Notes for Econ202A: Investment

Pierre-Olivier Gourinchas
UC Berkeley

Fall 2015

Disclaimer: These notes are riddled with inconsistencies, typos and omissions. Use at your own peril. Many thanks to Sergii Meleshchuk for spotting and removing many of them!

Contents

1 Introduction

Investment is important for macroeconomics:

- Investment determines the stock of capital. It increases the productive capacity of the economy, and therefore future standard of living
- Volatility of investment is high at business cycle frequencies. Therefore, investment matters a lot for business cycle fluctuations.

2 Investment and the Cost of Capital

2.1 The demand for capital

Denote r_K the rental rate of capital. Suppose we can write the firm's profits, after we optimize over *other* inputs (such as labor, intermediates, materials etc...) as $\Pi(K, X)$ where X denotes various other shifters of profits (such as cost of other inputs etc...) The firm maximizes profits, taking X as given, i.e.:

$$\max_K \Pi(K, X) - r_K K$$

The first order condition for the demand of capital is:

$$\Pi_K(K, X) = r_K$$

If the profit function exhibits diminishing returns to capital, and the usual Inada conditions, then the schedule $\Pi_K(.)$ is decreasing in K and there is a unique K that solves the above equation.

2.2 The User Cost of Capital

Most capital is not rented. How should we think of the rental rate r_K in a world where firms own capital? This is what the user cost of capital literature attempts to do. Consider a firm that purchases capital at price p_K. Capital depreciates at rate δ. The firm faces the following intertemporal problem (taking the entire sequence $\{X_t\}$ as given):

$$V(K_t) = \max_{I_t} \int_t^\infty e^{-\int_t^s r_u du}(\Pi(K_s, X_s) - p_{K,s} I_s) ds$$

where the law of motion of capital is:

$$\dot{K}_t = I_t - \delta K_t$$

and r_t is the risk free rate at time t.[1] We can solve this problem using the Maximum Principle we studied in class. Define the current period Hamiltonian:

$$\mathcal{H}_t = (\Pi(K_t, X_t) - p_{K,t} I_t) + \lambda_t (I_t - \delta K_t)$$

[1] This implicitly assumes that the owner of the firm is risk neutral. Otherwise, we would want to discount profits using the *stochastic discount factor* of the firm's owner.

According to the Maximum Principle, the optimality conditions are:

$$\begin{aligned} p_{K,t} &= \lambda_t \\ \Pi_K(K_t, X_t) - \delta\lambda_t &= r_t\lambda_t - \dot{\lambda}_t \\ \lim_{t\to\infty} K_t\lambda_t e^{-\int_0^t r_u du} &\leq 0 \end{aligned}$$

Combining these conditions, we obtain:

$$\begin{aligned} \Pi_K(K_t) &= (r_t + \delta)p_{K,t} - \dot{p}_{K,t} \\ \lim_{t\to\infty} K_t p_{K,t} e^{-\int_0^t r_u du} &= 0 \end{aligned}$$

By comparing the first order condition of the rental model with the condition above, this defines the user cost of capital $r_{K,t}$ as:

$$r_{K,t} = (r_t + \delta - \dot{p}_{K,t}/p_{K,t})p_{K,t} \tag{1}$$

Interpretation: The user cost of capital

- increases with the interest rate r_t (opportunity cost of investing p_K)
- increases with the depreciation rate of capital (δ)
- decreases with the increase in the price of capital goods (capital gain)

The user cost model is helpful to evaluate the effect of tax policies (see Hall and Jorgenson (1967)). But it is not very helpful to evaluate the dynamics of investment for two reasons:

- the model determines the stock of capital. Therefore any change in e.g. the user cost of capital would require an infinite investment rate as the stock of capital would 'jump' to its new level.
- Second, because the model does allow capital to 'jump', it means that decisions about the capital stock become static: they are determined by the current cost of capital, and are not forward looking

What is needed is something that slows down the adjustment of the capital stock in response to changes in the environment. The adjustment costs can be internal (e.g. firms face direct costs of adjusting their capital stock) or external (e.g. firms do not face costs of adjusting their stock of capital but face a higher price of capital goods).

3 A Model with Adjustment Costs

Consider the firm's problem, as before, but now assume that there are adjustment costs to capital. Specifically, if the firm wants to increase its capital stock by I_t units at time t at price $p_{K,t}$, it must purchase $I_t(1 + C(I_t, K_t))$ units of capital. $C(.)$ is the percentage increase in cost to install one unit of capital. We assume that it can potentially depend on the level of investment and the level of capital and satisfies the following assumptions:

$$C(I, K) \geq 0 \quad ; \quad C_{II} > 0 \quad ; \quad C_K < 0 \quad ; \quad C(0, K) = C'(0, K) = 0$$

That is, the adjustment cost is convex in investment. The fact that $C(0, K) = C'(0, K) = 0$ is important. It implies that the firm does not face much of an adjustment cost when it keeps investment infinitesimal. Hence, firms will respond by adjusting investment continuously and smoothly. We will see later models where firms face different forms of adjustment costs and, as a result, adjust their capital stock infrequently and in a lumpy way.

Example 1 *Examples of adjustment cost functions.*

- $C(I, K) = C(I)$ *if the adjustment costs does not depend on the level of capital;*
- $C(I, K) = D(I/K)$ *with* D *convex, if the adjustment cost depends on the ratio of investment to capital. That last formulation implies that the adjustment cost 'scales up' with the level of capital.*

3.1 The Hamiltonian

The firm problem becomes:

$$V(K_t) = \max_{I_t} \int_t^{\infty} e^{-\int_t^s r_u du}(\Pi(K_s, X_s) - p_{K,s}I_s(1 + C(I_s, K_s))ds$$

subject to the constraint:

$$\dot{K}_t = I_t - \delta K_t$$

As before, we can set-up the current value Hamiltonian:

$$\mathcal{H}(I_t, \lambda_t) = \Pi(K_t, X_t) - p_{K,t}I_t(1 + C(I_t, K_t)) + \lambda_t(I_t - \delta K_t)$$

The optimality conditions are:

$$p_{K,t}\left[1 + C(I_t, K_t) + I_t C_I(I_t, K_t)\right] = \lambda_t \quad \text{(2a)}$$

$$\Pi_K(K_t, X_t) - p_{K,t}I_t C_K(I_t, K_t) - \lambda_t\delta = r_t\lambda_t - \dot{\lambda}_t \quad \text{(2b)}$$

$$\lim_{t\to\infty} K_t\lambda_t e^{-\int_0^t r_u du} \leq 0 \quad \text{(2c)}$$

Consider the first equation. It is not the case anymore that the co-state variable λ_t equals the price of capital goods. The firm equates the value of one additional machine (λ_t) to the cost of an additional machine (the term on the right hand side), which includes the marginal effect of adjustment costs. Note in particular that the firm internalizes that adding one machine will also change the cost *per machine* for all existing machines purchased (this is the term in C_I).

This first equation can be expressed as:

$$\frac{\lambda_t}{p_{K,t}} = 1 + C(I_t, K_t) + I_t C_I(I_t, K_t) \tag{3}$$

and inverted to yield:

$$I_t = \phi(\frac{\lambda_t}{p_{K,t}}, K_t) \tag{4}$$

This determines an investment schedule. Since C_I is convex, investment is increasing in $\lambda_t / p_{K,t}$. Because this ratio is important, we give it a name: it is Tobin's marginal q, which we denote q_t:

$$q_t \equiv \frac{\lambda_t}{p_{K,t}}$$

Economically it is the ratio of the value of one unit of capital *inside the firm* (λ_t) and the value of one unit of capital *outside the firm* ($p_{K,t}$). Notice that investment is only a function of marginal q and of the level of capital. In particular, the firm does not need to know anything else about future demand etc... to figure out the optimal investment level.

The second equation can be rewritten as:

$$r_t = \frac{\Pi_K(K_t, X_t)}{\lambda_t} - \delta + \frac{\dot{\lambda}_t}{\lambda_t} + \frac{p_{K,t} I_t(-C_K(I_t, K_t))}{\lambda_t} \tag{5}$$

The left hand side is the risk-free interest rate. The right hand side is the return on investing a marginal unit. This return consists of three terms:

- the additional marginal profits generated by the extra unit of capital Π_K / λ_t, adjusted for depreciation ($-\delta$). This can be interpreted as a dividend yield.
- the second term is a capital gain on that unit (the term $\dot{\lambda}_t / \lambda_t$)
- the final term is new: it reflects the fact that adding one unit of capital reduces adjustment costs by C_K on all *inframarginal* units. Since we assumed $C_K < 0$, this term increases the return to capital.

Note that this expression can be re-arranged to give the 'user cost of capital' i.e. the rental rate that the firm would be willing to pay for this marginal unit of capital:

$$r_{K,t} = \Pi_K(K_t, X_t) = (r_t + \delta)\lambda_t - \dot{\lambda}_t + p_{K,t} I_t C_K(I_t, K_t)$$

Compared to the simple frictionless capital model, the user cost of capital features:

(a) a different value of capital (i.e. λ is potentially different from p_K);

(b) an additional term related to the savings on adjustment-costs as capital increases (the term in C_K).

We can integrate by parts the previous equation between time 0 and time T to obtain (this is a good exercise, make sure you know how to do it):

$$[\lambda_t e^{-\int_0^t (r_s+\delta)ds}]_0^T = \int_0^T (p_{K,t} I_t C_K(I_t, K_t) - \Pi_K(K_t, X_t)) e^{-\int_0^t (r_s+\delta)ds} dt$$

Now, we know from the TVC condition, that $\lim_{t\to\infty} K_t \lambda_t e^{-\int_0^t r_u du} = 0$. It must follow that $\lim_{t\to\infty} \lambda_t e^{-\int_0^t (r_u+\delta)du} = 0$.[2] Taking the limit as $T \to \infty$, yields:

$$\lambda_t = \int_t^\infty \left[\Pi_K(K_s, X_s) - p_{K,s} I_s C_K(I_s, K_s)\right] e^{-\int_t^s (r_u+\delta)du} ds$$

In other words, the marginal value of installed capital is given by the present discounted value of future marginal profits, adjusted for the dilution effect of capital on adjustment costs. The important point here is that Tobin's marginal q incorporates expectations about future profits. In the q-theory of investment, investment depends on expectations of future profitability of capital. q could be high (and therefore the firm could decide to invest) even if marginal profitability is currently low.

Example 2 *Consider the case where $C(I, K) = D(I/K)$ and assume that $D(0) = 0$, $D'(0) = 0$ and $D'' > 0$. Then the first equation yields:*

$$\frac{\lambda_t}{p_{K,t}} = q_t = 1 + D(I_t/K_t) + I_t/K_t D'(I_t/K_t)$$

which can be inverted to yield:

$$I_t/K_t = \phi(\lambda_t/p_{K,t})$$

with $\phi'(.) > 0$.

[2] To see this, observe that if this second condition were violated, then λ_t must tend to ∞. But from the first order condition, this requires that investment tends to infinity too and therefore capital tends to infinity as well therefore the TVC must fail too.

