Optimal Learning, Overvaluation and Overinvestment

Barbosa, António

ISCTE - Instituto Universitário de Lisboa

September 2019

Optimal Learning, Overvaluation and Overinvestment

António Barbosa*

Department of Finance

ISCTE-IUL - Instituto Universitário de Lisboa

This version: September 2019

Abstract

Periods of technological revolution are usually associated with overvaluation of and overinvestment by innovating firms. This paper develops a model that explains this behavior in a frictionless rational setting. When fully rational innovating firms face uncertainty about the returns to scale of their production functions, overinvestment emerges as the optimal way to learn about the returns to scale. The optimal learning is also shown to produce overvaluation. The model is also able to generate what an observer ex-post would identify as bubbles followed by overcorrection, negative excess returns in early periods, negative autocorrelation in excess returns and market-to-book and size effects.

*ISCTE Business School, Av. Forças Armadas, 1649-026 Lisbon, Portugal. Phone: +351 21 790 39 16. E-mail: antonio.barbosa@iscte.pt

1 Introduction

Historical evidence suggests that technological revolutions go hand in hand with overvaluation and overinvestment in the new technology sector. After an initial surge, both the stock price and the stock of capital of the new technology firms gradually converge to their long-run values. The electronics (early 1960's), biotechnological (early 1980's) and IT (late 1990's) booms are just the most recent examples of such episodes, which also include the railroad and electric power booms of the mid nineteenth and early twentieth centuries (see surveys by Malkiel, 1996 and Perez, 2003).[1]

The bubble-like pattern in the stock prices of new technology firms is commonly associated to market irrationality (e.g., Shiller, 2000 and Perez, 2003). For instance, Shiller (2000) advocates that systematic optimistic expectations regarding the new economy create "irrational exuberance" leading to stock price bubbles and subsequent burst when those expectations are corrected.

In turn, the branch of the literature that attempts to rationalize financial bubbles, not necessarily originating from a technological revolution, usually does so by relying on some form of market imperfection. For example, Allen et al. (1993) uses the lack of common knowledge, Abreu and Brunnermeier (2003) the lack of synchronization between rational arbitrageurs, Allen and Gorton (1993) a principal-agent problem between investor and fund managers and Allen and Gale (2000) a risk-shifting problem between stock market investors and credit providers to rationalize bubbles in the stock market. Although technological revolution is not a key ingredient in the explanation of financial bubbles in these models, they can act as a trigger, by introducing more uncertainty to the stock market.

The model developed in this paper departs from those in the main branch of the literature in that it explains the overvaluation and overinvestment by innovating firms without appealing to irrationality or market imperfections. In the model, of the partial equilibrium type, a fully rational innovating firm chooses its stock of capital in order to maximize the expected discounted sum of its cash flows. The key feature of the model is that the firm, due to the innovative nature of its activity, does not know a priori the value of the returns to scale parameter of its production function (α). Instead, the firm has to learn the value of α from the observation of the stream of cash flows it produces, which is a non-trivial task, given the existence of random productivity shocks that add noise to the information provided by the cash flows.

It is this uncertainty about α and subsequent optimal learning of its value that allows the model to generate overvaluation and overinvestment relatively to the full information (fundamental) values obtained in the same state of nature (same previous period stock of capital and α equal to the firm's belief), even if the firm's beliefs are correct. The intuition is straigthforward. First, given that both the full information firm's value and marginal value of capital are convex in α and that the production function has decreasing returns to scale, the existence of uncertainty about α alone implies that overvaluation and overinvestment obtain due to the Jensen's inequality, in what

[1] Given the unobservable nature of fundamental values, it is not consensual whether some of these episodes correspond to true bubbles or not (see for instance the Rappoport and White, 1993 and references therein). This debate, however, is outside the scope of this paper.

constitutes the convexity effect. Second, the optimization of the learning process generates further overvaluation but a mixed effect on the investment. On one hand, the firm has an incentive to speed up the learning of the value of α because the less uncertainty about its value, the closer the firm can get to choosing the full information stock of capital in the future and thus receive the optimal stream of cash flows. This has a positive impact on the future firm's value, regardless of what the value of α may be, and thus also on the firm's current value. On the other hand, the firm has an incentive to slow down the learning process because the less uncertainty about the value of α, the weaker the convexity effect which pushes the firm's value down. These effects correspond to the learning and strategic effects, respectively. Since the firm can only speed up (slow down) the learning process by investing more (less), which increases (decreases) the signal to noise ratio in the cash flow information, the learning effect contributes to overinvestment whereas the strategic effect to underinvestment.[2] It is found that in general the learning effect dominates and so overinvestment obtains, but underinvestment cannot be ruled out.

The strategic effect is not only able to explain why the optimal learning of α can sometimes generate underinvestment, but it is also crucial to obtain overvaluation when one takes the fundamental value as being the full information market-to-book ratio when the stock of capital is optimal instead of when the stock of capital is, suboptimally, the same as in the parameter uncertainty case. When the stategic effect is weaker, the larger stocks of capital and firm's values that obtain in the case of parameter uncertainty compensate each other, implying little to no MB overvaluation or even undervaluation. Only when the strategic effect is stronger and overinvestment is restricted, does the model produce MB overvaluation.

In addition to this static picture of overvaluation and overinvestment, the time-varying nature of firm's beliefs and uncertainty about α allows the model to provide rich dynamics in the firm's beliefs, overvaluation, overinvestment and returns. The main findings for the case of unbiased initial beliefs, obtained by simulating histories of productivity schocks, are as follows. First, the learning of α is very fast in the first 5 years after the firm's inception, both in terms of initial bias correction and reduction in the uncertainty about α, but quite slow after that. Second, both overinvestment and overvaluation relatively to the long-run (full information) values decrease very fast as time passes, disappearing in the first 3-5 years. Third, one-period excess returns in the first 3-5 years are significantly negative and about zero after that. As a consequence excess holding period returns are always negative. The pattern in the first years is remarkably consistent with what is found in the IPO literature (e.g. Ritter, 1991). Fourth, the cross sectional distributions of overvaluation, overinvestment and one-period excess returns are positively skewed, specially in the first years. Nevertheless, the odds are considerably favorable to detection of overvaluation and overinvestment in the initial years. The same naturally applies to the case of one-period excess returns. Fifth, excess returns display a considerable negative autocorrelation, documented in Fama and French, 1988, in the first years that gradually vanishes by year 5. Sixth, excess returns evidence a negative

[2] The strategic effect is consistent with the findings of Bertocchi and Spagat (1998) that experimentation can generate underinvestment when experimentation is more informative. In this model the strategic effect is stronger precisely in those circunstances.

correlation with previous period MB and firm's value. This implies that one can find what appears to be a MB and size effect as in Fama and French (1992). In this model those effects are nothing more than apparent since the firm's expected return is always equal to the constant riskless rate of return.

And what if the initial firm's beliefs are not correct? The effect of biases initial beliefs on the results is straightforward. Optimistic initial beliefs have an intuitive positive impact on overvaluation and overinvestment, and a negative impact on returns. The opposite is true if the initial beliefs are pessimistic. Given that correction of optimistic beliefs acts in the same way as reduction in the uncertainty about α, one obtain the same time series pattern of overvaluation, overinvestment and excess returns as in the unbiased case. They just are more pronounced. If the initial beliefs are pessimistic, it is found that those patterns are reversed only if the initial bias is large enough so that the bias effect dominates the bias effect. The negative autocorrelation in excess returns and their negative correlation with MB and firm's value are preserved, regardless of the initial bias.

One of the main contributions of the paper is the analysis of the impact of capital adjustment costs on these results. The impact of adjustment costs on the overvinvestment and overvaluation relatively to the full information values in the same state of nature is mixed: adjustment costs significantly reduce overinvestment, but have almost no effect on the overvaluation. The intuition is simple. Adjustment costs make the marginal cost of capital increasing in the stock of capital, whereas in their absence it was constant. As a result, a smaller increase of the stock of capital above its full information value is enough to equate the marginal value of capital to its marginal cost in the presence of adjustment costs, and overinvestment becomes smaller. In turn, the small impact of adjustment costs on overvaluation stems from the fact that adjustment costs have no effect on the long-run firm's value, the firm's value when the stock of capital reaches the level the firm would choose in the absence of adjustment costs, which makes for most of the firm's value.

In turn, the time series patterns of overvaluation and overinvestment obtained in the presence of adjustment costs depends on how overvaluation and overinvestment are measured. A naive observer takes the long-run values, which equal the full information values in the absence of adjustment costs, as being the fundamental values. The naive observer then identifies a pattern of initial underinvestment that eventually (if the adjustment costs are small enough) gives place to a modest overinvestment that gradually vanishes. This results from the sluggishness in the adjustment of the stock of capital introduced by adjustment costs. He also observes a pattern of MB overvaluation similar to the case of no adjustment costs, but with much larger and more persistent overvaluations. This is mainly due to a denominator effect induced by adjustment costs. Adjustment costs imply smaller initial stocks of capital but the firm's value remains mostly unchanged because the less valuable assets in-place component of the firm's value is partially compensated by a more valuable growth option component. However, the naive observer detects the same patterns even if the firm faces no uncertainty, thus failing to acknowledge that fundamental firms changes over time.

A sophisticated observer, using the full information values that would be obtained at the same point in time as fundamental values, finds completely different patterns. Initial periods are always characterized by overinvestment, as one would expect from the static analysis, which increases in the first couple of years and then gradually disappears. The initial periods are also associated with overvaluation, which is larger and more persistent than in the absence of adjustment costs, followed by a slight overcorrection motivated by excess capacity that gradually disappears as the firm slowly adjusts its stock of capital. All other results are robust to the introduction of adjustment costs and to whether the observer is naive or not.

Similar to this paper are the works of Johnson (2007) and Pástor and Veronesi (2009), which also use uncertainty about the prospects of the new technology to rationalize the overvaluation associated to technological revolutions. Johnson (2007) analyzes a general equilibrium model otherwise similar to the one used in this paper. A representative agent allocates his wealth between consumption and investment in each of the two sectors, the old economy with known returns to scale and a new economy with unknown returns to scale, and learns the returns to scale of the new economy from the observation of its cash flows. Johnson (2007) obtains overvaluation and overinvestment similar in magnitude to those obtained in this paper. However he does not consider the existence of adjustment costs and does not analyzes the time series dynamics of the firm's beliefs, overvaluation and overinvestment or the effect of biased firm's beliefs. This paper can thus be viewed as an extension of Johnson's (2007) results.

Pástor and Veronesi (2009) take on a different approach to the problem. They provide a general equilibrium model of a representative agent economy with two sectors, the old and new economy, where the representative agent implements the new technology in a small scale and learns about its productivity before deciding whether to irreversibly adopt it in substitution of the old technology or not. Initially, the small scale of the new economy and it small probability of adoption make the risk in the new economy mostly idionsincratic. As the representative agent learns that the new technology is more productive than the old one, the stock prices are pushed up by the good news, but pushed down as the increase in the probability of adoption shifts the risk from idiosincratic to systematic. At the beginning the former effect prevails but the latter prevails eventually and a boom and bust pattern in the stock prices is obtained when technological revolution occurs. Pástor and Veronesi's (2009) model, unlike this paper's model, is able to match the bubble's bust with large volatility. However their model is unable to match overvaluation with overinvestment, relies on the irreversibility of the technological change to obtain the results and is arguably more stylized than the one presented in this paper.

