W.C., and J. Tobin (1968), "Pitfalls in financial model building", American Economic Review 58(May):99–122.
Caballero, R.J. (1993a), "On the dynamics of aggregate investment", in: L. Serven and A. Solimano, eds., Striving for Growth After Adjustment, The Role of Capital Formation (The World Bank, Washington, DC) 81–106.
Caballero, R.J. (1993b), "Durable goods: an explanation for their slow adjustment", Journal of Political Economy 101(2, April):351–384.
Caballero, R.J. (1994a), "Small sample bias and adjustment costs", Review of Economics and Statistics 76(1, February):52–58.
Caballero, R.J. (1994b), "A reconsideration of investment behavior using tax reforms as natural experiments: a comment", Brookings Papers on Economic Activity 1994(2):62–68.
Caballero, R.J., and E. Engel (1999), "Explaining investment dynamics in U.S. manufacturing: a generalized (S, s) approach", Econometrica 67(4, July).
Caballero, R.J., and M.L. Hammour (1994), "The cleansing effect of recessions", American Economic Review 84(5, December):1350–1368.
Caballero, R.J., and M.L. Hammour (1996a), "On the timing and efficiency of creative destruction", Quarterly Journal of Economics 111(3, August):805–852.
Caballero, R.J., and M.L. Hammour (1996b), "On the ills of adjustment", Journal of Development Economics 51(1, October):161–192.
Caballero, R.J., and M.L. Hammour (1997b), "Improper churn: financial constraints and factor markets", mimeograph (MIT, May).
Caballero, R.J., and M.L. Hammour (1998a), "The macroeconomics of specificity", Journal of Political Economy 106(4, August):724–767.
Caballero, R.J., and M.L. Hammour (1998b), "Jobless growth: appropriability, factor substitution, and unemployment", Carnegie-Rochester Conference Series on Public Policy 48(June):51–94.
Caballero, R.J., and J. Leahy (1996), "Fixed costs: the demise of marginal q", Working paper No. 5508 (NBER, March).
Caballero, R.J., and R.S. Pindyck (1996), "Uncertainty, investment, and industry evolution", International Economic Review 37(3, August):641–662.
Caballero, R.J., E. Engel and J. Haltiwanger (1995), "Plant-level adjustment and aggregate investment dynamics", Brookings Papers on Economic Activity 1995(2):1–54.
Campbell, J.R. (1997), "Entry, exit, embodied technology, and business cycles", Working Paper No. 5955 (NBER, March).

Campbell, J.R., and J.D.M. Fisher (1996), "Aggregate employment fluctuations with microeconomic asymmetries", mimeograph (University of Rochester, August).
Caplin, A., and J. Leahy (1993), "Sectoral shocks, learning, and aggregate fluctuations", Review of Economic Studies 60(4, October):777–794.
Caplin, A., and J. Leahy (1994), "Business as usual, market crashes and wisdom after the fact", American Economic Review 84(3, June):549–565.
Chamley, C., and D. Gale (1994), "Information revelation and strategic delay", Econometrica 62: 1065–1085.
Chirinko, R.S. (1993), "Business fixed investment spending: a critical survey of modelling strategies, empirical results, and policy implications", Journal of Economic Literature 31(December):1875–1911.
Clark, J.M. (1917), "Business acceleration and the law of demand: a technical factor in economic cycles", Journal of Political Economy 25(March):217–235.
Clark, J.M. (1944), "Additional note on business acceleration and the law of demand", in: American Economic Association, Readings in Business Cycle Theory (Blackiston Company, Philadelphia, PA).
Cooley, T.F., J. Greenwood and M. Yorukoglu (1997), "The replacement problem", Working paper No. 444 (Rochester Center for Economic Research, August).
Cooper, R., J. Haltiwanger and L. Power (1994), "Machine replacement and the business cycle: lumps and bumps", mimeograph (Boston University).
Cummins, J.G., K.A. Hassett and R.G. Hubbard (1994), "A reconsideration of investment behavior using tax reforms as natural experiments", Brookings Papers on Economic Activity 1994(2):1–59.
Cummins, J.G., K.A. Hassett and R.G. Hubbard (1996a), "Tax reforms and investment: a cross country comparison", Journal of Public Economics 62(1/2, October):237–273.
Cummins, J.G., K.A. Hassett and S.D. Oliner (1996b), "Investment behavior, internal funds, and observable expectations", mimeograph (New York University, October).
Dixit, A. (1989), "Entry and exit decisions under uncertainty", Journal of Political Economy 97:620–638.
Dixit, A. (1993), The Art of Smooth Pasting (Harwood Academic Publishers, Langhorns, PA).
Doms, M., and T. Dunne (1993), "An investigation into capital and labor adjustment at the plant level", mimeograph (Center for Economic Studies, Census Bureau).
Eisner, R. (1969), "Tax policy and investment behavior: a comment", American Economic Review 59(June):379–388.
El Karoui, N., and I. Karatzas (1991), "A new approach to the Skorohod problem, and its applications", Stochastics 34:57–82.
Fallick, B.C., and K.A. Hassett (1996), "Investment and union certification", Discussion paper No. 1996–43 (FED, November).
Fazzari, S.M., R.G. Hubbard and B.C. Petersen (1988), "Financing constraints and corporate investment", Brookings Papers on Economic Activity 1988(1):141–195.
Gale, D. (1995), "What have we learned from social learning?", mimeograph (Boston University, August).
Gilchrist, S., and J.C. Williams (1996), "Putty-clay and investment: a business cycle analysis", mimeograph (Boston University, May).
Goolsbee, A. (1995a), "Investment tax incentives and the price of capital goods", mimeograph (Chicago GSB).
Goolsbee, A. (1995b), "Factor prices and the retirement of capital goods", mimeograph (Chicago GSB, July).
Goolsbee, A. (1997), "The incidence of investment tax subsidies: to the workers go the spoils?", mimeograph (Chicago GSB, February).
Goolsbee, A., and D.B. Gross (1997), "Estimating adjustment costs with data on heterogeneous capital goods", mimeograph (Chicago GSB, September).
Greenspan, A., and D. Cohen (1996), "Motor vehicles stocks, scrappage, and sales", Working paper No. 1996–40 (FED, October).