3.2 Marginal and Average q

q represents the increase (at the margin) in the firm's value from investing one more unit of capital. In practice, marginal q is difficult to measure. An easier measure is Tobin's average q, denoted Q and defined as the ratio of the market value of the firm to the replacement cost of its capital, that is:

$$Q_t = \frac{V(K_t)}{p_{K,t}K_t}$$

In general, average and marginal q may be quite different. However, Hayashi (1982) shows that the two are equal when:

1. $\Pi_{KK} = 0$, as obtains when there are constant returns to scale and factor markets are competitive.
2. $C(I, K)$ is homogenous of degree 0 in I, K, i.e. $C(\mu I, \mu K) = C(I, K)$. This is satisfied if $C(I, K) = D(I/K)$.
3. V is the PDV of cash flows (i.e. no bubbles, fads etc...)
4. There are no taxes

Hayashi (1982) also shows that if there are taxes, then:

$$\frac{I_t}{K_t} = \phi(\frac{q_t}{(1-\tau)(1-uD)})$$

where τ is the investment tax credit, D is the present value of depreciation allowances: $D = \int_0^\infty D(v)e^{-rv}dv$ where $D(v)$ is the allowed depreciation schedule for an asset of age v, and u is the profit tax.

With taxes, the relationship between average and marginal q is:

$$Q = q + \frac{A_0}{p_K K}$$

where $A_0 = u\int_{-\infty}^0 \left(\int_0^\infty D(v-s)e^{-rv}dv\right) I_s p_{Ks} ds$ is the present discounted value of current and future tax deductions attributable to past investments. It is not a decision variable (since it comes from investments before $t = 0$,) but it still affects the value of the firm.

The analysis shows the limits of using average Q instead of marginal q:

1. if the firm has market power (so that $\Pi_{KK} < 0$)
2. if V is different from the PDV of cash flows: the market does not value firms at their fundamental value. In that case, the firm can either:
   - ignore the market signals and invest based on the fundamental value;
   - if V is high, the market is the right place to fund investment (issue shares).

3.3 The Dynamics of the Model

To simplify things a bit (without any impact on the economic interpretation), let's assume that:

(a) the interest rate is constant and equal to r;

(b) the price of capital goods is constant and equal to 1, so that $q_t = \lambda_t$;

(c) the adjustment costs are homogenous in investment and capital: $C(I, K) = D(I/K)$;

The model can be summarized by the following equations:

$$\dot{K}_t = I_t - \delta K_t = (\phi(q_t) - \delta)K_t \tag{6a}$$

$$\dot{q}_t = (r + \delta)q_t - \Pi_K(K_t, X_t) - \Psi(q_t) \tag{6b}$$

where $\Psi(q_t) = -I_t C_K(I_t, K_t) = (I_t/K_t)^2 D'(I_t/K_t) = \phi(q_t)^2 D'(\phi(q_t))$. Observe that $\phi(1) = 0$ and $\Psi(1) = \Psi'(1) = 0$.

The first equation is the capital accumulation equation, where we substituted the fact that $I_t = \phi(q_t)K_t$; the second equation is the law of motion of $q_t = \lambda_t$ from the Maximum Principle. One of the variables, capital, is 'pre-determined' by historical conditions and cannot jump. The other, Tobin's q, is a 'jump' variable.

This system of two equations can be represented in a phase diagram. Let's analyze the two locii corresponding to $\dot{K} = 0$ and $\dot{q} = 0$.

1. Steady state capital stock. This locus corresponds to $\dot{K} = 0$. Substituting into (6a), we obtain:

   $$\phi(\bar{q}) = \delta$$

   Since $\phi'(q) > 0$, $\phi(1) = 0$ and $\delta > 0$, this implies that $\bar{q} > 1$. Observe that the value of q is such that $I = \delta K$, as expected in steady state. To establish the dynamics of K, observe that an increase in q above the $\dot{K} = 0$ schedule increases $\phi(q)$ so that $\dot{K} > 0$.

2. The second locus is given by (assuming that the variables X are constant too)

   $$\Pi_K(K, X) = (r + \delta)q - \Psi(q)$$

   This equation yields a relationship between K and q along which the marginal value of capital is constant. For q close to 1, we have $\Psi'(q)$ close to 0 and therefore the slope of that schedule is downward sloping.[3] To establish the dynamics, observe that an increase in K increases $\dot{q}$ since $\Pi_{KK} < 0$.

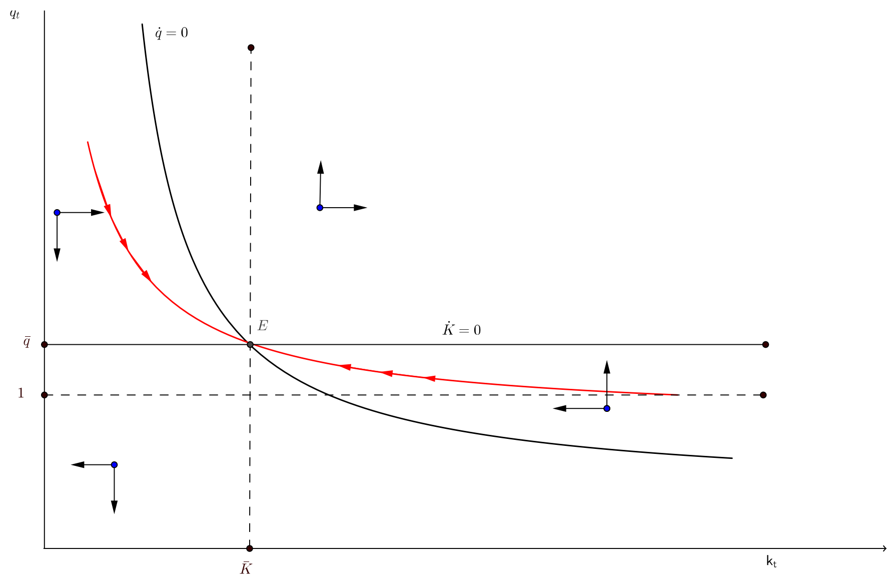

Figure 1: The dynamics of the model with adjustment costs

The dynamics are 'saddle-path stable.'[4] The only possible solution, for any given initial K_0, is for the marginal value of capital q_0 to 'jump' immediately to the saddle path that will converge to the steady state $(\bar{K}, \bar{q})$.

3.4 The Steady State

The steady state is characterized by the following conditions:

$$\begin{aligned} \phi(\bar{q}) &= \delta \\ \Pi_K(\bar{K}, X) &= r + \delta - \Psi(\bar{q}) = r + \delta - \delta^2 D'(\delta) \end{aligned}$$

The last term on the last equation represents the additional benefit that arises from investing in capital, i.e the dilution of adjustment costs. This term disappears in the case where $C_K = 0$.

The first equation indicates that Tobin's q steady state value exceeds unity because of depreciation. (You can check that $\bar{q} = 1$ if $\delta = 0$). This implies that the marginal value of capital exceeds its replacement value.

[3] To check this, take a full derivative to obtain: $\Pi_{KK} dK = [r + \delta - \Psi'(q)]dq$. The term on the right hand side is positive if $\Psi'(q) < r + \delta$ which will be the case for q close to 1. Since $\Pi_{KK} < 0$ this ensures the schedule is downward sloping. You can check that this is always the case if $C_K = 0$, i.e. there are no scale effects from capital. You can check that the system remains saddle path stable even if the $\dot{q} = 0$ schedule is upwards sloping.

[4] Technically, this means that the system has one root inside and one root outside the unit circle.

3.5 Using the model to explore the effect of shocks

First a general observation on what we mean by shocks here. The model was derived under the assumption that all the parameters are either constant or that their fluctuations are known ahead of time (e.g. the X_t). We now consider what happens if there is a sudden change in this environment.

If it seems a bit bizarre to you that we're allowing a change in the model that firms have never anticipated, it's because it is! There are ways to finesse this (for instance by assuming that these sort of shocks are both infrequent and small so that it is optimal for firms to discard them when solving for their optimal investment policy.[5] But if we follow the logic to its end, it means that the model cannot be used to tell us really about the real world where (a) business cycle fluctuations are not that infrequent and (b) are not necessarily that small.

Nevertheless, these 'phase diagram' are stock-full of economic intuition, so it is interesting to see what happens nonetheless. What this means is that these are not useful models to conduct any serious calibration and real world counterfactuals. But they will tell you a lot about the forces that drive firm's responses to changes in their environment.

Partly as a result of the 'perfect foresight' model's reliance on totally unanticipated shocks that will never happen again but just happened the literature has moved to models that encompass the stochastic structure of the environment in which firms operate. In these environments, firms know that changes may occur. They have rational expectations about these changes, in the sense that the sort of shocks that can occur are in the support of their beliefs about just such changes. In this sort of environment, firms adjust their behavior to take the associated risks into account. We will see models of that kind in the next class when we look at what happens if there are non-convex adjustment costs to capital. In these models, we can trace how the economy responds to a particular realization of a shock. Although the possibility of a shock is rationally anticipated by economic actors, they are still surprised by its realization, just like the fact that you know a recession may happen at anytime does not mean that you would not be surprised if one happened tomorrow. You will see models of this sort in the spring with Yuriy Gorodnichenko.

3.5.1 An unexpected permanent increase in demand

Consider the effect of a permanent unexpected increase in demand. This can be represented by a change in one of the shifters X in the profit function. Assume that for a given level of capital and production, the increased demand increases marginal profits. The resulting increase in Π_K shifts the $\dot{q}$ schedule to the right (why?).

At $t = 0$ (when the shock occurs), the economy is not on the new saddle-path. This requires an immediate jump in q: because profits are going to be higher in the future, the value of installed capital increases. This triggers an increase in investment and, over time, an

[5]The shocks need to be small because otherwise the uncertainty may cause firms to alter their behavior.

increase in capital.

Notice that while investment jumps, it remains finite and K itself does not jump. Finally, the increase in investment is highest immediately after the shock. Gradually, q returns to its steady state value, and as it does, so does investment. This is what is called an accelerator theory of investment: it responds to changes in output, not the level of output per se.

3.5.2 An unexpected transitory increase in demand

Consider the same thought experiment as above, but now the increase in demand is temporary, and will revert back at some time $T > 0$. The firm learns about the increase in demand and of their duration at time 0.

How can we find the dynamic path of the economy? The answer is that there cannot be a jump in q at time T. Why? Because at time T there is no news, therefore the value of installed capital should not change. Suppose it did, i.e. suppose that q jumped at T for a value of capital $K = K_T$. Recall that this is fully known as of time 0 after the news is announced. Suppose q drops down at T (this might seem plausible since at T the demand and therefore the profits of the firm decline). Then, the firm would prefer to reduce its investment in capital at $t = T$ so the conjectured (K, q) cannot be an equilibrium. Formally, remember that along the optimal path, the marginal value of the firm satisfies:

$$r_t = \frac{\Pi_K(K_t, X) + \Psi(q_t)}{q_t} - \delta + \frac{\dot{q}_t}{q_t}$$

The last term on the right would be infinity if there is a jump in q at T since the numerator is dq/dt and dq would not be infinitesimal. In other words, at that time the capital gain/loss on the marginal value of capital invested would be infinite. If the loss rate is infinite (i.e. q jumps down, it stands to reason that the firm would postpone installing the last unit of capital, to avoid realizing that loss. It follows that the conjectured path cannot be an equilibrium. This implies that the dynamics cannot be on the saddle path of the high-demand system. In fact, the only solution that is an equilibrium requires that the firm reaches the low-demand saddle path precisely at time T, while following the dynamics of the high demand system between $t = 0$ and $t = T$. The only solution is for q to increase less than in the case of a permanent increase in demand. This makes also sense since we know that q represents the PDV of future marginal profits minus the dilution component of adjustment costs. This PDV is lower now since the increase in demand is temporary.

The analysis tells us that even a temporary increase in demand raises investment (but less so than a permanent one). Finally, we note that the dynamic path for q crosses the line $q = \bar{q}$. This tells us that the initial investment will be divested later on: capital will first increase, then decrease its capital stock. However, the stock of capital starts shrinking even before we are back in the low demand system. Why? because firms know it is costly to adjust capital too rapidly, and should start even before demand declines.