The remainder of the paper is organized as follows. Section I describes the model, the firm's problem and discusses the assumptions made. Section II provides the results obtained from the solution to the firm's problem without adjustment costs and section III the results obtained when adjustment costs are present. Section IV concludes. All details about the numerical solution to the firm's problem is presented in the appendix.

2 The Model

This section presents a description of the model of the firm as well as a discussion of all relevant assumptions. The model is of the partial equilibrium type, where a firm chooses the optimal level of its production factor in order to maximize the expected discounted sum of cash flows.[3] Time is discrete and indexed by t, and the firm operates in a risk-neutral economy with a constant discount factor.

2.1 Cash Flows

The firm's cash flow has three components: (i) operational cash flow; (ii) gross investment; (iii) adjustment cost.

The operational cash flow (F) is generated from a production function with capital (K) as the sole production factor. As usual, it is assumed that only the stock of capital in place at the beginning of period t (K_t) contributes to the end-of-period t operational cash flow (F_t). The only other factor contributing to F_t is a random productivity shock (z_t), known only at the end of the period. The operational cash flow is given by

$$F_t\left(K_t, z_t\right) = e^{\theta + z_t} e^{\alpha \frac{K_t^{\beta} - 1}{\beta}}. \tag{1}$$

The second term is a generalization of the one-factor Cobb-Douglas production function that is used e.g. by Johnson (2007). The parameter α determines the level of returns to scale, and $\beta < 0$ is a shape parameter which governs how close this production function is to a Cobb-Douglas production function with the same α.[4,5] The first exponential in (1) is a scaling factor, related to the productivity of capital. The parameter θ is related to the average productivity of capital, while $z_t \sim NIID\left(0, \sigma^2\right)$ is the random productivity shock. It is assumed that the firm needs, at all times, a minimum stock of capital of 1 to be operational.

The economic rate of depreciation is given by δ. It follows that period t gross investment (I_t) is

$$I_t\left(K_t, K_{t+1}\right) = K_{t+1} - (1 - \delta) K_t.$$

Finally, I assume that each time the firm adjusts the size of its stock of capital, it pays an adjustment cost (A). The time t adjustment cost (A_t), following e.g. Alti (2003), is given by

$$A_t\left(K_t, K_{t+1}\right) = \frac{\phi}{2} K_t \left(\frac{K_{t+1} - K_t}{K_t}\right)^2, \ \phi \geq 0. \tag{2}$$

[3] Johnson (2007)and Pástor and Veronesi (2009) use representative agent general equilibrium models.

[4] In fact, α does not truly represent the returns to scale like in a Cobb-Douglas production function, in the sense that the production function is homogeneous of degree α in the production factor. However, α is closely related to the returns to scale of the production function and will henceforth be loosely referred to as the returns to scale parameter.

[5] It is easy to verify that $\lim_{\beta \to 0} e^{\alpha \frac{K_t^{\beta} - 1}{\beta}} = K_t^{\alpha}$.

Time t cash flow (CF_t) is the sum of these three components,

$$CF_t\left(K_t, K_{t+1}, z_t\right) = F_t\left(K_t, z_t\right) - I_t\left(K_t, K_{t+1}\right) - A_t\left(K_t, K_{t+1}\right).$$

For simplicity the prices of capital and output are assumed to be constant and, without loss of generality, equal to 1. Therefore, K_t, F_t, I_t and A_t have the double interpretation of quantities and dollar values.

Since the focus is not on the financing aspect of the firm's problem, I simply assume that the firm is exclusively financed by equity, which is is costlessly adjusted in order to match the value of the stock of capital. This implies that the firm's book value equals the value of its stock of capital and also that the firm pays out the net cash flow (CF_t) at the end of each period.

2.2 Parameter Uncertainty and Learning

The key assumption of the model is that the firm knows ex-ante all the functional forms and parameters related to the various components of cash flow, *with the exception of the returns to scale parameter* α.

Having an unknown α can be justified in the grounds of the innovative aspect of the firm's activity, either in the context of a broader technological revolution or of a firm level innovation. In both cases, the firm cannot rely on information about the returns to scale of similar firms or industries to identify its own returns to scale parameter. Instead, the firm has to make use of its own pre-investment knowledge about the innovative technology to form a prior belief on the value of α with a corresponding degree of uncertainty. After the initial investment is realized and the firm generates the first cash flows, new information arrives. The new information is then incorporated into the firm's beliefs by way of Bayesian updating. The weight given to the previous estimate of α and to the new information about α depends on the relative precisionss of each of these values (see Gill, 2002).

Rearranging the expression of operational cash flow (1), so that everything observable is on the left hand side, one obtains

$$S_t \equiv \frac{\ln F_t - \theta}{k_t} = \alpha + \frac{\sigma}{k_t}\varepsilon_t \tag{3}$$

where $k_t \equiv \frac{K_t^{\beta}-1}{\beta}$ and $\varepsilon_t \sim NIID\left(0,1\right)$. S_t is a summary variable for what is observed by the firm, and it will be referred to as the signal.

It is obvious that without the random productivity shocks the learning process of α would be trivial (the signal is noiseless). In that case the knowledge of the functional form and of the values of all other parameters of the operational cash flow function would allow the firm to exactly pin down the value of α after the observation of the first operational cash flow.

However, with a random productivity shock, the signal becomes noisy and the firm cannot recover the exact value of α just from one observation. Nonetheless, the noisy signal carries information about α. Knowing the distribution of the productivity shock, each observation of the signal is an observation of a Gaussian random process with unknown mean equal to α and standard deviation of $\frac{\sigma}{k_t}$. The mean of the signal process can then be estimated as an average of the observed signals. Following Bayes' rule, the optimal way to do this is to weight each of the observations by their precisions, defined as the inverse of the variance. Since the sum of normal random variables is also normally distributed, the assumption of a normal prior belief is useful, since the posterior distribution is normal (see Gill, 2002). Therefore, it is assumed that the firm's prior beliefs on α are normal with mean a_0 and variance v_0.

In this setup, the Bayesian update of beliefs is equivalent to the application of a Kalman filter to recover α from the noisy signal S_t. Direct application of the Kalman filter (see Liptser and Shiryaev, 2001) gives the following updating rules for a and v

$$a_{t+1} = \frac{\frac{1}{v_t}}{\frac{1}{v_t}+\left(\frac{k_{t+1}}{\sigma}\right)^2}a_t + \frac{\left(\frac{k_{t+1}}{\sigma}\right)^2}{\frac{1}{v_t}+\left(\frac{k_{t+1}}{\sigma}\right)^2}S_{t+1} = a_t + \frac{v_t}{v_t+\left(\frac{\sigma}{k_{t+1}}\right)^2}\left(S_{t+1}-a_t\right), \tag{4}$$

$$v_{t+1} = \frac{1}{\frac{1}{v_t}+\left(\frac{k_{t+1}}{\sigma}\right)^2} = v_t - \frac{v_t}{v_t+\left(\frac{\sigma}{k_{t+1}}\right)^2}v_t. \tag{5}$$

Equation (5) tells us two important things. First, the uncertainty about the estimate of α converges monotonically to zero. So, at some point in time the firm will (almost) know the exact value of α. Second, the larger the firm's stock of capital at a given point in time, the faster the uncertainty decreases. It is easy to see why this occurs. From equation (3), the larger the stock of capital, the less noise there is in the signal. This follows because with a small stock of capital, the difference between the operational cash flow obtained with different values of α is small enough to be disguised by the productivity shock. However, as the stock of capital increases, that difference increases and becomes more perceptible, despite the noise introduced by the productivity shock. Therefore, the larger the stock of capital the more informative is the signal, and so the more v_t decreases, that is, the faster is the learning of α.[6] This constitutes an *incentive for the firm to overinvest relative to the full information case* since the less uncertainty about α, the closer the firm can get to choosing the full information stock of capital and thus receive the optimal stream of cash flows.

Substituting equation (3) for S_{t+1} in equation (4), one obtains

$$a_{t+1} = a_t + \frac{v_t}{v_t+\left(\frac{\sigma}{k_{t+1}}\right)^2}\left(\alpha - a_t + \frac{\sigma}{k_{t+1}}\varepsilon_{t+1}\right). \tag{6}$$

We can see that, on average, with the observation of each signal, a_t is going to converge to α, that is, if a_t is below (above) α, on average the update is going to be positive (negative). However that convergence is not monotonic, due to the productivity shock which makes a_{t+1} a random variable.

[6] In the limit, as K goes to infinity, the noise goes down to zero and the firm can exactly observe the value of α. From equations (4) and (5), the uncertainty would reduce instantaneously to zero, and the estimate of α (a_t) would be updated to the value of the observed noiseless signal, α.

It is straightforward to derive the distribution of a_{t+1}. Recursive substitution of equations (4) and (5) into themselves yields

$$\begin{aligned}
a_{t+1} &= \frac{v_{t+1}}{v_0} a_0 + v_{t+1} \sum_{i=1}^{t+1} \left(\frac{k_i}{\sigma}\right)^2 S_i \\
v_{t+1} &= \frac{1}{\frac{1}{v_0} + \sum_{i=1}^{t+1} \left(\frac{k_i}{\sigma}\right)^2}.
\end{aligned}$$

Therefore, a_{t+1} is distributed as

$$\begin{aligned}
a_{t+1} &\sim N\left(\frac{v_{t+1}}{v_0} a_0 + \alpha v_{t+1} \sum_{i=1}^{t+1} \left(\frac{k_i}{\sigma}\right)^2, v_{t+1}^2 \sum_{i=1}^{t+1} \left(\frac{k_i}{\sigma}\right)^2\right) \\
&\sim N\left(\frac{v_{t+1}}{v_0} a_0 + \left(1 - \frac{v_{t+1}}{v_0}\right) \alpha, v_{t+1} \left(1 - \frac{v_{t+1}}{v_0}\right)\right) \qquad (7)
\end{aligned}$$

The expected value of a_{t+1} is a weighted average of the initial belief (a_0) and α, with the weight on a_0 given by the ratio of current to initial uncertainty about the true parameter value. Since v_{t+1} decreases monotonically, a_{t+1} is expected to converge to α. The expected percentage of the initial bias between a_0 and α that is covered at any point in time can be computed as

$$1 - \frac{E\left[a_{t+1} - \alpha\right]}{a_0 - \alpha} = 1 - \frac{\frac{v_{t+1}}{v_0} a_0 + \left(1 - \frac{v_{t+1}}{v_0}\right) \alpha - \alpha}{a_0 - \alpha} = 1 - \frac{v_{t+1}}{v_0}. \qquad (8)$$

It depends only on the ratio $\frac{v_{t+1}}{v_0}$ and is monotonically increasing with time.