Greenwald, B., and J. Stiglitz (1993), "Financial market imperfections and business cycles", Quarterly Journal of Economics 108(1, February):77–114.
Gross, D. (1994), "The investment and financing decisions of liquidity constrained firms", mimeograph (Chicago GSB).
Grout, P.A. (1984), "Investment and wages in the absence of binding contracts: a Nash bargaining approach", Econometrica 52(2, March):449–460.
Hall, R.E., and C.I. Jones (1996), "The productivity of nations", mimeograph (Stanford University, August).
Hall, R.E., and D.W. Jorgenson (1967), "Tax policy and investment behavior", American Economic Review 57(3, June):391–414.
Hart, O. (1995), Firms, Contracts and Financial Structure, Clarendon Lectures in Economics (Oxford University Press, Oxford).
Hassett, K.A., and R.G. Hubbard (1996a), "New evidence concerning the openness of the world market for capital goods", mimeograph (Board of Governors of the Federal Reserve System, June).
Hassett, K.A., and R.G. Hubbard (1996b), "Tax policy and investment", Working Paper No. 5683 (NBER, July).
Hayashi, F. (1982), "Tobin's marginal Q and average Q: a neoclassical interpretation", Econometrica 50(1, January):213–224.
Hopenhayn, H.A. (1992), "Entry, exit, and firm dynamics in long run equilibrium", Econometrica 60:1127–1150.
Hubbard, R.G. (1995), "Capital-market imperfections and investment", mimeograph (Columbia University).
Jorgenson, D.W. (1963), "Capital theory and investment behavior", American Economic Review 53(2, May):247–259.
Jovanovic, B. (1997), "Obsolescence of capital", mimeograph (New York University, February).
Kaplan, S.N., and L. Zingales (1997), "Do investment-cash flow sensitivities provide useful measures of financing constraints?, Quarterly Journal of Economics 112(1, February):169–216.
Kiyotaki, N., and J. Moore (1997), "Credit cycles", Journal of Political Economy 105(2, April):211–248.
Koyck, L.M. (1954), Distributed Lags and Investment Analysis (North-Holland, Amsterdam).
Krieger, S. (1997), "The general equilibrium dynamics of investment, scrapping and reorganization in an economy with firm level uncertainty", mimeograph (Chicago, July).
Leahy, J. (1993), "Investment in competitive equilibrium: the optimality of myopic behavior", Quarterly Journal of Economics 108:1105–1133.
Meyer, J.R., and E. Kuh (1957), The Investment Decision: An Empirical Study (Harvard University Press, Cambridge, MA).
Nickell, S.J. (1978), The Investment Decisions of Firms (Cambridge University Press, Oxford).
Oliner, S.D., G.D. Rudebusch and D. Sichel (1995), "New and old models of business investment: a comparison of forecasting performance", Journal of Money, Credit and Banking 27:806–826.
Pindyck, R.S. (1988), "Irreversible investment, capacity choice, and the value of the firm", American Economic Review 78(5, December):969–985.
Ramey, G., and J. Watson (1996), "Contractual fragility, job destruction and business cycles", mimeograph (University of California at San Diego, June).
Rivers, D., and Q.H. Vuong (1991), "Model selection tests for nonlinear dynamic models", Working paper No. 91–08 (University of Toulouse).
Rotemberg, J.J. (1987), "The new Keynesian microfoundations", in: O.J. Blanchard and S. Fischer, eds., NBER Macroeconomics Annual 1987 (MIT Press, Cambridge, MA) 69–104.
Rothschild, M. (1971), "On the cost of adjustment", Quarterly Journal of Economics 85(November): 605–622.
Salter, W.E.G. (1960), Productivity and Technical Change (Cambridge University Press, Cambridge).
Shapiro, M.D. (1986), "Investment, output, and the cost of capital", Brookings Papers of Economic Activity 1986(1):111–152.

Simons, H.C. (1944), "Some reflections on syndicalism", Journal of Political Economy 52:1–25.
Stock, J.H., and M.W. Watson (1993), "A simple MLE of cointegrating vectors in higher order integrated systems", Econometrica 61(4, July):783–820.
Tinbergen, J. (1939), "A method and its application to investment activity", in: Statistical Testing of Business Cycle Theories, vol. 1 (Economic Intelligence Service, Agathon Press, New York).
Tobin, J. (1969), "A general equilibrium approach to monetary theory", Journal of Money, Credit and Banking 1:15–29.
Williamson, O.E. (1979), "Transaction-cost economics: the governance of contractual relations", Journal of Law and Economics 22(2, October):233–261.
Williamson, O.E. (1985), The Economic Institutions of Capitalism (Free Press, New York).