3.5.3 An anticipated permanent increase in demand

In that case, for the same reasons as before, there cannot be a jump at T. So the economy cannot remain in steady state. It must be on the path that leads to the new saddle path at T. This means that investment must jump at $t = 0$ and investment must increase. This shows that investment will respond to expectations of higher demand at some point in the future: news or beliefs about future high demand times can be sufficient to trigger a boom in investment, even if current profitability remains unchanged.

3.5.4 Anticipated Temporary Increase in demand

In that case, the increase occurs at $t_1 > 0$ and ends at T. The dynamics are easy to characterize: there is a limited investment boom, followed by a reduction in investment and a return to the original equilibrium.

3.5.5 Effect of Interest rate movements

A permanent decrease in interest rates leaves the $\dot{K}$ schedule unchanged and shifts the $\dot{q} = 0$ schedule to the right (and steepens it). The shift is similar to a permanent increase in output. Note however, that it is the entire path of future interest rates that matters for investment. In other words, it is more likely to be a long term interest rate than a short term one.

3.5.6 Effect of taxes

With an investment tax credit, the equilibrium consists in replacing $p_K = 1$ with $p_K(1-\tau) = (1-\tau)$. The first order condition becomes:

$$\frac{I_t}{K_t} = \phi(\frac{q_t}{1-\tau})$$

From this, it follows that an increase in τ lowers the $\dot{K} = 0$ schedule. If $C_K(.) = 0$, then this is the only effect and q drops: the value of installed capital is 'diluted' by the additional investment, so the value of the marginal projects declines. In the more general case where $C_K \neq 0$, the $\dot{q} = 0$ curve also shifts. It is likely to shift to the right, i.e. there are more after tax profits.

So both a permanent and temporary investment tax credit can boost investment and therefore aggregate demand. Consider the case where $C_K = 0$ (or where the K in C_K refers to aggregate capital and therefore is not taken into account by the firm when investing). The $\dot{q} = 0$ schedule shifts down. The new steady state value would be $q = \bar{q}(1-\tau)$. The tax credit stimulates investment, which lowers the profitability of firms and therefore lowers q.

Now observe that with a temporary investment tax credit q does not fall as much. Therefore, investment is higher than if the tax credit was permanent. Why? because a temporary tax credit creates a strong incentives to firms to invest while the credit is in place. We even have an investment boom as the credit is about to expire (i.e. as the tax credit is about to expire, notice that the optimal path turns up: q increases and so does I).

4 Empirical Evidence on the q model

q-theory makes a very strong prediction: aggregate investment should depend on q only: $I_t = K_t\phi(q_t)$. There is the slight difficulty that we don't observe marginal q, but many people rely on Hayashi's (1982) result to use average Q instead of the marginal one, adjusted for taxes, as discussed above. It is a bit of a risky exercise, because the conditions for marginal and average q to be equated are probably not satisfied (i.e. firms do have some market power, factor markets are not necessarily competitive, and adjustment costs are not necessarily homogenous of degree zero in K and I).

But if we brush asides these considerations, what does the empirical literature show?

- **Summers (1981, Brookings)** assumes a quadratic adjustment costs with constant returns. This yields the following empirical specification:

$$I_t/K_t = c + b(q_t - 1) + \epsilon_t$$

The coefficient b in this regression is the inverse of the constant term in the cost function (i.e. $D(I/K) = 1/(2b)(I/K)$).[6] Figure 2 reports the results of this regression. The benchmark estimate is specification 4-6 (the specifications differ in the number of lags they include on the right hand side and the treatment of autocorrelation of the errors). The results indicate $\hat{b} = 0.031(0.005)$ which is significant, but very low: investment is not very responsive to q. What this implies is that the adjustment costs need to be very high (i.e. $D(I/K) = 1/(2\hat{b})(I/K) = 16(I/K)$). This implies that if $I/K = 0.2$ then the average investment cost per unit of capital $ID(I/K)/K = 16(0.2)^2 = 64\%$ a very large number. This very low $\hat{b}$ may be the result of (a) measurement error on q which attenuates the estimates, or (b) the result of –for instance– omitted variable bias. Suppose, for instance, that times of high investment demand increase interest rates. This would lead to a lower q since it is the PDV of future marginal profits; (c) the quadratic model of investment costs is not the right one!

- **Cummmins, Hassett and Hubbard (1994, Brookings)** instrument q using changes in the tax code. The idea is that changes in taxes can have large effects on a firms valuation and will differ across industries depending on capital intensity. So using changes in the tax code, they estimate a $\hat{b}$ close to 0.5 on firm level data (Compustat), which implies that the adjustment costs are more reasonable, around 4% of capital. However, it is unclear how much this result carries over to aggregate investment: (a) to the extent that the supply of investment goods is not infinitely elastic, the effect of an increased demand for capital may be mostly to raise the price of investment goods. This is what Goolsbee (1998) finds in a very nice paper. If so, this suggests that the component of adjustment costs that matters may be external, i.e. related to the price response of investment goods; (b) the R^2 of the regressions are quite low, i.e. q still explains a small fraction of investment at the firm level. In fact, the R^2 increase

[6] Notice the the cost is $ID(I/K)$ so it is quadratic in investment, as needed.

significantly once we add cash flow or other current variables (current profits, current sales) as a right hand side variable, the fit improves markedly.

- **Fazzari, Hubbard and Petersen (1988, Brookings)** Models with some forms of financial friction imply that internal funds are cheaper than external funds, i.e. firms will tend to rely on retained earnings to fund investment before they turn to external funds (bonds, loans or equity). If that is the case, perhaps it is not surprising that investment increases with higher cash flow or retained earnings. The problem with a simple regression of investment on cash flow is that cash flow may contain information about future profitability. This is likely to be true both in the cross section. The idea of FHP is similar to that of Zeldes for households: split the sample into firms that are likely to be constrained and firms that are likely to be unconstrained. If cash flow is a proxy for profitability, it should matter for both groups identically. But if financial frictions are important, the first group should be more sensitive to cash flows. FHP divide firms based on the size of dividends distributed (i.e. distributed earnings vs. retained earnings). The coefficient on cash flow is 0.230 (0.010) for the high dividend firms and 0.461 (0.027) for the low dividend one. The hypothesis that it is the same is strongly rejected. The empirical support for large effects of cash flow on firms and financial frictions is very strong.

- **Kaplan and Zingales (1997)**. Kaplan & Zingales (1997) critique FHP on two fronts. First, theoretically, they claim that financially constrained firms may not, in fact, be necessarily more sensitive to cash flows, even if internal finance is cheaper. The issue is that, although firms may make more investment when they have more cash flows, the question is whether this is more the case for more financially constrained firms. Theoretically, this is unclear (it involves the third derivative of the profit function). Empirically, they also question the validity of the sample of firms that are in the constrained group (there are only 49 of them in that group, compared to 334 in the unconstrained group). First there are many reasons that lead firms to choose a high or low dividend level and this may have little to do with credit constraints (for instance, a firm may have a low dividend policy, but have a credit line, or a firm may have a high dividend policy but may be unable to cut it down even in times of crisis).

5 Investment in a model with Uncertainty

Until now, we assumed that there was no uncertainty and we characterized the optimal investment policy. But uncertainty is a powerful force that firms are facing and we need to model it if we want to understand the drivers of investment dynamics.

There are two ways to proceed here. One would be to revert to a discrete time set-up and use the tools from dynamic programming that we used when we looked at the consumption problem under uncertainty and precautionary saving. I will start with that approach. The other approach would be to introduce a stochastic dimension in the continuous time model

Table 4. q **Investment Equations, 1932–78**[a]

		Independent variable		Summary statistic		
Equation[b]	*Constant*	$q-1$	Q	*Rho*	*Standard error of estimate*	*Durbin-Watson*
4-1	0.119 (0.006)	−0.038 (0.019)	...	...	0.039	0.29
4-2	0.096 (0.008)	...	0.026 (0.007)	...	0.036	0.21
4-3	0.104 (0.035)	0.039 (0.016)	...	0.944	0.017	1.27
4-4	0.096 (0.025)	...	0.017 (0.004)	0.923	0.016	1.12
4-5	0.084 (0.033)	0.013 (0.018)	0.015 (0.005)	0.933	0.016	1.11
4-6	0.088 (0.024)	...	0.031 (0.005)	0.922	0.016	1.11
4-7	0.230 (0.039)	−0.106 (0.036)	...	...	0.044	0.43
4-8	0.076 (0.012)	...	0.051 (0.013)	...	0.040	0.34

Source: Estimations by the author.

a. The dependent variable is I/K. Equations in which rho is omitted were estimated without autocorrelation correction. The numbers in parentheses are standard errors.

b. For equation 4-6, the coefficient on Q is the sum of the coefficient on Q and lagged Q. Equations 4-7 and 4-8 were estimated using as instruments the lagged values of the tax variables, θ, c, τ, Z, and ITC.

Figure 2: Summers (1981): Table 4

Table 4. Effects of Q and Cash Flow on Investment, Various Periods, 1970–84[a]

Independent variable and summary statistic	*Class 1*	*Class 2*	*Class 3*
		1970–75	
Q_{it}	−0.0010	0.0072	0.0014
	(0.0004)	(0.0017)	(0.0004)
$(CF/K)_{it}$	0.670	0.349	0.254
	(0.044)	(0.075)	(0.022)
$\overline{R}^2$	0.55	0.19	0.13
		1970–79	
Q_{it}	0.0002	0.0060	0.0020
	(0.0004)	(0.0011)	(0.0003)
$(CF/K)_{it}$	0.540	0.313	0.185
	(0.036)	(0.054)	(0.013)
$\overline{R}^2$	0.47	0.20	0.14
		1970–84	
Q_{it}	0.0008	0.0046	0.0020
	(0.0004)	(0.0009)	(0.0003)
$(CF/K)_{it}$	0.461	0.363	0.230
	(0.027)	(0.039)	(0.010)
$\overline{R}^2$	0.46	0.28	0.19

Source: Authors' estimates of equation 3 based on a sample of firm data from Value Line data base. See text and Appendix B.

a. The dependent variable is the investment-capital ratio $(I/K)_{it}$, where I is investment in plant and equipment and K is beginning-of-period capital stock. Independent variables are defined as follows: Q is the sum of the value of equity and debt less the value of inventories, divided by the replacement cost of the capital stock adjusted for corporate and personal taxes (see Appendix B); $(CF/K)_{it}$ is the cash flow–capital ratio. The equations were estimated using fixed firm and year effects (not reported). Standard errors appear in parentheses.

Figure 3: Fazzari et al (1988): Table 4

we used to characterize optimal investment dynamics in the model with perfect foresight. I will then do that. That way, we will see how both optimization methods work, and we will also build some tools for stochastic optimization in continuous time.

5.1 The model in discrete time with quadratic adjustment costs

Consider the model with constant returns to scale adjustment costs of section 3, but cast in discrete time. The firm starts at the beginning of period t with a capital stock K_{t-1} inherited from the period $t-1$. We assume that there is no depreciation and that the firm can produce immediately with newly installed capital. This means that if the firm chooses to invest I_t, it earns profits $\Pi(K_t, \theta_t)$ in period t, where $K_t = K_{t-1} + I_t$. θ_t is a random variable, such as productivity, or the price of the domestic good, or of inputs... It is observed at the beginning of period t before investment decisions. We assume θ_t follows a Markov process, so that knowing θ_t is the only relevant piece of information for forecasting θ_s for $s > t$. The firm discounts profits at the constant gross interest $R = 1 + r$. We also simplify slightly the problem by assuming that the price of investment goods is constant $p_{Kt} = 1$.