The variance of a_{t+1} increases with time when $v_{t+1} > \frac{v_0}{2}$ and decreases when $v_{t+1} < \frac{v_0}{2}$. The last case is the most frequent one.

2.3 The Firm's Problem

The firm's objective is to maximize the expected discounted sum of future cash flows. To achieve that, in each period the firm chooses the stock of capital (K_{t+1}) that is going to be used to produce next period's output (F_{t+1}), given the values of the state variables: (i) the stock of capital in place at the beginning of the period (K_t); (ii) the current estimate of the value of α (a_t); and (iii) the current uncertainty about the estimate of α (v_t). As discussed before, the riskless discount factor is ρ. The firm faces no financial constraints when choosing its stock of capital.

The firm solves the following dynamic programming problem:

$$\begin{aligned}
V\left(K_t, a_t, v_t, z_t\right) &= \max_{K_{t+1}} \left\{CF_t\left(K_t, K_{t+1}, z_t\right) + \rho E_t\left[V\left(K_{t+1}, a_{t+1}, v_{t+1}, z_{t+1}\right)\right]\right\} \qquad (9) \\
s.t.\ a_{t+1} &= a_t + \frac{v_t}{v_t + \left(\frac{\sigma}{k_{t+1}}\right)^2} \left(S_{t+1} - a_t\right) \qquad (10) \\
v_{t+1} &= v_t - \frac{v_t}{v_t + \left(\frac{\sigma}{k_t+1}\right)^2} v_t. \qquad (11)
\end{aligned}$$

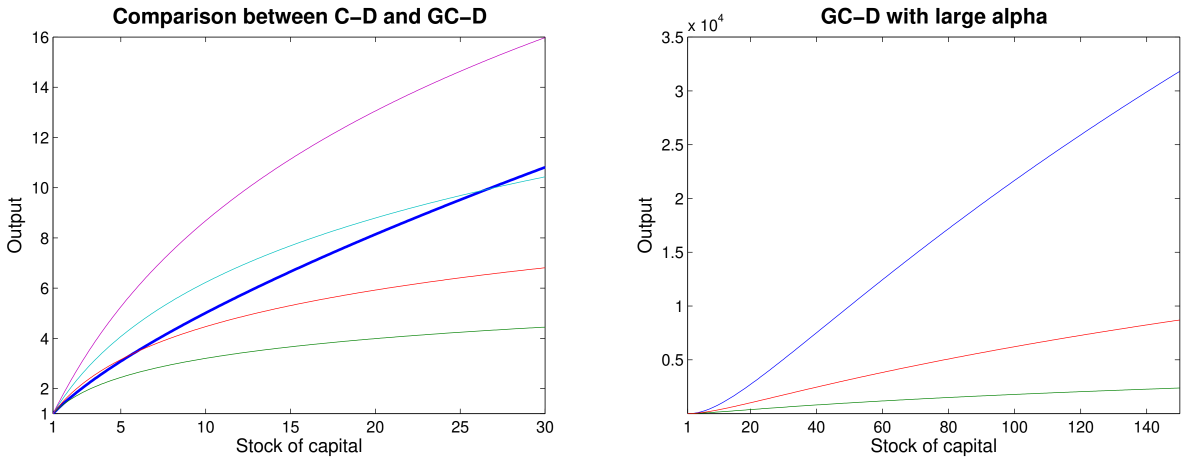


Figure 1: Properties of the GC-D production function. The left panel compares the GC-D production function with $\beta = -0.3$ and α ranging from 0.7 to 1.3 in 0.2 increments (thinner lines, with α increasing from bottom to top) with the C-D production function with $\alpha = 0.7$ (thicker line). The right panel plots the GC-D production function with $\beta = -0.3$ and $\alpha = \{3, 3.5, 4\}$ (α increasing from bottom to top)

Note that the expectation is taken using the firm's beliefs instead of the actual probabilities. Referring back to equation (6), the actual distribution of a_{t+1} depends on the current bias in the firm's beliefs. However, the firm has no way to assess the actual bias. Instead, the firm takes α as a random variable for which it has beliefs $\alpha \sim N(a_t, v_t)$. Then, from the point of view of the firm, $S_{t+1} \sim N\left(a_t, \left(\frac{\sigma}{k_{t+1}}\right)^2 + v_t\right)$ when in reality $S_{t+1} \sim N\left(\alpha, \left(\frac{\sigma}{k_{t+1}}\right)^2\right)$.

2.4 Discussion of Assumptions

Most of the aspects of the model described in the previous subsections are standard in the literature (e.g. Alti, 2003). However, some of the assumptions made deserve further explanation.

First, the Generalized Cobb-Douglas (GC-D) production function used here, although not novel (e.g. Johnson (2007)), is much less common than the mainstream Cobb-Douglas (*C-D*) production function. Its choice over the *C-D* production function was dictated mostly by the distributional assumption used for the beliefs. With a normal belief on the value of α, there is always the possibility of α being above 1, in which case one would obtain increasing returns to scale with a *C-D* production function. This is problematic because it is not economically sensible to have increasing returns to scale in the whole range of the production function, and also because it would generate an infinite firm's value and optimal stock of capital. To avoid this problem it would be necessary to use other distributional assumptions for the beliefs on α, like a truncated normal or uniform distribution. However, in both cases the conjugacy between the prior and posterior distributions is lost and the posterior distribution has no known form.[7]

The *GC-D* production function doesn't exhibit this problem. As can be seen from the right panel of figure 1, which plots the GC-D production function with $\beta = -0.3$ and $\alpha = \{3, 3.5, 4\}$, the GC-D production function exhibits decreasing returns to scale even with α significantly above 1.

[7] See Gill (2002) for a list of prior-signal distribution pairs for which conjugacy holds.

The left panel provides a comparison between the GC-D production function with different α's and the C-D production function with $\alpha = 0.7$.

Next, the requirement of a minimum stock of capital of 1 stems from the fact that the production function behaves differently for stocks of capital above and below 1. When the stock of capital is above 1, the higher the returns to scale parameter, the higher the output. However, the opposite is true for stocks of capital below 1. This feature, at best not very interesting economically, causes severe problems in the accuracy of the numerical approximation to the firm's value function. Hence, the lower bound of 1 for capital is justified.[8]

A more relevant assumption is that the firm knows everything about the model except the value of the parameter α. In particular, it may seem odd why the firm knows the mean (θ) and the standard deviation (σ) of the productivity shock and does not know the returns to scale parameter (α). Following Johnson (2007), the knowledge of θ and σ can be motivated by small-scale experimentation with a stock of capital (say, the minimum stock of capital) for which the differences between the operational cash flow obtained with different values of α is small (or null in the case of $K = 1$). For that reason, however, such small-scale experimentation is of no use to learn about α. Alternatively, if the productivity shock is not firm and sector specific, it can be learned from the observation of the productivity of other firms in different sectors. However, the returns to scale parameter is assumed to be sector specific, and so it cannot be learned from the observation of the returns to scale of other sectors.

Mine is a model of symmetric information where learning and value maximization endogenously generate overvaluation and overinvestment. This stands in contrast to a strand of the literature where informational asymmetries are the major driver of overvaluation. Nonetheless, the addition of informational asymmetries of information to my model should only contribute to reinforce overvaluation.

The model also excludes any kind of strategic interaction between firms, namely the strategic behavior of firms when they are able to free-ride on the externality created by the learning process undergone by similar firms, or when there is a first-mover advantage. The inclusion of such features would require a completely different and considerably more complex model. In this paper we can imagine that the firm is somehow able to protect the benefits it gets from the learning of its returns to scale: the information does not spillover; the firm is protected by patents; the products are differentiated; or the first-mover advantage offsets free-riding.

One aspect that is oversimplified in the model and may be worth exploiting in future work is the financing side of the firm's problem. Considering costs of external funding should have a similar effect in the results as the introduction of adjustment costs. Limited financing ability is expected to generate an obstacle to increases in the stock of capital, specially in the first years. This in turn would contribute to a smaller overinvestment and more persistent overvaluation, through more persistent growth options. In addition, the history of productivity shocks becomes more important, given its impact in the firm's autofinancing. But, above all, the introduction of a more realistic

[8] Johnson (2007)uses the first argument to justify the introduction of a similar minimum stock of capital.

financing side to the firm's problem will allow the analysis of dividend policies and capital structures. The drawback is that the added complexity may make it very hard to obtain numerical solutions to the model.

The choice of a partial equilibrium model of the firm over a general equilibrium model of a two sector economy as used in Johnson (2007) is justified by the simplicity and flexibility of the former. The simplicity of the model developed in this paper allows a clearer understanding of the mechanisms at play. The numerical solution to the problem of the firm is considerably easier than the solution of the problem of the representative agent, because in the first case there is only one control (investment) instead of two (investment in the new sector and consumption). In addition, firm's values in this model are obtained directly from the solution of the problem of the firm. By the contrary, in Johnson's (2007) model the solution of the problem of the representative agent yields expected utilities. Sector's values are obtained indirectly through numerical integration, which difficults the simulations needed for time series analysis, perhaps reason why Johnson (2007) only analyzes the time series of returns. Also, by modelling the firm instead of the economy, one obtains the cross sections of firms instead of cross sections of economies. Finally, although adjustment costs could have just as easily been introduced in Johnson's (2007) model, the same is not true for the extensions discussed above. The drawback in using the partial equilibrium model is that one loses the ability to analyze endogenously determined expected returns. However, Johnson (2007) does not take advantage of this feature.

3 Frictionless Solutions to the Firm's Problem

3.1 Parameter Values for the Solutions

The criterion for the choice of the parameter values was to ensure comparability of the results with those of Johnson (2007). Thus, the parameter values of this model, shown in table 1, were chosen to closely resemble those of Johnson (2007), with some adjustments needed to ensure convergence and reasonable precision of this model's numerical solution.

Thus, I used a slightly smaller β, to decrease the impact of increases in the returns to scale (α) on the production level. This yields more precise solutions. Since the smaller β also implies a lower level of production, I used a larger average productivity shock to compensate for that effect. A more volatile productivity shock was also used, which implies a slightly slower learning than in Johnson (2007). Nevertheless, upon experimentation with a σ of 20%, the impact of this larger σ on the results is small, whereas with the larger σ it is possible to obtain solutions to larger values of initial uncertainty with a more reasonable level of precision.

Finally, it should be stressed out that these parameter values, like those of Johnson (2007), are not a result of any specific calibration exercise. Starting at reasonable parameter values, the only experimentation that took place was trial and error to achieve convergence and an acceptable level of precision in the numerical solutions.

Parameter	Value
Riskless discount factor (ρ)	0.95
Depreciation rate (δ)	0.1
Average productivity shock (θ)	-1.445
Volatility of productivity shock (σ)	0.3
Shape of production function (β)	-0.3
Period of time (t)	1 year

Table 1: Parameter values used in the solutions.