Summing up, the firm solves the following problem:

$$V(K_{t-1}, \theta_t) = \max_{\{I_s\}_t^\infty} E_t \left[\sum_{s=t}^{\infty} R^{-(s-t)} \left(\Pi(K_s, \theta_s) - I_s(1 + D(I_s/K_{s-1})) \right) \right]$$

subject to the following accumulation equation:

$$K_t = K_{t-1} + I_t$$

and where $D(0) = D'(0) = 0$, $D'(.) \geq 0$ and $D''(.) > 0$.

Observe that in this model, the user cost of capital (in the absence of adjustment costs) is simply $r_K = r/R$.[7] The difference with the previous case is that we are taking expectations of future discounted profits. The other change is that the value function is a function of both inherited capital K_{t-1} and the current realization of the stochastic variable θ_t. The latter is here because it helps to predict future realizations of the shocks.[8] Finally, we also assume that the adjustment cost is defined in terms of I_t/K_{t-1}.

We can write the Bellman equation:

$$V(K_{t-1}, \theta_t) = \max_{I_t} \Pi(K_t, \theta_t) - I_t(1 + D(I_t/K_{t-1})) + R^{-1} E_t \left[V(K_t, \theta_{t+1}) \right]$$

[7] In the continuous time limit, this would yield $r_K = r$, as previously. The denominator term arises because we assume that capital is immediately productive.

[8] This implies that if the shocks are iid, the value function is only a function of K_{t-1} as in the deterministic case.

and the first order condition is:

$$1 + D(I_t/K_{t-1}) + (I_t/K_{t-1})D'(I_t/K_{t-1}) = \Pi_K(K_t, \theta_t) + R^{-1}E_t\left[V_K(K_t, \theta_{t+1})\right] \quad (7)$$

while the Envelope condition with respect to capital yields:

$$V_K(K_{t-1}, \theta_t) = \Pi_K(K_t, \theta_t) + (I_t/K_{t-1})^2 D'(I_t/K_{t-1}) + R^{-1}E_t\left[V_K(K_t, \theta_{t+1})\right] \quad (8)$$

These equations look ugly, but in fact the interpretation is very similar to the certainty case. First, define $q_t = \Pi_K(K_t, \theta_t) + E_t[V_K(K_t, \theta_{t+1})]$. This is Tobin's marginal q in period t (equal to the marginal value of one unit of capital installed inside the firm).

From the first order condition, we obtain:

$$I_t = K_{t-1}\phi(q_t)$$

just as in the deterministic model (see equation (6a); The $\phi(.)$ function is the same as in that equation).

From (8) we can then infer:

$$V_K(K_{t-1}, \theta_t) = q_t + \Psi(q_t)$$

where $\Psi(.)$ is defined as in equation (6b). Substituting into the definition of q_t we obtain the law of motion:

$$q_t = \Pi_K(K_t, \theta_t) + R^{-1}E_t[q_{t+1} + \Psi(q_{t+1})]$$

which is the discrete-time equivalent of equation (6b).

The modifications to the model are minimal: it is still the case that firms will set their investment level based on q, but they will take uncertainty into account and replace q with its expected future value.

Notice that if there are no adjustment costs (so that $D(I/K) = 0$) then the equations simplify to:

$$1 = q_t$$

and

$$\Pi_K(K_t, \theta_t) = r_K$$

as expected.

5.2 Discrete time and non-convex adjustment costs

5.2.1 Motivation

We now consider what happens when the adjustment costs, instead of being quadratic (i.e. smooth around 0) are non-convex. This is relevant for a number of reasons:

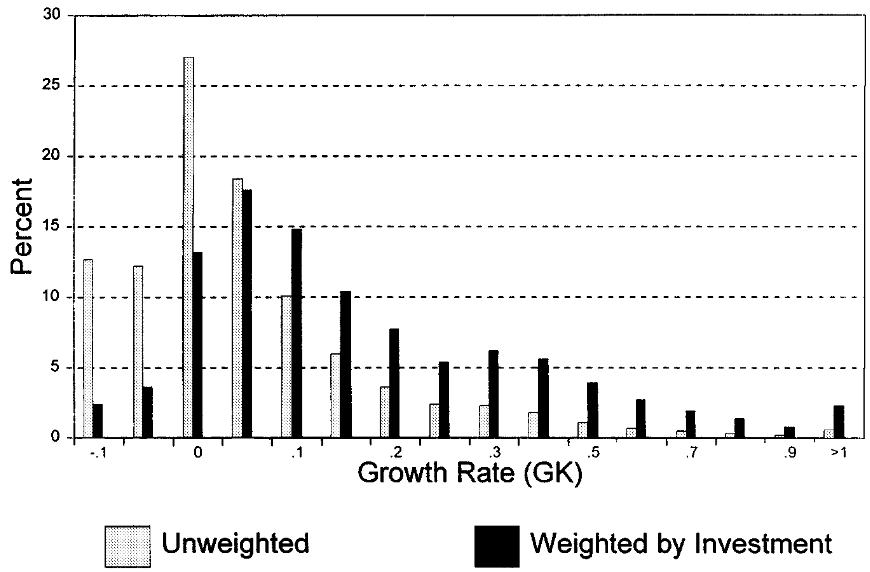

FIG. 1. Capital growth rate (GK) distributions: Unweighted and weighted by investment.

Figure 4: Doms & Dunne (1998): Figure 1

- Empirically, investment at the microeconomic level appears to be quite lumpy and irreversible. A landmark study by Doms and Dunne (1998) at the Census, found that investment at the plant level is both infrequent and 'spiky'. Doms and Dunne look at a sample of 12000 manufacturing plants over the period 1972-1989. They find that on average, the largest investment episode accounts for 25% of the overall investment over the entire period, and represents 50% of investment for more than half the establishments (see figures 4 and 5).

- This 'lumpiness' would not matter much if it was randomly distributed over plants and time, so that a model of aggregate investment with smooth adjustment costs could still account for the empirical evidence. But this does not appear to be the case: Doms and Dunne find that 18% of investment is accounted for by top projects: there is granularity in the data and the structure of investment at the microeconomic level seems to matter.

- We know that the q theory does not perform very well when it comes to explaining aggregate investment dynamics. Some of this is probably due to financial frictions, but some of it is most certainly due to the importance of heterogeneity

- It will allow us to explore some cool new tools!

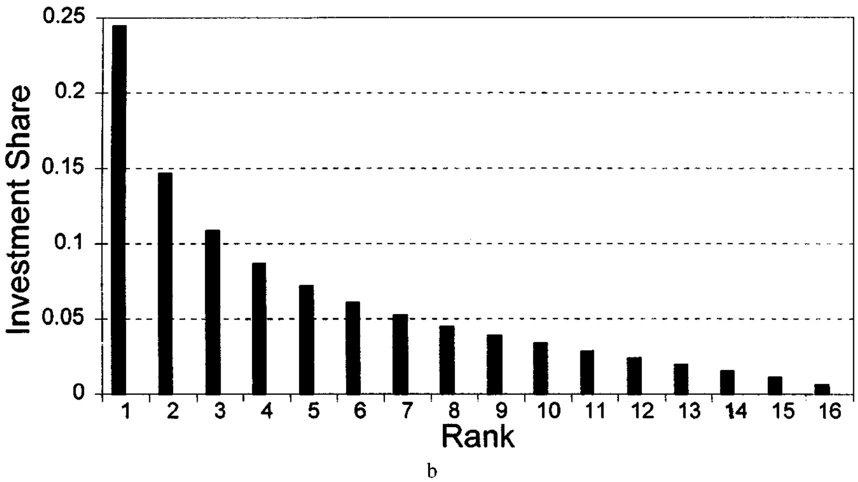

FIG. 2. Capital growth rates (GK) by rank, means, and medians. (b) Mean investment shares by capital growth rate rank.

Figure 5: Doms & Dunne (1998): Figure 2b

5.2.2 A detour by the frictionless model

It is useful to define the 'target' level of capital as the level of capital that the firm would choose in the absence of adjustment costs. To fix ideas, suppose that we can write:

$$\Pi(K, \theta) = K^{\alpha}\theta.$$

θ represents productivity and α is related to the market power of the domestic firm.

The preceding analysis indicates that, in the absence of adjustment costs, the choice of capital would satisfy:

$$r_K = \Pi_K(K_t, \theta_t) = \alpha K_t^{\alpha-1}\theta_t$$

We can solve this expression for the desired capital stock K_t^*:

$$K_t^* = \left(\frac{\alpha\theta_t}{r_K}\right)^{1/(1-\alpha)}$$

We can then define the capital gap as the ratio $Z_t = K_t/K_t^*$. In the frictionless model K_t and K_t^* are always equal and $Z_t = 1$. But this is no longer necessarily the case when there are adjustment costs. Nevertheless, we should expect (in a sense to be made clear) that firms' investment decisions will 'tend' towards $Z = 1$, i.e. that they will aim to close the gap between current and desired capital.

For instance, in the quadratic adjustment cost model, it is easy to rewrite the optimal investment policy as (using the fact that $I_t = K_t - K_{t-1}$):

$$\frac{Z_t}{Z_{t-1}} = \frac{K^*_{t-1}}{K^*_t}(1+\phi(q_t)) = \left(\frac{\theta_{t-1}}{\theta_t}\right)^{1/(1-\alpha)}(1+\phi(q_t))$$

This equation shows that –in general– the capital gap will not be equal to 1. Instead, it will vary with (a) the previous capital gap; (b) the change in productivity which is not predictable and tells us how desired capital changes; and (c) q_t, which controls how desirable investment is. The equation tells us that, if the shocks remain constant between two periods, $Z_t > Z_{t-1}$ if $q_t > 1$ and $Z_t < Z_{t-1}$ otherwise.

5.2.3 Non-convex adjustment costs

We now consider what happens when the firm faces non-convex adjustment costs. Instead of postulating an adjustment cost function $C(I, K)$ that is smooth, we will assume that the firm potentially faces both fixed and flow variable costs. More specifically, let's assume that the firm faces the following costs:

- **Fixed Costs**: Assume that the firm has to pay a fixed cost $C_l K^*_t$ if it adjusts upwards at time t and $C_u K^*_t$ if it adjusts capital downwards.
- **Variable costs**: Assume that the firm has to pay a flow cost $c_l I_t$ if investment is positive ($I_t > 0$) and a flow cost $-c_u I_t$ if investment is negative ($I_t < 0$).
- **no cost if no adjustment**

The overall adjustment cost function is then:

$$C(\eta, K^*) = K^*\left[(C_l + c_l\eta)1_{\{\eta>0\}} + (C_u - c_u\eta)1_{\{\eta<0\}}\right]$$

where $\eta = I/K^*$. The cost function is represented on figure 6.

Consider what happens as a result of these costs. First, it should be pretty obvious that the fixed costs C_u, C_l are going to induce a range of inaction: it makes more sense to bunch investment and pay the fixed cost less frequently. This is true even with uncertainty.

But in fact the same is true with the variable costs c_u and c_l. Suppose that there are no fixed costs, but strictly positive variable costs. The firm may be cautious about investing one extra unit because it is possible that tomorrow the desired capital will decrease, forcing the firm to disinvest. In that case, capital is unchanged, but the firm ends up paying $c_u + c_l > 0$.[9]

The upshot is that both types of costs in presence of uncertainty will induce the firm to delay investing over a certain range. Intuitively, this range will be close to the desired capital, i.e. $Z = 1$. Over that range, since K will not adjust, Z will be moving as a result to shocks to K^*.