3.2 The Full Information Case

From the last section, we know that the firm will, at some point in time, learn the true value of α. Once the value of α is known to the firm, it will choose the full information optimal stock of capital which will remain unchanged thereafter. This can be seen from the firm's problem described by equations (9) to (11). If at time τ the value of α becomes known then $a_t = \alpha$ and $v_t = 0$, $\forall t \geq \tau$ and the firm's problem reduces to

$$
\begin{aligned}
V(K_t, z_t) &= \max_{K_{t+1}} \{F_t(K_t, z_t) - I_t(K_t, K_{t+1}) - A_t(K_t, K_{t+1}) + \rho E_t[V(K_{t+1}, z_{t+1})]\} && (12)\\
&= F_t(K_t, z_t) + \max_{K_{t+1}} \{-I_t(K_t, K_{t+1}) - A_t(K_t, K_{t+1}) + \rho E_t[V(K_{t+1}, z_{t+1})]\}. && (13)
\end{aligned}
$$

Notice that the current realization of the productivity shock (z_t) has no impact on the firm's choice, since the productivity shock is a realization of an i.i.d. process, and as such the value of z_t gives no information about the value of z_{t+1}. Thus, the firm chooses K_{t+1} given only K_t and so the full information stock of capital is non-stochastic. As a consequence, the firm's value is also non-stochastic.[9]

Given that the optimal stock of capital and firm's value converge to the non-stochastic full information values, the full information values provide the perfect benchmark against which to measure the over(under)investment and over(under)valuation arising in the presence of parameter uncertainty. First, the full information firm's value is its fundamental value, and the full information stock of capital is the optimal stock of capital associated to the firm fundamentals. Second, the full information values are the ones that a market observer would, after α is known, use to compare the previous years' values, and judge the existence and extent of the overinvestment and overvaluation.

To solve the firm's problem and obtain the firm's value and optimal stocks of capital I used a numerical procedure described in the appendix. The left panel of figure 2 plots the average time t firm's value (V_t), including the current period's cash flow (CF_t), for different combinations of α and K_t.[10] It is not surprising that the firm's value is increasing in α and K_t, since both imply

[9] The exception is in the period of time between the knowledge of the current productivity shock's value and the distribution of the current cash flow. At all other points in time, the firm's value is the average over the realization of the future productivity shocks, which is non-stochastic.

[10] The average is taken over the current period's operational cash flow (F_t), which depends on the stochastic productivity shock (z_t). This is the same as the firm's value at the end of period t but immediately before knowing the realization of the productivity shock.

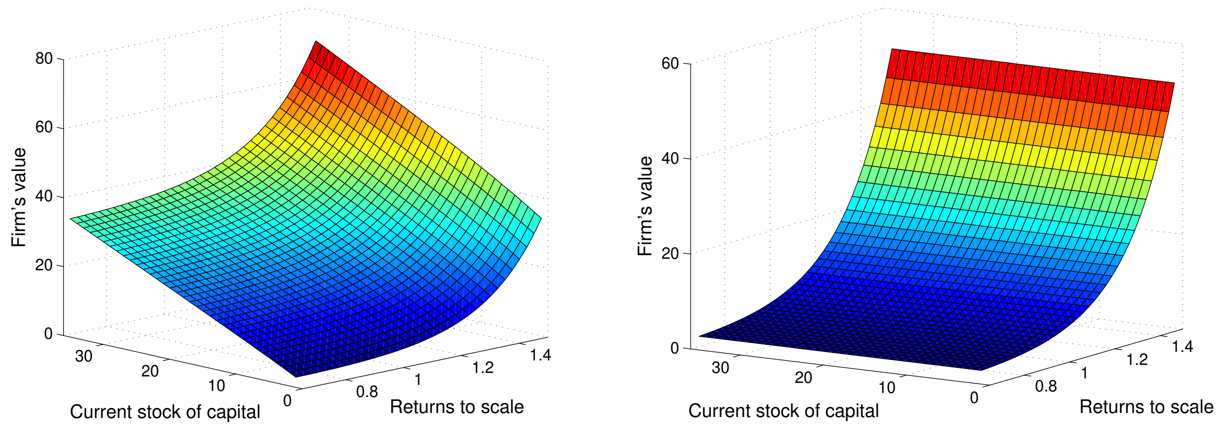


Figure 2: Full information firm's value. The full information firm's value is plotted for values of the returns to scale parameter (α) ranging from 0.65 to 1.5 and for values of the current stock of capital (K_t) ranging from 1 to 36. In the left panel the firm's value includes the current cash flow (CF_t), and the value is the average of the firm's value over the current productivity shock (z_t). In the right panel the firm's value excludes CF_t.

larger operational cash flows. The most interesting feature is that, for a given K_t, the firm's value is convex in α. This feature is at the heart of the overvaluation and overinvestment that one observes in the presence of parameter uncertainty. As we shall see in the next subsection, even without considering learning and strategic effects, parameter uncertainty generates overvaluation due to the Jensen's inequality.

Another important aspect is that the firm's choice of K_{t+1} is independent of K_t due to the absence of adjustment costs. This holds because the firm chooses K_{t+1} equating the marginal value of capital, $\rho \frac{\partial V_{t+1}}{\partial K_{t+1}}$, to the marginal cost of capital, $1 + \phi \left(\frac{K_{t+1} - K_t}{K_t} \right)$, which, in the absence of adjustment costs, reduces to 1 and eliminates the effect of K_t in the choice of K_{t+1}. As a result, V_{t+1} is the same regardless of the value of K_t, and so V_t only depends on K_t through CF_t. This can be seen from the right panel of figure 2 which plots the time t firm's value excluding CF_t.

Obviously, the larger the K_t, the larger the CF_t and V_t. The relation between V_t and K_t is almost linear but slightly concave. I_t is a linear function of K_t, but the decreasing returns to scale of the production function introduces a small amount of concavity through F_t.

From the above we can see that the absence of adjustment costs has in this model the usual implication: the value added by the growth option is exhausted once the firm revises its stock of capital, since the firm adjusts the stock of capital to its long-run value in just one period. This is an important feature that, as we will see shortly, prevents more persistence in the dynamics of the market-to-book ratio (MB) and returns when parameter uncertainty is taken into account.

The full information optimal stock of capital, optimal firm's value and optimal market-to-book ratio that an observer would use as a benchmark to identify and measure overinvestment and overvaluation are plotted in figure 3. Similarly to the firm's value, the optimal stock of capital (K_{t+1}) is also increasing and convex in α, as can be seen from the left panel. In fact, the shape of the optimal stock of capital plot is quite similar to the one of the optimal firm's value (firm's value

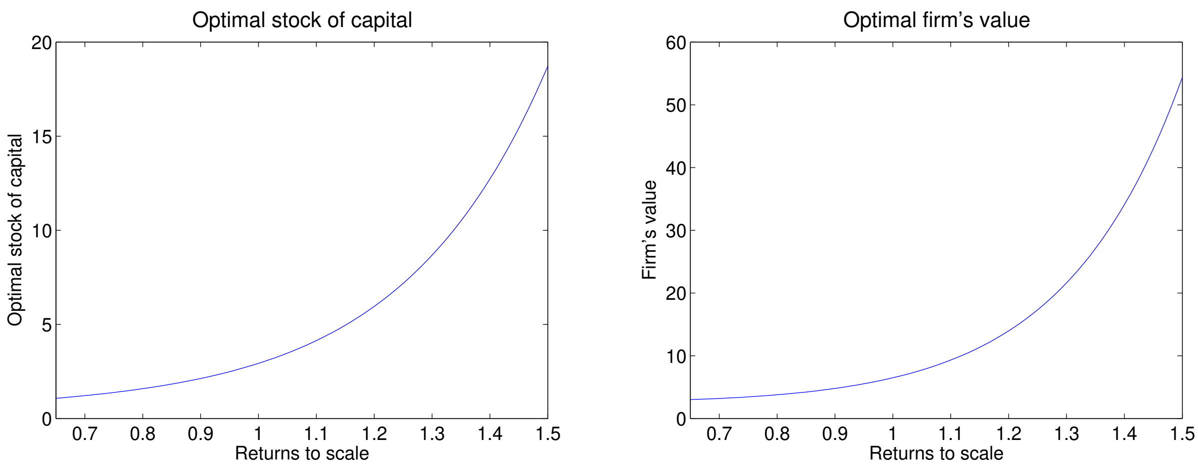


Figure 3: Full information optimal stock of capital, optimal firm's value and optimal market-to-book. The left panel plots the full information optimal stock of capital for values of α ranging from 0.65 to 1.5. The values plotted are independent of the previous period stock of capital due to the absence of adjustment costs. The center and right panels plot the full information firm's value and market-to-book ratio, respectively, after the payment of CF_t and the adjustment to the optimal stock of capital, as shown in the left panel.

immediately after the payment of CF_t and the adjustment of the stock of capital to its optimal value), as plotted in the center panel. The parabola shape of the optimal MB ratio is an important feature that, as we will see below, implies overvaluation when the bias, either positive or negative, in the firm's beliefs is large.

3.3 The Case of Parameter Uncertainty

The analysis to the solution of the firm's problem with parameter uncertainty is divided into two subsections. In the first I analyze the overvaluation and overinvestment observed at any given moment in time as well as the factors behind those observations. In the second, I consider the dynamics of the overvaluation, overinvestment and returns over time, using simulation. There I also analyze the cross-sectional distribution of the simulated variables of interest.

Once again, all details about the numerical procedure used to solve the firm's problem are provided in the appendix.

3.3.1 Static Analysis

From the discussion of the previous sections and the results of Johnson (2007) (albeit obtained with a different model), one should already expect that the addition of parameter uncertainty generates overvaluation and overinvestment relative to the full information values that would prevail in the same state of nature (that is, same K_t and $\alpha = a_t$ but $v_t = 0$). Three main effects are behind these results. The first is the convexity of the full information firm's value with respect to α (*convexity effect*). In the presence of parameter uncertainty, and ignoring other effects, the firm's value is the average over α of the corresponding full information firm's values with the same stock of capital. Then, by Jensen's inequality, the higher the uncertainty about α, the higher the firm's value relative to its full information value. Precisely the same happens with the marginal value of capital, since the full information marginal value of capital is also convex in α (as shown below). This, together with the decreasing returns to scale, implies that the higher the uncertainty about α, the larger the stock of capital that is needed to equate the expected marginal value of capital to its marginal cost, which is independent of α. This produces overinvestment relative to the full information case.

To verify the assertion that the full information marginal value of capital is also convex in α, recall that in the full information case and in the absence of adjustment costs, the stock of capital and cash flows from time $t+1$ afterwards are always the same (the latter only up to the random productivity shock). This implies that the continuation value, $\rho E_t\left[V\left(K_{t+1}\right)\right]$, can be computed as a perpetuity,

$$\rho E_t\left[V\left(K_{t+1}\right)\right]=\rho\frac{E_t\left[F_{t+1}\left(K_{t+1},z_{t+1}\right)-I_t\left(K_{t+1},K_{t+2}\right)\right]}{\frac{1}{\rho}-1}=\rho\frac{e^{\theta+\frac{\sigma^2}{2}}e^{\alpha\frac{K_{t+1}^{\beta}-1}{\beta}}-\delta K_{t+1}}{\frac{1}{\rho}-1}.$$

Then, the marginal value of capital (MVK), and its derivatives with respect to α are

$$MVK\equiv\rho E_t\left[\frac{\partial V\left(K_{t+1}\right)}{\partial K_{t+1}}\right]=\rho\frac{e^{\theta+\frac{\sigma^2}{2}}e^{\alpha\frac{K_{t+1}^{\beta}-1}{\beta}}\alpha K_{t+1}^{\beta-1}-\delta}{\frac{1}{\rho}-1}$$

$$\frac{\partial MVK}{\partial\alpha}=\rho\frac{e^{\theta+\frac{\sigma^2}{2}}e^{\alpha\frac{K_{t+1}^{\beta}-1}{\beta}}K_{t+1}^{\beta-1}}{\frac{1}{\rho}-1}\left(1+\alpha\frac{K_{t+1}^{\beta}-1}{\beta}\right)>0$$

$$\frac{\partial^2 MVK}{\partial\alpha^2}=\rho\frac{e^{\theta+\frac{\sigma^2}{2}}e^{\alpha\frac{K_{t+1}^{\beta}-1}{\beta}}K_{t+1}^{\beta-1}}{\frac{1}{\rho}-1}\frac{K_{t+1}^{\beta}-1}{\beta}\left(2+\alpha\frac{K_{t+1}^{\beta}-1}{\beta}\right)>0.$$

The second effect corresponds to the learning incentives (*learning effect*). As shown in the previous section, overinvestment provides the firm faster learning about α, which in turn contributes positively to the firm's value, since the firm can get closer from choosing the full information optimal stock of capital and thus receive the optimal stream of cash flows. However, this is true only because of the convexity of the firm's value and of the marginal value of capital. If both were linear in α, that would not be the case. The key to understand why the convexity is crucial for the existence of learning effects, lies in the fact that in the convex case the optimal stock of capital and firm's value are also a function of the level of uncertainty about α, whereas in the linear case all that matters is the estimate of α. As a result, in the linear case the firm's current value is independent of the next period's level of uncertainty about α and so there is no incentive to decrease the uncertainty or, in short, there is no learning effect at all. By the contrary, in the convex case, reductions in the uncertainty about α lead to revisions of the optimal stock of capital that place it closer to its optimal full information value. As a result, the next period's firm's value has the tendency to increase with reduction in the parameter uncertainty and so does the firm's current value, due to a chain effect. This implies that in the convex case there exists a learning effect: it is optimal to overinvest to some extent as a means to achieve faster learning which increases the firm's value. One can then conclude that the learning effect only exists together with the convexity effect, and that the former effect acts as an amplification of the latter.

But the reduction in parameter uncertainty has another effect in firm's value: it "weakens the convexity effect" which pushes the firm's current value down. That is, if on one hand the reduction in parameter uncertainty implies an increase in the next period's firm's values in the entire range of a_t values, on the other hand it implies that less weight be given to those firm's values corresponding to a_t values in the tails of the distribution associated to the firm's beliefs in the computation of the firm's current value.[11] The former obviously has a positive impact on the firm's current value, but the latter has a negative impact on the firm's current value due to the convexity of the firm's value to a_t/α relation. This negative impact to the reduction of parameter uncertainty in firm's value naturally gives the firm an incentive to strategically slow down the learning process, accomplished by decreasing investment, in order to preserve uncertainty and the associated overvaluation. This constitutes what I denominate as the *strategic effect*, which generates overvaluation and underinvestment. The strategic effect is consistent with the findings of Bertocchi and Spagat (1998) that "when more investment produces more information, experimentation can push towards a lower level of investment, thus reducing the information acquisition". As we will see later, in this model the strategic effect is stronger precisely when the cash flows are more informative.

Depending on which of the two effects dominates, the learning or the strategic effect, one may either observe over or underinvestment, respectively. As is apparent, the strategic effect can only subsist in the presence of convexity and so it is only a multiplicative effect.

The left panel of figure 4 plots the ratio of the firm's value with parameter uncertainty ($\sqrt{v_t}$) ranging from 5% to 25% in 5% steps and $K_t = 4.1438$, to the full information firm's value in the same state ($K_t = 4.1438$, $\alpha = a_t$ and $v_t = 0$).[12] It confirms the intuition built up above: (i) parameter uncertainty always implies overvaluation; (ii) for the same returns to scale parameter, the overvaluation is monotonically increasing in v_t; and (iii) in general, for the same level of parameter uncertainty, the larger the returns to scale parameter, the higher the overvaluation. This last observation follows from the fact that the convexity of firm's value with respect to α is increasing in α. Therefore the convexity effect is stronger, which, given the multiplicative nature of the learning and strategic effects, implies that these effects are also stronger (in the sense that they produce more overvaluation).

The right panel of figure 4 provides the same plot but for the case where the firm behaves myopically. That is, the firm does not take in consideration that its future beliefs will change as a response to the arrival of new information. Therefore, a myopic firm does not take in account the effect of its current actions on the precision of its future beliefs, which is the same as to say that there is no learning or strategic effect. We can see that there still exists overvaluation, attributable

[11] The later is what is loosely meant as a weaker convexity effect.

[12] $K_t = 4.1438$ is the full information optimal stock of capital when $\alpha = 1.1$, the value of the returns to scale parameter that will be used extensively next. The same plot with the full information and parameter uncertainty firm's values with K_t equal to either the full information optimal stock of capital or the optimal stock with parameter uncertainty are very similar in shape and magnitudes and so are omitted.

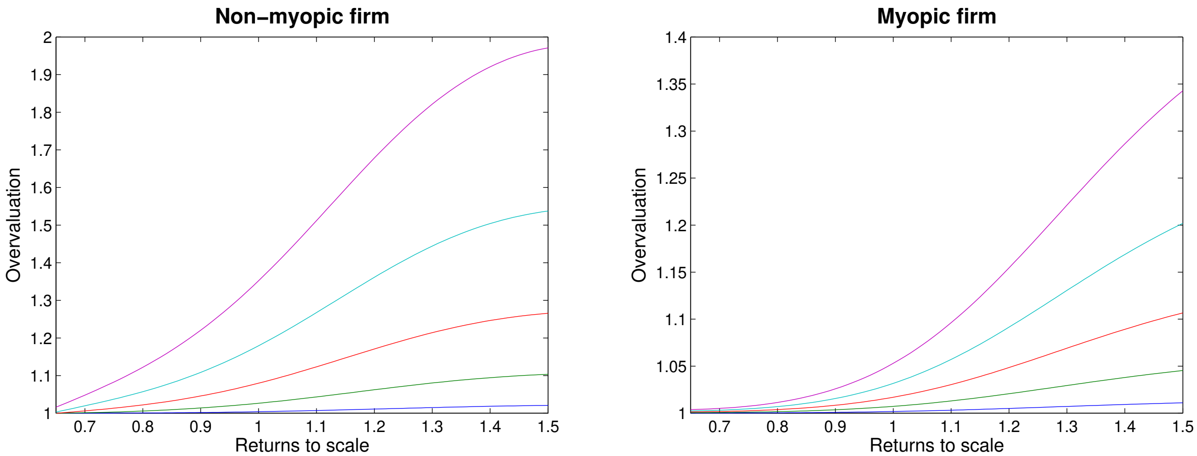


Figure 4: Overvaluation against the full information firm's value in the same state (variable level of parameter uncertainty, fixed K_t). The figure plots the ratio of the firm's value with parameter uncertainty to the full information firm's value in the same state [same current stock of capital (K_t) and α equal to its current estimate (a_t)], for values of a_t/α ranging from 0.65 to 1.5, $K_t = 4.1438$ (the full information optimal stock of capital when $\alpha = 1.1$) and level of parameter uncertainty ($\sqrt{v_t}$) ranging from 5% to 25% in 5% increments. Both the full information and the parameter uncertainty firm's values include the expected value of CF_t. In the left panel it is considered a non-myopic firm and so the overvaluation plotted includes the learning and strategic effects. In the right panel it is considered a myopic firm and so the overvaluation plotted is only a result of the convexity effect. In both panels, the outermost lines correspond to larger levels of parameter uncertainty.

only to the convexity effect, but to a much smaller extent than in the case of a non-myopic firm. This implies that learning and strategic effects are responsible for the majority of the overvaluation. Unfortunately, it is impossible to separate the learning from the strategic effect and thus to know which fraction of the extra overvaluation is attributed to each effect.

To complete the characterization of the overvaluation, it remains to see how does it respond to the current stock of capital. For that, the left panel of figure 5 presents a surface plot of the (relative) overvaluation for K_t ranging from 1 to 36, α ranging from 0.65 to 1.5 and $\sqrt{v_t}$ fixed at 25% (the shape of the surface plot is similar for different values of $\sqrt{v_t}$). From its observation we can see that the overvaluation is decreasing in K_t. This result follows because the value added by the parameter uncertainty is independent of K_t, as can be seen from the right panel (which plots the absolute overvaluation), and the firm's value is increasing in K_t. In turn, the independence between the value added by the parameter uncertainty and K_t is a result of the non existence of adjustment costs. As discussed in the previous section, the difference between the time t firm's values with different K_t is only in CF_t, because the firm chooses the same K_{t+1} regardless of K_t. Since the parameter uncertainty does not have any effect on CF_t, it follows that the value added by parameter uncertainty does not depend on K_t.

From the left panel of figure 5 one can also see that, although at the beginning the relative overvaluation is increasing in the belief about α, at some point this relation inverts. However, this only means that for lower values of α, the absolute overvaluation increases more rapidly than the full information firm's value but that after some point the opposite is true. In fact, looking at the right panel, we can see that the absolute overvaluation is not only increasing in α but also increasing

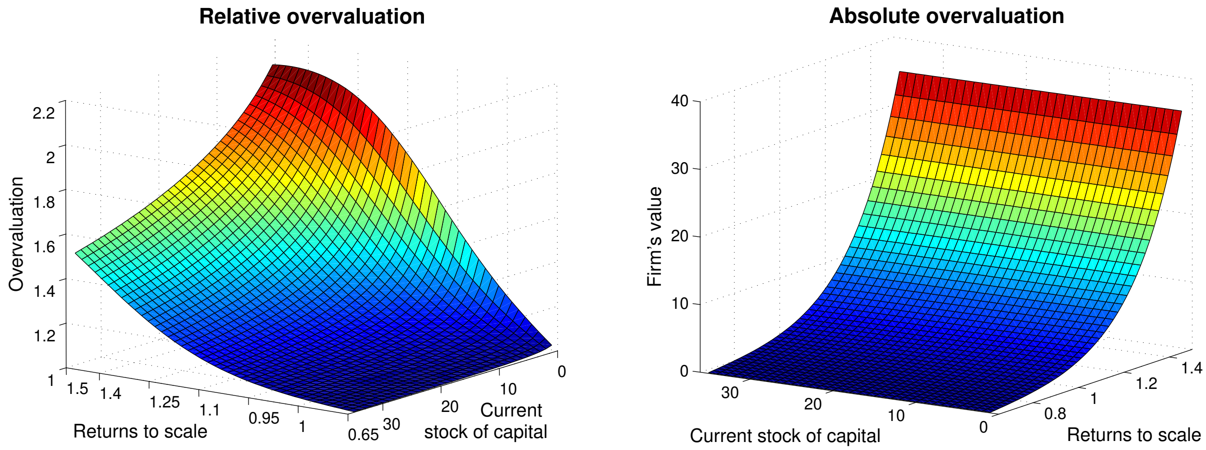


Figure 5: Overvaluation against the full information firm's value in the same state (fixed level of parameter uncertainty, variable K_t). In the left panel it is plotted the equivalent of the left panel of figure 3, with parameter uncertainty fixed at 25% and with K_t ranging from 1 to 36. In the right panel it is plotted the absolute overvaluation instead of the relative overvaluation plotted in the left panel. All the rest is equal to the figure in the left panel.

at an increasing rate. Therefore, very large values of α are sure not to produce the opposite results (undervaluation) as the values in the range considered here.