[9] By contrast, with convex adjustment costs, that cost would be 0.

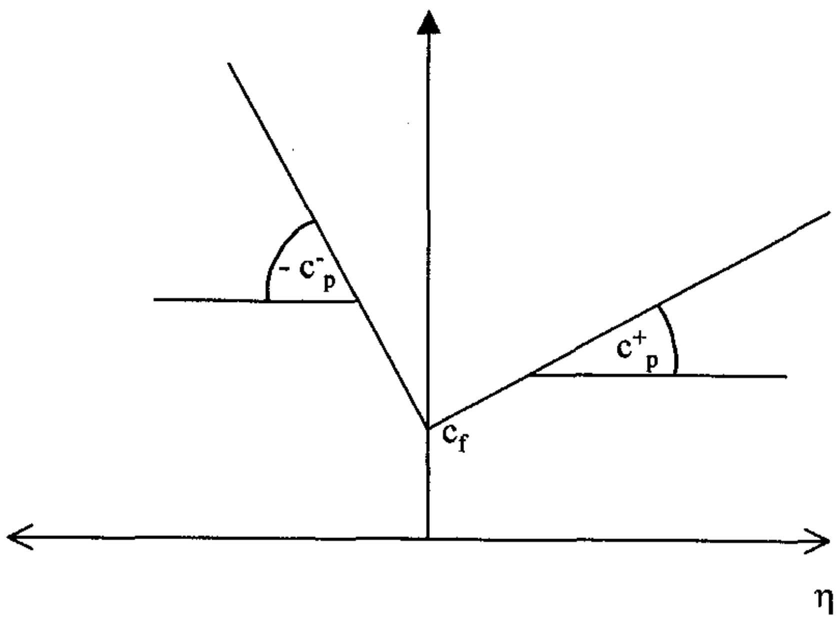

Figure 6: Non-Convex Adjustment Costs. In that figure, $c_p^+ = c_l, c_p^- = c_u$ and $c_f = C_u = C_l$

5.2.4 Characterizing the Solution

Let's now derive the shape of the optimal solution. The rigorous way to do this would be to first set-up the problem in continuous time and apply optimal control theory for stochastic processes. I will discuss later how this is done at a more general level. But the intuition can be obtained quite easily and without fancy maths from the discrete time set-up, so this is what I do here.

The first step is to express the profit function in terms of the capital gap Z:

$$\Pi(K_t, \theta_t) = \frac{r_K}{\alpha} Z_t^{\alpha} K_t^*$$

We will express the problem in terms of the state variables Z and K^* instead of K and θ. We know from the preceding discussion that the firm will adjust capital infrequently. The way to model this is to consider two value functions. The first value function $V(Z, K^*)$ when the firm is not adjusting its capital stock. The second value function $\tilde{V}(Z, K^*)$ denotes the value of the firm which can choose whether to adjust capital.

Now, we consider the Bellman equation over a very small interval of length Δt. We consider first the region where the firm is not adjusting its stock of capital, so $\eta_t = I_t = 0$.

The Bellman equation takes the following form:[10]

$$V(Z_t, K_t^*) = \frac{r_K}{\alpha} Z_t^\alpha K_t^* \Delta t + (1 - r\Delta t) E_t[\tilde{V}(Z_{t+\Delta t}, K_{t+\Delta t}^*)]$$

This equation says that the value of the firm is equal to the flow profits in the current period, plus the expected discounted value of the firm tomorrow, taking into account that the firm may decide to adjust its capital stock at time $t + \Delta t$ (hence the value function inside the expectation operator is $\tilde{V}$).

The other equation chooses whether to adjust and by how much:

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

where $C(\eta, K^*)$ is the non-convex adjustment cost function described earlier. This second equation says that (a) the firm will choose to adjust its capital stock only if it yields a higher value to the firm than doing nothing and (b) the value of the firm immediately after adjusting is the value at the new level of capital, net of the adjustment costs, and net of the purchase cost of capital (the last term ηK_t^*).

Since both adjustment costs function and profits are linear in K^*, the value functions are also homogenous of degree one in K^*. This implies that the range of inaction is going to be invariant in the space of imbalances and we can define normalized value functions $v(Z) = V(Z, K^*)/K^*$ and $\tilde{v}(Z) = \tilde{V}(Z, K^*)/K^*$. These normalized value functions satisfy the following Bellman equations:

$$\begin{aligned} v(Z_t) &= \frac{r}{\alpha} Z_t^\alpha \Delta t + (1 - r\Delta t) E_t[v(Z_{t+\Delta t})] \\ \tilde{v}(Z_t) &= \max \left\langle v(Z_t), \max_\eta (v(Z_t + \eta) - c(\eta) - \eta) \right\rangle \end{aligned}$$

where

$$c(\eta) = \left[(C_l + c_l \eta) 1_{\{\eta > 0\}} + (C_u - c_u \eta) 1_{\{\eta < 0\}} \right]$$

Notice that what makes the problem much simpler here is that we made all the right assumptions to 'scale' the problem by desired capital K^*.[11]

Since we know that the solution will feature a range of inaction, it will be characterized by four parameters:

- the points L and U at which the firm will adjust(triggers);
- the points $l > L$ and $u < U$ where it will return (targets).

[10] This is where we are using our 'knowledge' of continuous time stochastic optimization.
[11] Since K^* is a function of the shock θ this simplification is very useful.

What equilibrium conditions must v satisfy? First, in the inaction range, the continuous time version of the Bellman equation characterizes a second order ordinary differential equation in v. The novelty of the problem is that the boundary conditions of this equation are themselves endogenous: they are the four parameters above that define the range of the value function. Typically a second order differential equation requires two parameters. This means we have a total of six conditions to satisfy to characterize fully this problem. What are these six conditions?

First, it must be the case that the firm is indifferent between adjusting and not adjusting at the boundary:

$$\begin{aligned} v(L) &= v(l) - (C_l + (1 + c_l)(l - L)) \\ v(U) &= v(u) - (C_u + (1 - c_u)(u - U)) \end{aligned}$$

These conditions are called value matching. The only difference between the trigger and target points must be the adjustment costs.

Now the other four conditions are obtained by optimizing over η, the size of the adjustment, conditional upon adjustment:

$$\begin{aligned} v'(l) &= 1 + c_l \\ v'(u) &= 1 - c_u \end{aligned}$$

Similarly, we must ensure that the there is no advantage to delaying adjustment:

$$\begin{aligned} v'(L) &= 1 + c_l \\ v'(U) &= 1 - c_u \end{aligned}$$

These conditions are called smooth pasting. They ensure that there is no kink in the value function at the points at which the adjustment occurs.

This provides us with the six conditions we need to determine both the shape of the value function as well as the range of inaction and the optimal adjustment policy.

Observe that :

- if the firm for some reason found itself outside the range $[L, U]$, it would adjust immediately. This implies that the value function for $Z \leq L$ (for instance) is $v(Z) = v(l) - C_l - (1 + c_l)(l - Z)$ and is linear in Z with slope c_l.
- if there are no variable costs of adjustment ($c_u = c_l = 0$), then $l = u = c$, i.e. the adjustment is complete on either side. However, it is not necessarily the case that $c = 1$ (i.e. it's possible for the adjustment to be such that $K \neq K^*$, in particular if there is a drift in the shock process –e.g. if K^* increases over time).
- if there are no fixed costs of adjustment ($C_u = C_l = 0$), then the process is regulated: there is no reason no to adjust infinitesimally, once the boundaries are reached. This means $L = l$ and $U = u$.[12]

[12]In that case, we lose 2 boundary conditions. However, one can show that smooth pasting requires that $v''(L) = v''(U) = 0$.

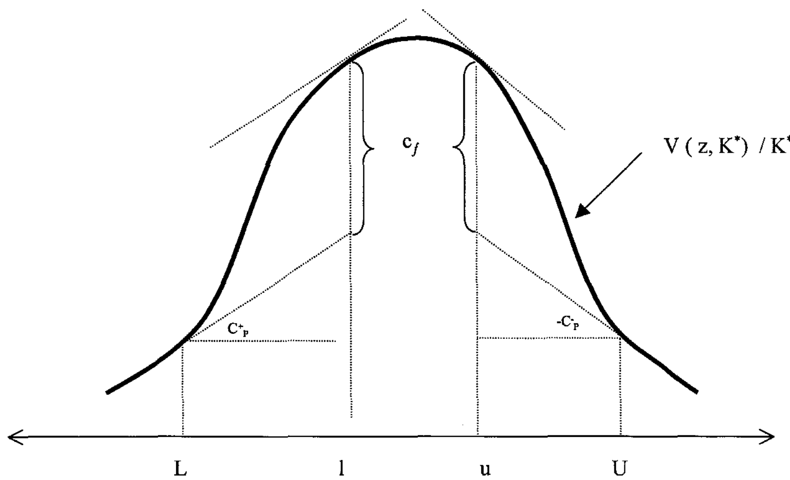

Figure 7: Optimal Value Function. From Caballero's Handbook chapter. Note: the value function defined in Caballero corresponds to $v(z) - z$ in terms of our notation.

5.2.5 Non-Convex Adjustment Costs and q

Heuristically, we can define q as $v'(Z)K^*$ in this model.[13] Figure 8 plots the value of q as a function of the capital gap Z. It is clear that there is no monotonous relationship between q and investment (or the capital gap). Since q takes the same value at the trigger and target points, but investment is large at the trigger point and zero at the target point, it is going to be difficult to obtain a meaningful relationship between q and investment.

5.3 Stochastic Dynamic Programming

Let us now fill in some of the blanks by considering a full fledged stochastic optimization problem. We will do this in a slightly more general context than the one studied above, and then derive the appropriate implications for the investment problem. Consider the following optimization problem, denoted (P), which is a continuous time analog of the discrete time set-up we considered above.

$$V(x_t) = \max_{dA} E\left[\int_t^{\infty} e^{-\rho(s-t)}\left(\tilde{g}(x_s)\ ds - dC_s\right)|x_t\right] \tag{9a}$$

$$dx_s = \mu(x_s)\ ds + \sigma(x_s)\ dw_s + dA_s \tag{9b}$$

$$dC_s = \phi(dA_s) \tag{9c}$$

In the problem above, $V(x)$ is the value function, equal to the discounted value of some flow payoff $\tilde{g}(x)$ which depends on the state variable x. The second equation describes the

[13]To see this, note that the value of the firm is $K + V(K, \theta)$.