These results imply that if the firm has the choice between two similar technologies except maybe for the unknown α, for which it has the same expectation of the value of α, it prefers to implement the riskier one, the technology for which is has less certainty about the value of α.

How about the overinvestment? The left panel of figure 6 plots the ratio of the optimal stock of capital (K_{t+1}) with parameter uncertainty ($\sqrt{v_t}$) ranging from 5% to 25% in 5% steps, to the full information firm's value in the same state ($\alpha = a_t$ and $v_t = 0$).[13] The right panel provides a similar plot for the myopic case. We can observe that, for the range of α values considered here, parameter uncertainty produces overinvestment. Moreover, the learning effect is always stronger than the strategic effect and so the overinvestment is larger for a non-myopic firm than it is for a myopic firm. However, the effect of parameter uncertainty in the optimal stock of capital is not as straightforward as its effect on the firm's value, since in this case the strategic effect acts in the opposite direction of the other two effects. From the analysis of figure 6, it is possible to assess the strength of the strategic effect relatively to the other two effects: the strategic effect becomes stronger as (i) α increases and (ii) as the level of parameter uncertainty increases. In fact, when $a_t = \alpha = 1.5$ and $\sqrt{v_t} = 25\%$, the strategic effect almost dominates the learning effect. The first of these relations is explained by the fact that convexity is increasing in α. Since the larger the convexity, the larger the downward impact of reductions in parameter uncertainty on the firm's value, the strategic effect is stronger the larger the α. The second relation follows from the fact that the larger the parameter uncertainty, the more informative is the signal (observation of the operational cash flow), and so the larger the reduction in the parameter uncertainty after the observation of the signal, as can be seen from equation (5). In turn, for the same degree of

[13] Recall that without adjustment costs, K_{t+1} is independent of the initial stock of capital K_t.

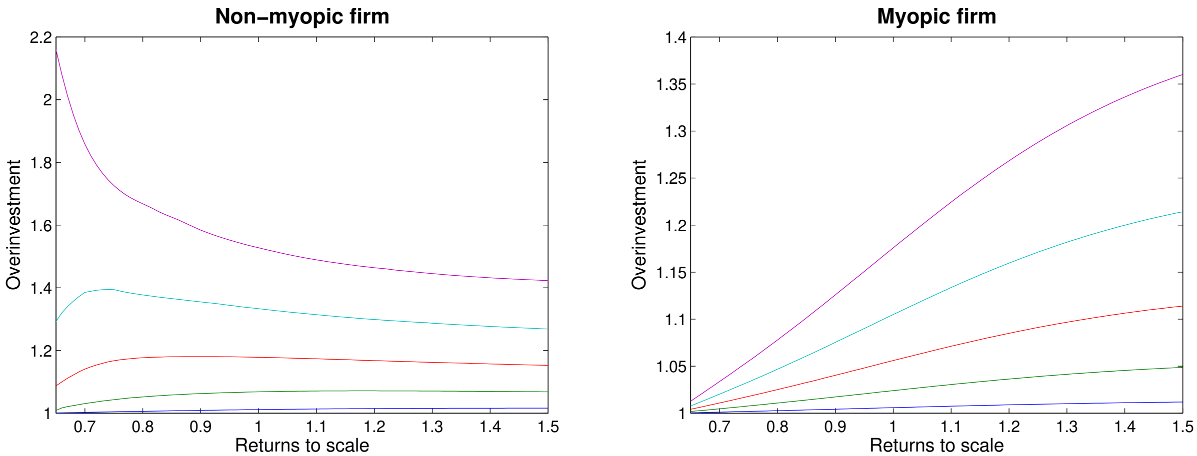


Figure 6: Overinvestment against the full information optimal stock of capital in the same state. The figure plots the ratio of the optimal stock of capital (K_{t+1}) with parameter uncertainty to its full information value in the same state, for values of a_t/α ranging from 0.65 to 1.5 and level of parameter uncertainty ($\sqrt{v_t}$) ranging from 5% to 25% in 5% increments. In the left panel a non-myopic firm is considered and so the overinvestment plotted includes learning and strategic effects. In the right panel a myopic firm is considered and so the overinvestment plotted is only a result of the convexity effect. In both panels, the values plotted are independent of the current stock of capital (K_t) due to the absence of adjustment costs. Also, in both panels, the outermost lines correspond to larger levels of parameter uncertainty.

convexity, the larger the reduction in the parameter uncertainty, the larger the reduction in firm value, which makes the strategic effect stronger.

Thus, it is perfectly possible that for a_t/α and v_t large enough the strategic effect dominates the learning effect and that a non-myopic firm overinvests less than a myopic one, or even that it underinvests. This is an interesting aspect of this model and that is not observed in Johnson (2007), where the underinvestment arises only due to budget constraints.

Up until now it was always assumed unbiased beliefs on the value of α. Thus, the overvaluation and, in general, the overinvestment observed are only a result of a direct effect of parameter uncertainty (disaggregated in convexity, learning and strategic effects) over the firm's value and optimal stock of capital. However, in the presence of parameter uncertainty only by chance will beliefs be unbiased. Since the relevant benchmarks by which over(under)valuation and over(under)investment are measured, are the full information values for the true α, the bias effect must also be taken into account. The bias effect is straightforward: optimistic beliefs about α increase overinvestment and overvaluation generated by the parameter uncertainty effect and pessimistic beliefs have the opposite effect. If the negative bias is large enough, one can even observe undervaluation and underinvestment.

Figure 7 provides a picture of the bias effect. It plots the overvaluation and overinvestment relative to the full information values when α is 1.1 (and K_t is 4.1438 in the case of the left panel), and thus shows the parameter uncertainty and the bias effect together.[14] From the comparison of these

[14]For other values of α the shape of both plots is essentially the same. The same is true of the plot in the left panel for different values of K_t, although the smaller the K_t, the sharper is the increase in the overvaluation as a response to an increase in the positive bias. Recall that the optimal stock of capital is independent of K_{t+1} and so the plot in the right panel is the same regardless of the value of K_t.

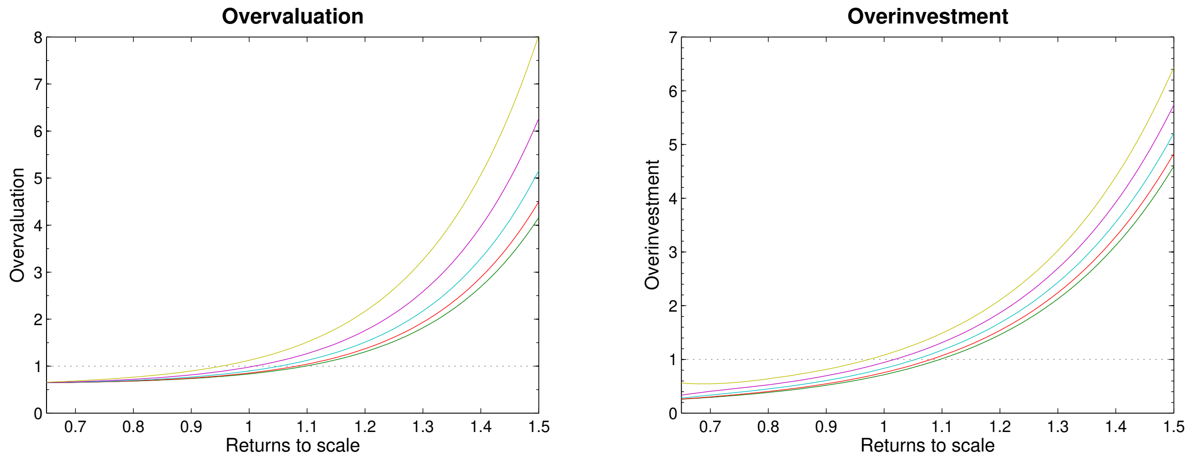


Figure 7: Overvaluation and overinvestment against the full information values with $\alpha = 1.1$. This figures plots the overvaluation (left panel) and overinvestment (right panel) measured as the ratio of the parameter uncertainty values to the respective full information values when $\alpha = 1.1$, therefore considering the bias effect in addition to the parameter uncertainty effect as in the left panel of figures 4 and 6, respectively. All the rest is as in those figures. In both panels, the outermost lines correspond to larger levels of parameter uncertainty.

figures with the left panels of figures 4 and 6 we can see that the bias effect is quite strong. Positive biases result in dramatic overvaluations and overinvestments. In turn, negative biases substantially reduce the overvaluation and overinvestment and, if large enough, produce undervaluation and underinvestment. Also, since on average positively (negatively) biased beliefs are followed by still positively (negatively) biased beliefs, although closer to the true value of α (see equation 7), the bias of the firm's initial belief is crucial to the over(under)valuation and over(under)investment observed ex-post. A larger initial positive bias will, on average, result in larger overvaluation and overinvestment over the years, and a larger initial negative bias will, on average, result in smaller overvaluation and overinvestment or larger undervaluation and underinvestment over the years. One curious aspect is that the range of beliefs for which there is overinvestment is slightly narrower (broader) than the one for which there is overvaluation when $\alpha > (<)\, 1$. This implies that ex-post overvaluation is observed more (less) frequently than overinvestment when $\alpha > (<)\, 1$.

Up to this point, the overvaluation has been measured against the full information (fundamental) firm's value for the same current stock of capital (K_t). However, there is a caveat in measuring the overvaluation in this way: it is assumed that the firm in the full information case has the same stock of capital as in the parameter uncertainty case, when in fact only at $t = 0$, when the stock of capital is 0 in both cases, is this true. Otherwise, one is comparing the firm's value with parameter uncertainty to the full information firm's value that would prevail if somehow the stock of capital had deviated from its optimal value, or to the firm's value that would prevail if the firm had just learned the value of α. More relevant is then to compare the MB in the case of parameter uncertainty with the optimal full information MB, which reflect different optimal stocks of capital. I shall refer to this as MB overvaluation to help the reader differentiate this definition of overvaluation from the one used before, where overvaluation is measured against the full information firm's value in the same state.[15]

In addition, this is essentially the way a naive observer would, ex-post, measure the overvaluation. He would pick the present day MB and use it as a benchmark against which to compare the past MB ratios. If the analysis is done sufficiently after $t = 0$, the MB is already close to its long-run value which, in the absence of adjustment costs, equals the full information MB that prevailed since $t = 0$.