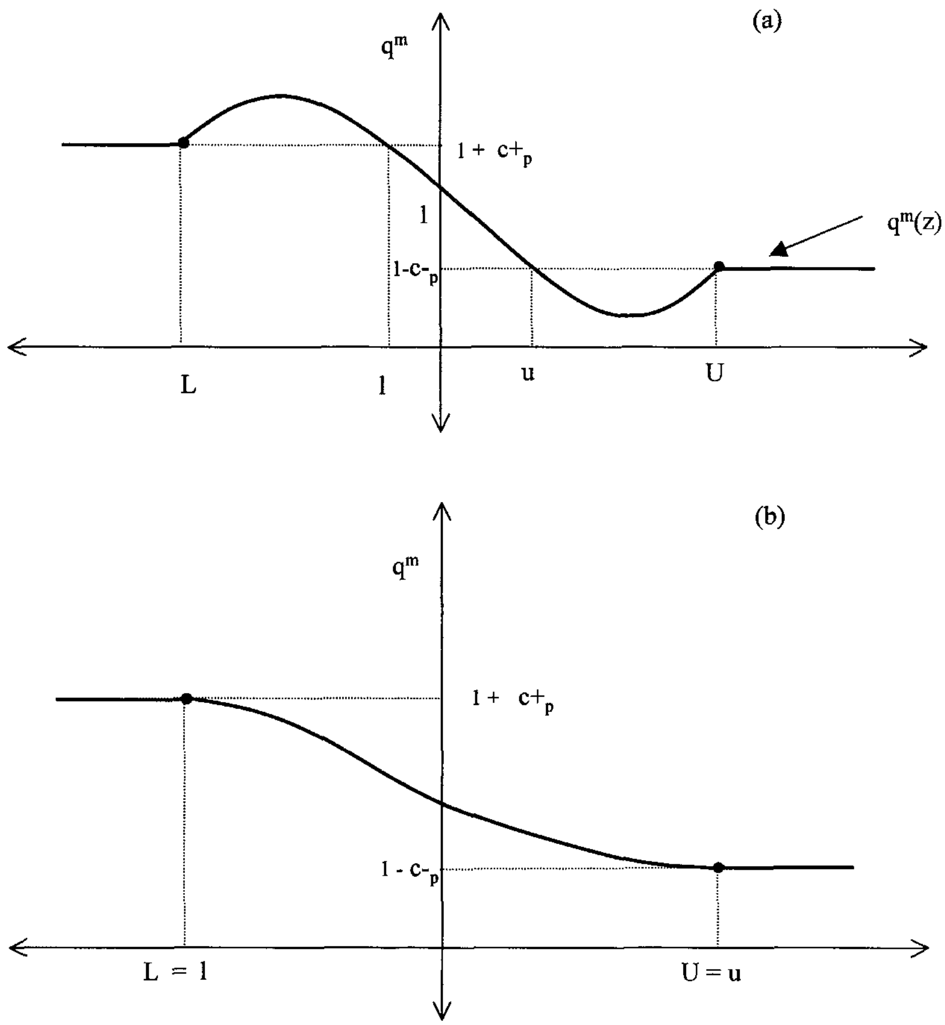

Figure 8: Tobin's marginal q. From Caballero's Handbook chapter. The top graph corresponds to the case with $C_u = C_l = C_f > 0$. The bottom graph corresponds to the case $C_u = C_l = 0$.

law of motion of the state variable. w_t is a standard Brownian Motion. For those of you who are not familiar with Brownian motions, they are the basic building bloc of continuous time stochastic processes. A Brownian motion w_t is a stochastic process such that:

- the increments dw between t and $t + dt$ are i.i.d
- the increments are normally distributed with mean 0 and standard error $\sqrt{dt}$.

The variance of the increments is what makes Brownian motions special: heuristically, the variance of the innovation is dt, i.e. loosely speaking '$(dw)^2$ is of order dt'.

The second equation specifies that over an interval of time dt, the state variable x changes because of a 'drift' term $\mu(x)$, which would correspond to $\dot{x}$ in the deterministic case. In addition to the drift term, there is also a stochastic adjustment coming from the innovation dw to the Brownian motion. This is the stochastic volatility component. This volatility term complicates things because it implies that the usual time derivative dx/dt is not well defined any more: if you look at the change $x_{t+\Delta t} - x_t$, it is equal to $\mu(x_t)\Delta t + \sigma(x_t)(w_{t+\Delta t} - w_t)$. If you divide by Δt and take the limit as $\Delta t \to 0$, you can check that the ratio $\lim_{\Delta t \to 0}(w_{t+\Delta t} - w_t)/\Delta t$ diverges (again, in a heuristic sense because $w_{t+\Delta t} - w_t$ is of order $\sqrt{\Delta t}$ so the ratio is of order $1/\sqrt{\Delta t}$). In short, while the process x (or w) is continuous, it is nowhere differentiable!! Stochastic calculus develops the tools we need to be able to manipulate processes like this.

dA_s is the **control variable** and represents the change in the state variable x. Thus, A_s represents the cumulative adjustment up to time s and dC_s represents the cost of adjusting by dA_s.

Specified this way, the problem is quite general and encompasses the usual case of quadratic adjustment costs as well as the non-smooth optimization problems. Note also that the adjustment cost nor the adjustment itself need not be infinitesimal in the time interval dt. In particular, if we shift x discretly (i.e. $dA > 0$), then dA_s/ds is infinite, corresponding to an infinite rate of adjustment.

5.3.1 Quadratic Adjustment Cost Case;

We start by describing the solution method and concepts when the adjustment cost is convex in the rate of adjustment. Assume that:

$$dC_s = \psi\left(\frac{dA}{ds}\right) ds \tag{10}$$

where $\psi(.)$ is convex, with $\psi(0) = \psi'(0) = 0$. In this situation, adjusting x is reversible for small adjustments. Given the convexity in ψ, we will never want to adjust by a discrete amount instantly: this would entail an infinite cost. Thus, we can define the following control variable:

$$i_s = \frac{dA_s}{ds}$$

i_s represents the rate of adjustment. It is akin to 'investment.' We can then rewrite Problem ($\mathcal{P}$) as:

$$\begin{aligned} V(x_t) &= \max_{i_s(.)} E_t \left[\int_t^{\infty} e^{-\rho(s-t)} g(x_s, i_s)\ ds \right] \\ dx_s &= f(x_s, i_s)\ ds + \sigma(i_s)\ dw_s \end{aligned}$$

where $f(x_t, i_t) = \mu(x_t) + i_t$ and $g(x_t, i_t) = \tilde{g}(x_t) - \psi(i_t)$. In order to solve this problem, we would like to apply the *Bellman Principle* to derive the *Bellman Equation*, as we did in the discrete time case. Remember that the Bellman Principle states that if a policy function is optimal for the original problem, it must be optimal for any sub-problem along the path.

In continuous time, the equivalent of period $t+1$ is period $t+dt$. So we would like to write the Bellman Principle between t and $t+dt$. Before we do this, we need one piece of machinery: Itô's Lemma.

5.3.2 Itô's lemma:

Itô's Lemma tells how to write the 'stochastic derivative' of a function of stochastic process. Consider a stochastic process of the form:

$$dx_t = \mu(x_t)dt + \sigma(x_t)dw_t \tag{11}$$

and suppose that we are interested in a function of x: $y = f(x)$. What is the stochastic process followed by y? The answer is given by Itô's lemma:

Proposition 1 (Itô's Lemma) *If x follows the stochastic process (11), then $y = f(x)$ follows:*

$$dy_t = \left[f'(x_t)\mu(x_t) + \frac{1}{2} f''(x_t)\sigma(x_t)^2 \right] dt + f'(x_t)\sigma(x_t)dw_t \tag{12}$$

Notice that Itô's Lemma tells us that the 'usual' rule of differentiations needs to be modified. The usual chain rule of differentiation would tell us that $dy = f'(x)dx$. But this is incorrect according to Itô's lemma: there is an additional term on the right hand side that involves the second derivative of the function f: $1/2 f''(x)\sigma(x)^2$.

First note that Itô's lemma gives a different answer from the usual rules of calculus only when the function f has some curvature, i.e. when $f''(.) \neq 0$. To get some intuition for this term, let's use a second-order Taylor expansion of y_{t+dt} around y_t. We can write:

$$\begin{aligned} dy_t &= y_{t+dt} - y_t = f(x_{t+dt}) - f(x_t) \\ &= f'(x_t)dx_t + \frac{1}{2}f''(x_t)(dx_t)^2 + o(||dx||^2) \\ &= f'(x_t)\mu(x_t)dt + f'(x_t)\sigma(x_t)dw_t + \frac{1}{2}f''(x_t)(dx_t)^2 + o(||dx||^2) \end{aligned}$$

Now, the key thing is to collect all the terms of order dt or below in this expression. In figuring out the order of a term, we use the 'convention' that terms in dw are of order $\sqrt{dt}$. The first term on the right is the one we would obtain by the usual chain rule of differentiation. It involves one term of order dt and one term of order $\sqrt{dt}$ so we keep both.

What about $(dx_t)^2$? We can write it as:

$$\begin{aligned} (dx_t)^2 &= (\mu(x_t)dt + \sigma(x_t)dw_t)^2 \\ &= \mu(x_t)^2(dt)^2 + 2\mu(x_t)\sigma(x_t)dtdw_t + \sigma(x_t)^2(dw_t)^2 \end{aligned}$$

Notice that the last term in this expression is, in fact, of order dt. So we need to keep that term too. All the others terms are of order higher than dt and can discarded. If we put things back together, we obtain Itô's Lemma.

Observe that if we evaluate the conditional expectation of dy_t (where the expectation is conditional on information available at time t), we obtain (since $E_t[dw_t] = 0$):

$$E_t[dy_t] = \left[f'(x_t)\mu(x_t) + \frac{1}{2}f''(x_t)\sigma(x_t)^2\right] dt$$

It follows that the expected change in y is differentiable and we can define the expected rate of change of y as:

$$\frac{E_t[dy_t]}{dt} = f'(x_t)\mu(x_t) + \frac{1}{2}f''(x_t)\sigma(x_t)^2$$

Now, that we know how to use Itô's lemma, let's apply it to $V(x)$. Given that V is only a function of x, we can write:

$$dV(x) = \left(V'(x)\ f(x,i) + \frac{1}{2}\ V''(x)\ \sigma^2(x)\right)\ dt + V'(x)\ \sigma(x)\ dw$$

Thus:

$$\frac{E[dV]}{dt} = V'(x)\ f(x,c) + \frac{1}{2}\ V''(x)\ \sigma^2(x)$$

More generally, we can define the operator $\mathcal{D}$, for any function $G(x,t)$:

$$\mathcal{D}G(x,t) = \frac{\partial G(x,t)}{\partial t} + \frac{\partial G(x,t)}{\partial x} \quad f(x,i) + \frac{1}{2}\frac{\partial^2 G(x,t)}{\partial x^2} \quad \sigma^2(x)$$

and summarize the previous expression as:

$$\frac{E[dV]}{dt} = \mathcal{D}V$$

5.3.3 The Hamilton-Jacobi-Bellman Equation:

We are now in a position to apply apply the Bellman Principle. We write the Bellman equation between times t and $t+dt$ as we did in the previous note, and expand, using the rule of stochastic calculus we just learned:

$$\begin{aligned}
V(x_t) &= \max_{i_s(.)} E_t\left[\int_{s=t}^{\infty} e^{-\rho(s-t)}\, g(x_s, i_s)\ ds\right] \\
&= \max_{i_t(.)}\left\{g(x_t, i_t)\ dt + e^{-\rho\ dt} E_t\left[\int_{s=t+dt}^{\infty} e^{-\rho(s-t-dt)}\, g(x_s, i_s)\ ds\right]\right\} \\
&= \max_{i_t}\left\{g(x_t, i_t)\ dt + e^{-\rho\ dt} E_t\left[V(x_{t+dt})\right]\right\} \\
\rho\, V(x_t) &= \max_{i_t}\left\{g(x_t, i_t)\ + \frac{E_t[dV(x_t)]}{dt}\right\}
\end{aligned}$$

where the last equation follows from the Taylor expansion.