In what follows, I will use the MB values computed immediately after the payment of the current cash flow (CF_t) and the adjustment of the stock of capital to its optimal value. The MB computed in this way has the advantage of depending only on the current beliefs. By the contrary, the MB computed immediately before the payment of CF_t and the adjustment of the stock of capital depends also on the current stock of capital, which can be virtually any value depending on the previous period's beliefs.

Figure 8 provides the picture of the MB overvaluation measured as described above. It plots the ratio between the MB with parameter uncertainty to the respective full information value, with $\sqrt{v_t}$ ranging from 5% to 25% in 5% increments and a_t ranging from 0.65 to 1.5. In the top (bottom) panels the case of a (non-)myopic firm is considered , and in the left (right) panels the case without (with) bias effect.

In the myopic case without bias effect (top left panel), one can see that MB over(under)valuations are almost meaningless, which agrees with the findings of Johnson (2007). The bias effect (top right panel), though, generates a significant MB overvaluation when the bias, either positive or negative, is large enough. This follows because the relationship between the MB and a_t inherits the parabola shape that it has in the full information case (figure 3), as is apparent. The parabolas are just shifted up or down depending on the value of α considered. The closer α is to 1, the larger the overvaluation for all the range of α.

Considering now the figures in the bottom panel, which also account for the learning and strategic effects, one can see that the additional MB overvaluation over the myopic case (convexity effect only) must be due to the strategic effect. The additional MB overvaluation is verified for higher a_t and v_t which is precisely when the strategic effect is stronger. In those cases, although the learning and strategic effects inflate firm values and stocks of capital, the increase in the latter is smaller than that in the former due to the strategic effect. As a result MB overvaluation increases. For smaller values of a_t, where the strategic effect is weaker, the learning and strategic effect imply a MB undervaluation, which means that the learning effect implies larger increases in the stocks of capital than in the firm values. This is essentially a result of the decreasing returns to scale.

Comparing the overvaluation measured in this way with the overvaluation as measured by the ratio of the firm's value for the same current stock of capital (K_t) prior to the payment of CF_t and the adjustment of the stock of capital (left panels of figures 4 and 6), we can see that most of the overvaluation is now gone. This ought to be expected, since in the absence of adjustment costs the value of the growth options disappears once the stock of capital is adjusted. Thus, in the absence

[15] Technically, overvaluation measured against the full information firm's value in the same state is also a MB overvaluation, since the stock of capital is the same in the parameter uncertainty and full information cases.

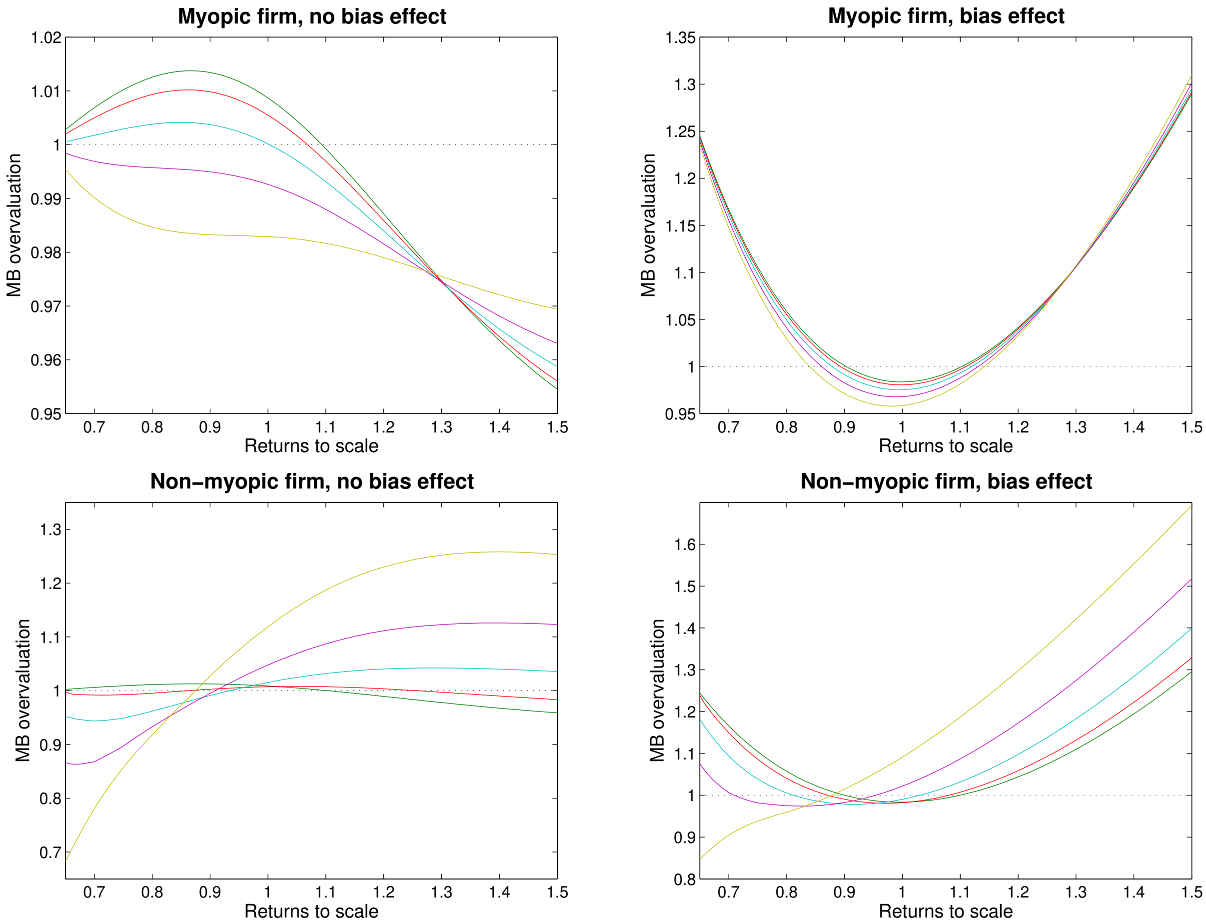


Figure 8: Market-to-book overvaluation. This figure plots the ratio of the MB with parameter uncertainty to the full information MB, with α ranging from 0.65 to 1.5 and $\sqrt{v_t}$ ranging from 5% to 25% in 5% increments. The time t MB is measured immediately after the payment of CF_t and the adjustment of the stock of capital. The top (bottom) panels plot the (non-)myopic case, and the left (right) panels exclude (include) the bias effect. In the right panels it is used $\alpha = 1.1$. In all panels, the outermost lines at $a_t/\alpha = 1.5$ correspond to larger levels of parameter uncertainty.

of adjustment costs, the overvaluation that one can expect an observer to detect ex-post is much more modest than what it looked like using the initial metric.

3.3.2 Dynamic Analysis

In this subsection the focus is on the time series dynamics of the firm's beliefs (a_t and v_t), optimal stock of capital (K_{t+1}), value (V_t), market-to-book (MB_t), one-period excess returns (r_t) and annualized excess holding period returns from time 0 (HPR_t). Both V_t and MB_t are computed immediately after the distribution of CF_t and the adjustment of the stock of capital. The time series of these values are generated through simulation. Given an initial set of beliefs (a_0, v_0) the firm chooses the initial optimal stock of capital, K_0, without any constraints. At each of the subsequent periods, a realization of the random productivity shock is drawn, the firm pays out the current cash flow and updates its beliefs, based on which it revises the stock of capital. The simulations span 40

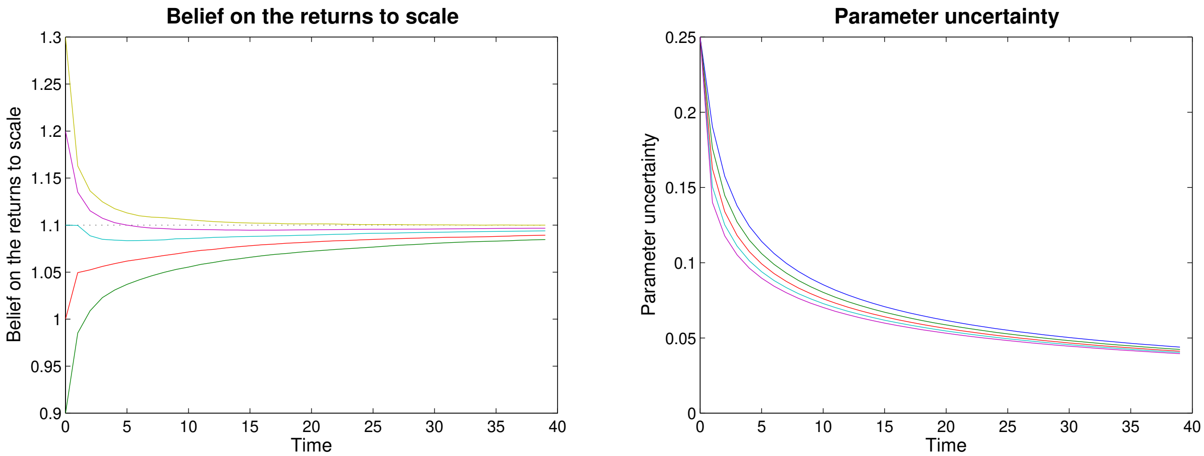


Figure 9: Time series of firm's beliefs. This figure plots the time series of the average a_t (left panel) and the average $\sqrt{v_t}$ (right panel) over 39 years for different initial beliefs. The average is taken over 10,000 simulations and the initial beliefs have a_0 ranging from 0.9 to 1.3 in 0.1 increments and $\sqrt{v_0} = 25\%$. The true value of α is 1.1. In the left (right) panel the outermost lines correspond to larger (smaller) initial beliefs.

periods/years (including year 0) and are repeated 10,000 times.

Firm's Beliefs Figure 9 provides the time series of average a_t and $\sqrt{v_t}$ over the 10,000 simulations, with $\alpha = 1.1$, $\sqrt{v_0} = 25\%$ and a_0 ranging from 0.8 to 1.3 in 0.1 steps. To begin with, consider the case where the initial firm's belief is unbiased ($a_0 = \alpha = 1.1$, which corresponds to the center line). It stands out immediately that the time series of average a_t is slightly biased downwards after year 2. It should be stressed out that this is not a result of the arguably small size of the simulation.[16] Looking at equation (6), it is obvious that $E[a_1] = \alpha$ since at time 1 everything except ε_{t+1} is equal across simulations. However this is not the case from period 2 onwards. The different realizations of z_t imply different beliefs which in turn imply different stocks of capital. In particular, the smaller a_t the smaller K_{t+1} is, as shown in the previous subsection, which means that next period's signal is not so informative. Thus, when $a_t > \alpha$ the convergence toward α (learning) is slightly faster than when $a_t < \alpha$, which explains the result.