Using Itô's lemma, we obtain the *Continuous Time Hamilton-Jacobi-Bellman Equation*:

$$\boxed{\begin{aligned}
\rho\, V(x) &= \max_i \{g(x,i) + \mathcal{D}V(x)\} \\
&\text{or} \\
\rho\, V(x) &= \max_i \left\{g(x,i) + V'(x)\ f(x,i) + \tfrac{1}{2}\, V''(x)\ \sigma^2(x)\right\}
\end{aligned}} \qquad (13)$$

Notice the similarity with the deterministic case: the only difference is the 'curvature term' $V''(x)\sigma^2(x)/2$ on the right hand side. The interpretation is straightforward: if we think of the value function as the price of an asset, the Bellman equation is simply an arbitrage equation:

$$\rho = \max_i \left\{\frac{g(x,i)}{V(x)} + \frac{\mathcal{D}V(x)}{V(x)}\right\}$$

The left hand side is the relevant discount rate. The first term on the right hand side represents the *flow payment* divided by the price of the asset. It is the equivalent of a *dividend price ratio*. The second term represents the *expected capital gain*.

5.3.4 Euler Equation

We now write the First Order condition of the maximization problem (13):

$$\boxed{g_i\,(x,i) + V'(x)\;\; f_i\,(x,i) = 0} \tag{14}$$

This First-Order Condition is only *necessary* and defines $i^*\,(x)$, the optimal adjustment function.

The optimal policy function entails adjustment in every period. Going back to the definition of g and f:

$$\psi'\,(i) = V'(x)$$

Thus the optimal policy is such that the *ratio* $\psi'\,(c)\,/V'(x)$ *is kept equal to 1 at all times*.

This result is very general: you adjust so as to stay on the margin. Here the left hand side represents the marginal cost of adjusting by 1 unit, and the right hand side represents the marginal benefit.

Specializing the results even further, assume that $\psi\,(i) = i + \frac{1}{2}\,i^2$. Assume further that $\mu = -\delta x$ and σ is constant. Then it is easy to see that:

$$\begin{aligned} V'\,(x) &= E\left[\int_0^\infty e^{-(\delta+\rho)s}\;\tilde{g}'\,(x)\;\, ds\right] \\ i &= V'(x) - 1 \end{aligned}$$

and this is the traditional q-theory of investment, with $q = V'(x)$. The marginal value of the firm is the discounted expected marginal product of capital, and investment takes place when it exceeds the price of the investment good (1).

5.3.5 Envelope Theorem

Now take a derivative with respect to the state variable x. According to the Envelope Theorem, we do not need to consider the induced variations in i^*: they are of second order. Hence:

$$\rho\,V'(x) = g_x\,(x,i^*) + V''(x)\;\; f\,(x,i^*) + V'(x)\;\; f_x\,(x,i^*) + \frac{1}{2}V'''\,(x)\;\sigma^2\,(x) + V''\,(x)\;\sigma\,(x)\;\sigma'\,(x)$$

Note that we do not have the max operator on the right hand side since we are at the optimum i^*. This expression looks ugly, but you might observe that this is equivalent to:

$$(\rho - \; f_x\,(x,i^*))\;\; V'(x) = g_x\,(x,i^*)\; + V''\,(x)\;\;\sigma\,(x)\;\;\sigma'\,(x) + \mathcal{D}V'\,(x) \tag{15}$$

This is the equivalent of the differential equation for Tobin's q in the investment model. Formally, if we define $\tilde{g}(x) = \Pi(x)$, $f(x) = -\delta x$, σ constant and $\rho = r$, we obtain

$$(r + \delta)\ q = \Pi_x(x) + \mathcal{D}q$$

which corresponds to equation (5b).

5.4 Non Smooth Optimization Problems:

5.4.1 Different types of costs:

As mentioned above, in the quadratic adjustment cost case, infinitesimal adjustments are both costless and reversible. On the contrary, large adjustment shifting the state variable discretly are extremely costly.

We now consider a somewhat polar case where adjustment -however infinitesimal- is only partially reversible, if at all.

We consider now these different types of costs:

- **Fixed Costs** (C_u, C_l): every time you adjust upward (resp. downward), you pay the fixed (i.e. independant of dt and c) cost C_l (resp. C_u).
- **Kinked Linear Costs** (c_u, c_l), with $c_u \neq c_l$potentially: The cost is proportional to the adjustment. Formally, the cost is:

  $$\begin{array}{r}\text{if } dA > 0 : \phi\,(dA) = c_l\ dA \\ \text{if } dA < 0 : \phi\,(dA) = -c_u\ dA\end{array}$$

  In the case where $c_u = -c_l$, we have a perfectly reversible adjustment cost and the previous technique will apply: it is optimal to adjust continuously.

  When the cost curve is kinked (i.e. $c_u + c_l > 0$), you cannot reverse totally your adjustment. This is a case of *partially reversible adjustment*. This situation occurs when there is some specificity in the asset you buy, or when there are signalling problems in the market for used goods. Typically, you can only resell at a discount.

With fixed costs or kinked variable costs, it will be optimal not to adjust every period: the solution will feature an inaction range.

In what follows I will assume that we have both fixed and kinked adjustment costs. Thus:

$$\begin{array}{r}\text{if } dA > 0 : \phi\,(dA) = C_l + c_l\ dA \\ \text{if } dA < 0 : \phi\,(dA) = C_u - c_u\ dA\end{array} \tag{16}$$

We can then rewrite ($\mathcal{P}$):

$$
\begin{aligned}
V(x_0) &= \max_{\{dA_t\}} E_0 \left[\int_0^\infty e^{-\rho t} \left(\tilde{g}(x_t)\ dt - 1_{\{dA_t > 0\}} (C_l + c_l\ dA_t) - 1_{\{dA_t < 0\}} (C_u - c_u\ dA_t) \right) \right] \\
dx_t &= \mu(x_t)\ dt + \sigma(x_t)\ dw_t + dA_t
\end{aligned}
$$

In technical terms, this problem is a *free-boundary problem*: we have to find simultaneously the value function and the optimal boundaries of the inaction range.

5.4.2 Structure of the optimal policy function:

As discussed earlier, it should be clear that it is not optimal to adjust continuously. The general rule will be one of *inaction*, interspersed by adjustments.

As a result, marginal costs and marginal benefit will typically differ as long as no action is taken. An action will be triggered by large imbalances between the relevant marginal benefit and marginal cost, to take into account the presence of the fixed or kinked adjustment cost structure.

As in the discrete time example, the most general rule consists of 4 points (L, l, u, U) around the optimal value x^* solving $\tilde{g}'(x) = 0$. U and L are respectively the upper and lower *trigger* points, while u and l are the upper and lower *target* points. When the state variable reaches U, it jumps instantaneously back to u, where $L < u \leq U$, and when the system reaches L, it jumps to l, where $L \leq l < U$.

In some cases, it will appear that only 1, 2 or 3 of these trigger and return points are relevant.

5.4.3 When it is optimal not to adjust:

Given a postulated rule, we can ask the question: what is the value function inside the inaction range $[L, U]$? Since there is no adjustment (by definition) in the inaction range, we know that the Bellman Equation is:

$$
\rho\, V(x) = \tilde{g}(x) + V'(x)\ \mu(x) + \tfrac{1}{2}\, V''(x)\ \sigma^2(x) \tag{17}
$$

This is a second order differential equation. Its general solution is the sum of a particular solution and the solution to the homogenous equation (without the $\tilde{g}$ term). An educated guess is to try a solution of the form of $\tilde{g}$ for the particular solution. Typically, the solution will depend on two integration constants, A_1 and A_2. These two constants must be determined by the boundary conditions of the problem, to which we now turn.

5.4.4 When it is optimal to adjust:

Value Matching: Given the rule (L, l, u, U) that we postulated, when the state variable reaches L, it immediately jumps to l. Thus the value of being at L is exactly the value of being at l minus the adjustment cost to go there. A similar reasoning at the upper boundary provides the following two boundary conditions:

$$\boxed{\begin{array}{rcl} V(L) &=& V(l) - C_l - c_l \ (l - L) \\ V(U) &=& V(u) - C_u - c_u \ (U - u) \end{array}} \tag{18}$$

We can solve for the integration constants A_1 and A_2 that satisfy these *Value Matching* conditions. Note that no optimality is involved in these conditions. They are conditions that define the value at the trigger and return points, given these points. By a similar reasoning, we also know that:

$$\begin{array}{rcll} V(x) &=& V(l) - C_l - c_l \ (l - x); & \text{for } x \leq L \\ V(x) &=& V(u) - C_u - c_u \ (x - u) & \text{for } x \geq U \end{array} \tag{19}$$

Smooth Pasting: We now ask the following question: what is the optimal rule in that family? Consider what it means for a rule to be optimal: no other rule in the same family can yield a higher value. In particular, it cannot be optimal to adjust when $x \neq L$ or $x \neq U$. Thus, if we adjust say from x to y, then it must be true that:

$$\begin{array}{rcll} V(x) &\geq& V(y) - C_l - c_l \ (y - x); & \text{for } x < y \\ V(x) &\geq& V(y) - C_u - c_u \ (x - y) & \text{for } x > y \end{array}$$

Now, let us concentrate on the first line: take x close to L, and y close to l. We can expand and rewrite the equation as:

$$V(L) + V'(L) \ (x - L) \geq V(l) + V'(l) \ (y - l) - C_l - c_l \ (y - x)$$

Using the Value Matching condition, we rewrite:

$$\left(V'(L) - c_l\right) \ (x - L) + \left(V'(l) - c_l\right) \ (y - l) \geq 0$$

This as to be satisfied for any $x < y$, hence we must have the *Smooth Pasting* conditions:

$$\boxed{\begin{array}{rcccl} V'(L) &=& V'(l) &=& c_l \\ V'(U) &=& V'(u) &=& -c_u \end{array}} \tag{20}$$

These 4 conditions allow to identify the remaining 4 unknowns: L, l, u, U, characterizing fully the equilibrium. See the graphical interpretation.

Another way of deriving the Smooth Pasting Conditions might be more illuminating. Define ξ as the adjustment when a trigger point is reached. We can rewrite the value matching condition as:

$$V(x) = V(x+\xi) - \phi(\xi)$$

at any point where there is an adjustment and $\phi(\xi)$ is the cost function. Now, we have to optimize on the *size* of the adjustment ξ. Thus, at any trigger point, we must have:

$$V'(x+\xi) = \phi'(\xi)$$

or, in our case,

$$V'(l) = c_l; \qquad V'(u) = -\, c_u \tag{21}$$

This gives 2 conditions. To get the last 2 ones, consider equation 19 and differentiate to the left of L and to the right of U. We get:

$$V'\left(L^-\right) = c_l; \qquad V'\left(U^+\right) = -c_u$$

Now, one can show that V has to be differentiable at L and U, hence the result.

5.4.5 Special Cases;

1. Fixed Costs only (i.e. $c_u = c_l = 0$): in that case, $u = l$ and we have the familiar (S, s) model.

2. No fixed cost (i.e. $C_u = C_l = 0$); Supercontact conditions;

   In the situation where $C_u = C_l = 0$, the results turn out to be slightly different. Without fixed cost, the only impediment to continuous adjustment is the presence of partial irreversibility associated with the kink in the cost schedule. However nothing prevents adjustment, when it occurs, to be infinitesimal. This will indeed be the optimal solution, and $L = l$, $U = u$.[14] The problem of course is that there are now only 2 boundary conditions and 4 unknowns (as the Value Matching condition does not bring any information). The trick is to work instead with V'. Defining $v = V'$, we can rewrite the Envelope Condition as:

   $$\left(\rho - \mu'(x)\right)\, v(x) = \tilde{g}'(x) + v'(x)\,\sigma(x)\,\sigma'(x) + \mathcal{D}v(x)$$

   This is a second order differential equation in v that we can -hopefully- integrate as before. Now the boundary conditions on v are, on one hand:

   $$v(L) = c_l \quad ; \quad v(U) = c_u$$

   and on the other hand (by a reasoning similar to the one leading to the Smooth Pasting condition):

[14] This can also be seen directly from the Value Matching and Smooth Pasting conditions. $L = l$ and $U = u$ satisfies identically the Value Matching condition and does not violate the Smooth Pasting ones.

$$\boxed{v'(L) = 0 \quad ; \quad v'(U) = 0}$$

These last conditions are called the *Super Contact* conditions. In this situation, the state variable x follows a *regulated* Brownian motion: adjustment occurs marginally so that x never moves outside of the band. This is sometimes dubbed the "corridor model".

3. More general problems: when the per period payoff depends on some exogenous process: $\tilde{g}(x, y)$, then the optimal value for the state variable x^* varies over time.The trick is to make the problem stationary again by defining a new state variable. Typically, one can use the ratio marginal benefit/marginal cost, or the deviation from the optimum: $z_t = x_t - x_t^*$.

5.5 Aggregation

Models with lumpy investment can capture important aspects of investment dynamics at the micro economic level. One important question, though, is how to go from the micro to the macro dimension. At one level, we'd expect some of the lumpiness to wash out as we aggregate: at anty given point in time, some firms are adjusting by a large amount, many others are not adjusting at all. The overall level of investment may not change much from one period to the next.

At the same time, when aggregate shocks hit the economy, they may 'push' the entire distribution of firms to the left (leading to more investment) or to the right (leading to less investment): in such cases, aggregate investment too might spike, simply because the number of firms that are undergoing a large investment changes. In that case, the microeconomic lumpiness may matter for aggregate dynamics. What this suggests is that the aggregate dynamics themselves are likely to depend on the interaction between aggregate shocks and the cross section distribution of capital gaps Z_{it} across microeconomic units since this determines the mass of firms that are likely to adjust at a given point in time. If this distribution is uniform, then we'd expect aggregate investment to be quite smooth: at any point in time, there would only be a small number of firms close to the thresholds. Conversely, if all firms are identical, then we'd expect aggregate investment to be as lumpy as individual investment.

The cross sectional distribution of capital gaps is an empirical object, but it is also an endogenous one: that distribution will reflect the history of shocks that firms experience and how many of them adjust etc... In practice, we can characterize how this cross section distribution evolves over time, but to do so, it makes sense to have a somewhat more convexified representation of the decision process of each individual firm. In the model we studied above, firms either adjust by a large amount with probability one, or they don't adjust at all, also with a probability one. If the model is rigorously correct, one could, in principle recover the boundaries of the inaction range simply by observing firms' behavior. But in practice, there are many reasons why firms, with an identical capital gap, may decide to adjust, or not. Caballero and Engel allow for this by assuming that the fixed costs C_u, C_l that the firm faces are themselves random and i.i.d. (so that they do not themselves become a state variable). This means that at any given point in time, some firms with identical capital gaps may make different decisions in terms of adjustment and only a certain fraction of them will adjust, those for which the adjustment cost is low enough. For a given distribution of adjustment costs, this generates a convex 'hazard rate', i.e. a convex probability of adjustment, as a function of the capital gap.

To explore this idea further, consider the case where $c_u = c_l = 0$ (i.e. there are no variable costs) and $C_u = C_l = C$ (i.e. the fixed costs are the same when adjusting up or down). Going back through the steps we followed in section 5.2.4, we see that there is a unique return point c such that the value function satisfies:

$$v(L) = v(U) = v(c) - C$$

Suppose now that C is randomly distributed with a distribution G, iid across microeconomic units and over time. We can define the largest fixed cost $\Omega(Z)$ such that the firm with a capital gap Z would decide to adjust. It satisfies:[15]

$$\Omega(Z) = v(c) - v(Z)$$

We can then define a adjustment hazard as the proportion of firms with a capital gap Z that are adjusting:

$$\Lambda(Z) = G(\Omega(Z))$$

The average investment of firms with capital gap Z is then simply obtained as:

$$I_t(Z) = \Lambda(Z)(\frac{c}{Z} - 1)\bar{K}_t(Z)$$

where $\bar{K}_t(Z)$ is the average capital stock of firms with a capital gap Z at time t.[16]

If we define the cross sectional density of firms by $f(Z, t)$, then we can express aggregate investment as:

$$I_t = \int (\frac{c}{Z} - 1)\bar{K}_t(Z)\Lambda(Z)f(Z, t)dZ$$

Caballero and Engel use this expression to approximate the aggregate investment rate as:

$$\frac{I_t}{K_t} = \int (\frac{c}{Z} - 1)\Lambda(Z)f(Z, t)dZ$$

[15] Conversely, the inverse of this function defines an upper and lower bounds $L(C)$ and $U(C)$ such that a firm experiencing a realization of its adjustment cost C will only adjust if it finds itself outside the range $[L(C), U(C)]$.

[16] Observe that the investment of a firm with capital stock K_t that invests is $(c - Z)K_t^* = (c/Z - 1)K_t$.

The model can be closed by characterizing the law of motion of the cross section distribution, $f(Z,t)$. Assume that the productivity shock θ_t consist of an aggregate component, common to all microeconomic units, and an idiosyncratic shock: $\theta_t = \bar{\theta}_t \epsilon_t$. To characterize this law of motion, we need to be specific with the timing of events. Let's assume that starting with last period's cross section distribution $f(Z,t-1)$, firms first receive the aggregate shock $\bar{\theta}_t$. This shifts the desired capital stock K_t^* and therefore shifts the capital gap, conditional on an initial capital stock Z: $Z_t' = Z_t(\epsilon_t \bar{\theta}_t / \bar{\theta}_{t-1})^{1/(1-\alpha)}$. Next comes the adjustment, if $C < \Omega(Z')$. Finally, firms get a new draw of their adjustment cost C'. [To be continued]

While these models suggest that non-convex adjustment costs *can* play an important role, it does not guarantee that it does. Thomas (JPE 2002) shows that there can be some powerful general equilibrium forces that impose that the aggregate implications of lumpy investment dynamics at the micro level can remain negligible, even in a framework similar to the one just studied. In her paper, this result obtains because in a closed economy wages and interest rates are not constant and will respond to changes in investment demand. For instance, if consumers wish to smooth consumption, they may not be willing to increase savings when investment spikes because of a mass of firms want to invest. Instead, what happens is that the interest rate rises. This increase in interest rates deters firm from increasing their investment. If the consumption smoothing motive is strong, this means aggregate investment must remain smooth too. This is an interesting result, but it is also a *model* result. It does not prove that investment lumpiness does not matter. It simply shows a closed economy environment where it does not matter.

5.6 Application

One application of the framework with non-convex adjustment costs is to the question posed by Hsieh and Klenow in their 2009 QJE paper: why do standards of living vary so much across countries? The answer, according to Hsieh and Klenow is that resources may be more misallocated in some countries than others and this misallocation lowers overall aggregate productivity. The key insight in that paper was that revenue productivity should be equated across firms in a world without distortions. Using plant level data to measure revenue TFP (or TFPR in the lingo), they found significant evidence of misallocation in China and India, as compared to the U.S.

Consider a firm i producing a differentiated variety of a good, with production function

$$Y_i = A_i K_i^{\alpha} L_i^{1-\alpha}$$

where A_i is the firm-level productivity and α is the capital share, common to all firms in that country/sector. The firm faces a demand for its differentiated product with an elasticity $\sigma > 1$ (this can arise from a model of monopolistic competition with Dixit-Stiglitz preferences over varieties). The profits of the firm are:

$$\pi_i = (1 - \tau_i^Y) P_i Y_i - w L_i - (1 + \tau_i^K) r_K K_i$$

where τ_i^Y is an output distortion, specific to the firm, while τ_i^K is a capital distortion, also firm specific. τ_i^Y distorts symmetrically the demand for capital and labor. τ_i^K distorts the demand for capital relative to labor. w is the real wage and r_K is the rental rate of capital.

Assuming there are no adjustment costs to capital and labor, the optimization problem solved by the firm yields the following prices and quantities:

$$
\begin{aligned}
P_i &= \frac{\sigma}{\sigma-1}\left(\frac{r_K}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{(1+\tau_i^K)^{\alpha}}{A_i(1-\tau_i^Y)} \\
\frac{K_i}{L_i} &= \frac{\alpha}{1-\alpha}\frac{w}{r_K}\frac{1}{1+\tau_i^K} \\
MRPL_i &= w\frac{1}{1-\tau_i^Y} \\
MRPK_i &= r_K\frac{1+\tau_i^K}{1-\tau_i^Y}
\end{aligned}
$$

where $MRPL_i$ is the marginal revenue product of labor and $MRPK_i$ is the marginal revenue product of capital, defined respectively as $P_i \times MPL_i$ and $P_i \times MPK_i$. Notice that the last two equations indicate that the marginal revenue products net of taxes are equated across firms. Hsieh and Klenow then emphasize the difference between physical productivity, also denoted $TFPQ$, and revenue productivity, $TFPR$:

$$
\begin{aligned}
TFPQ_i &= \frac{Y_i}{K_i^{\alpha}L_i^{1-\alpha}} = A_i \\
TFPR_i &= \frac{P_iY_i}{K_i^{\alpha}L_i^{1-\alpha}} \propto (MRPK_i)^{\alpha}(MRPL_i)^{1-\alpha} \propto \frac{(1+\tau_i^K)^{\alpha}}{1-\tau_i^Y}
\end{aligned}
$$

What these formula indicate is that, while $TFPQ$ varies if firms' productivity differs, $TFPR$ varies only because of the distortions. Variations in $TFPQ$ are absorbed into factor demands, so that revenue products are equated, in the absence of distortions. Intuitively, more productive firms will demand more capital and more labor, up to the point where $MRPL$ and $MRPK$ are the same as that of other firms.

Hsieh and Klenow construct estimates of τ_i^K and τ_i^Y using:

$$
\begin{aligned}
1+\tau_i^K &= \frac{\alpha}{1-\alpha}\frac{wL_i}{r_KK_i} \\
1-\tau_i^Y &= \frac{\sigma}{\sigma-1}\frac{wL_i}{(1-\alpha)P_iY_i}
\end{aligned}
$$

The first equation imputes a capital distortion when the ratio of firm-level compensation of labor to compensation of capital exceeds the industry one. The second equation imputes an output distortion when the labor share is low relative to the industry one (adjusted for rents).

Hsieh and Klenow find that the cross sectional distribution of TFPR is larger in China and India, compared to the U.S. and quantify the impact this misallocation has on aggregate TFP. One question is where this variation in TFPR is coming from. One possible interpretation is that it captures permanent distortions in these economies: informal frictions, low enforcement of property rights, poor infrastructure, credit frictions, political frictions..... This is a structural view.

Asker et al (JPE 2014) propose an alternative explanation: the dispersion in the cross section distribution of $TFPR$ could simply reflect the presence of adjustment costs. Adjustment costs imply that marginal revenue products are not equated at any given point in time, either because of convex or non-convex adjustment costs. Consider two countries A and B with similar technology and similar adjustment costs. Suppose now that A faces a more volatile productivity process so that desired capital is more volatile. In that environment, the cross section of $TFPR$ is going to be larger in country A than country B (in models with convex adjustment costs, firms are chasing a target that moves faster; in models with non-convex adjustment costs, the inaction bands are larger so there is more dispersion in the cross section). In that interpretation, the dispersion arises naturally from a more volatile environment (or one with larger adjustment costs) due to the dynamic process of adjustment. Asker et al find reduced form evidence linking the dispersion in $MRPK$ and the volatility of $TFPQ$. The conclusion from Asker et al is that the observed dispersion in $TFPR$ may be ‘efficient’ (to the extent that the adjustment costs reflect technological frictions) and not distortions.