When the initial beliefs are biased, the convergence of a_t to α is particularly fast in the first 3-5 years, when the parameter uncertainty is also large. After those initial years the convergence of a_t to α is much slower. This is explained by the exponential decrease of the parameter uncertainty with time, as reported by the right panel of figure 9, and by the Bayesian update of beliefs. As the parameter uncertainty decreases, the precision of the current belief increases relative to the precision of the new signal. As a result, the less informative is the new signal which, according to equation (4), implies that less weight be given to the new signal and thus the smaller is the belief's update. Since the parameter uncertainty decreases very fast in the first 5 years, the convergence of a_t to α slows down after that point in time.

As noted above, the parameter uncertainty decreases very fast in the first 5 years and much

[16] Even though the simulation size is only 10,000, it takes 8 hours to complete.

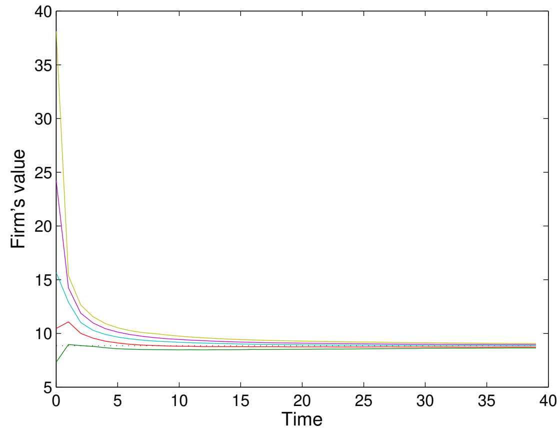


Figure 10: Time series of average firm's value. This figure plots the time series of the average firm's value (after the distribution of CF_t and the adjustment in the stock of capital) over 39 years for different initial beliefs. The average is taken over 10,000 simulations and the initial beliefs have a_0 ranging from 0.9 to 1.3 in 0.1 increments and $\sqrt{v_0} = 25\%$. The true value of α is 1.1. The outermost lines correspond to larger initial beliefs. The dashed line corresponds to the full information firm's value.

slower after that. When the initial belief is unbiased, the parameter uncertainty is reduced to about one half by the third year, but it takes 11 more years to reduce it further to one quarter of its initial value. This is also a result of the decrease in the information content of the new signals as parameter uncertainty decreases which, according to equation (5), implies that the less the parameter uncertainty decreases with observations of new signals. The precision of the signal, which depends only on the standard deviation of the productivity shock and on the current stock of capital, can be improved only by increasing the stock of capital. However, as the parameter uncertainty decreases, the learning incentive to overinvest decreases, and if anything the precision of the signal degrades.

When the initial beliefs are biased, the picture does not change by much. The only difference is that the larger the positive (negative) bias, the larger (smaller) the stock of capital which implies that the faster (slower) is the learning.

Overvaluation and Overinvestment Before analyzing the time series of MB overvaluation and overinvestment relative to the full information values, it is useful to give a look at the time series of average firm's values, plotted in figure 10. The plot should be read carefully, though. It does not give the correct picture of the overvaluation since the stocks of capital with parameter uncertainty are not the same as in the full information case.

Considering first the case where the initial beliefs are unbiased (center line), the pattern of convergence of the V_t to its full information value (depicted as the horizontal dashed line) is similar to the one observed for the convergence of $\sqrt{v_t}$. The convergence in the first 5 years is very fast, following from both the sharp reduction in $\sqrt{v_t}$ in that period of time and the sharp reduction of the overvaluation with the reduction of $\sqrt{v_t}$ (as can be seen from the left panel of figure 4). In the following years the convergence of V_t to its full information value is much slower. What may seem

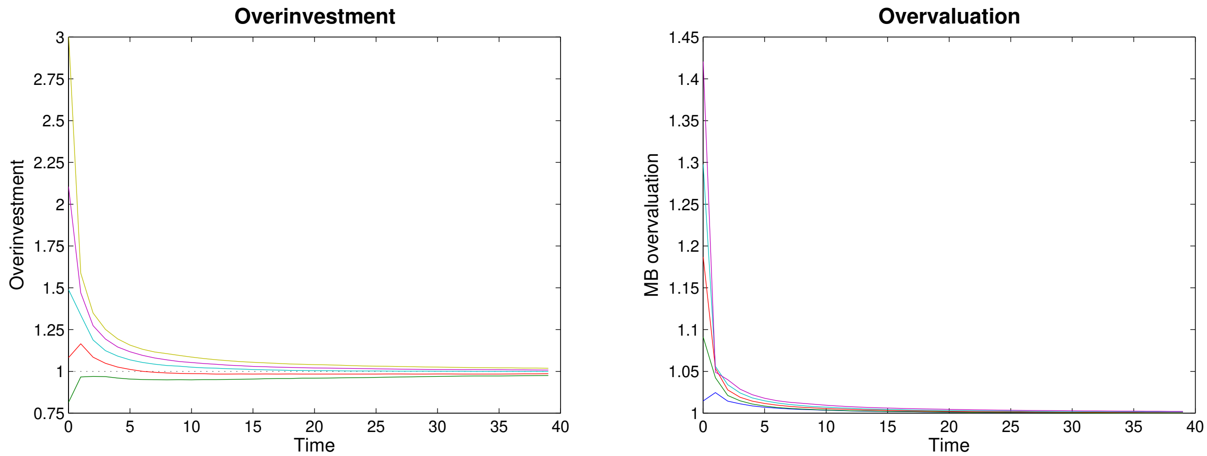


Figure 11: Time series of average overinvestment and market-to-book overvaluation. This figure plots the time series of the average overinvestment (left panel) and MB overvaluation (right panel) over 39 years for different initial beliefs. The average is taken over 10,000 simulations and the initial beliefs have a_0 ranging from 0.9 to 1.3 in 0.1 increments and $\sqrt{v_0} = 25\%$. The true value of α is 1.1. The outermost lines correspond to larger initial beliefs.

surprising is that after years 15-20 the firm's value is very close to the full information value when the parameter uncertainty is still a bit above 5% (one fifth of its initial value). However, as figure 6 shows, the overvaluation when $\sqrt{v_t} = 5\%$ is pretty much meaningless.

When the initial beliefs are positively biased, the convergence to the full information value in the first years is even sharper. Adding to the effect of the sharp decrease of the parameter uncertainty, the strong bias correction that occurs in the first 3-5 years contributes to the decrease in the firm's value. In contrast, when the initial beliefs are negatively biased, the bias correction and the decrease in the parameter uncertainty act in opposite directions. Depending on the relative strength of the bias and parameter uncertainty effects on the firm's value one may observe a monotonic convergence of the firm's value to its full information value from above (very slight negative bias), increase of the firm's value followed by a convergence to the full information value from above (moderate negative bias) or a monotonic convergence to the full information value from below (large negative bias).

Looking now at the time series of average overinvestment, plotted in the left panel of figure 11, one can see that, except for its scale, it is identical to the plot of the time series of average firm's values. Moreover, everything said above about the firm's value fits here entirely. In the case of unbiased initial beliefs, by year 5, the original overinvestment of 45% is drastically reduced to about 10%. After year 10 the overinvestment is at a meager 3%.

The right panel of figure 11 plots the time series of average MB overvaluation (the ratio between the average MB to its full information value). The MB overvaluation is more modest and ephemeral than what a reader who skipped the last paragraphs of the previous subsection would expect. But it fits nicely what was discussed there. The MB overvaluation is modest because it does not include the value of the growth option which is exhausted after the adjustment of the stock of capital in the absence of adjustment costs. In the unbiased case, if we observe any MB overvaluation at all, it is thanks to the strategic effect, as discussed previously. Since a strong strategic effect

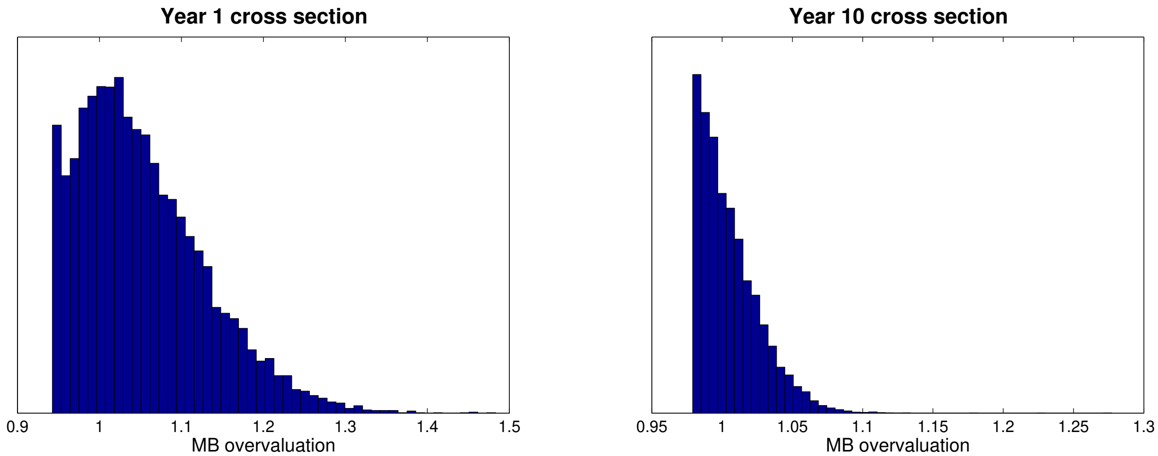


Figure 12: Cross section of market-to-book overvaluation. This figure plots the cross sectional distribution of MB overvaluation ratios, when the initial beliefs are unbiased ($a_0 = \alpha = 1.1$ and $\sqrt{v_0} = 25\%$), based on 10,000 simulations. The left panel shows the year 1 and the right panel the year 10 cross sectional distribution.

depends crucially on a high $\sqrt{v_t}$, the sharp decrease in $\sqrt{v_t}$ that occurs in the first couple of years implies that the strategic effect almost disappears. This results in a very fast vanishing of the overvaluation. With adjustment costs it is expected a significantly more persistent MB overvaluation. Not surprisingly, when the initial beliefs are biased, the bias effect contributes to larger (smaller) MB overvaluations when it is positive (negative).[17]

Next I look into the cross-sectional distribution of the MB overvaluation and overinvestment. Figure 12 shows the histogram of the year 1 and 10 MB overvaluation for the case of unbiased initial beliefs. The distinctive aspects are that the cross-sectional distribution of MB overvaluation is positively skewed and that the skewness increases with time (the decrease of the dispersion with time was obvious from the beginning). The skewness follows from the symmetric/slightly negatively skewed cross-sectional distribution of the firm's belief and from the parabola shape of the MB overvaluation to firm's belief relation as documented in the bottom right panel of figure 8.[18] The skewness increases with time because as parameter uncertainty falls, the firm's beliefs become more concentrated in the flat area of the parabola. The introduction of bias into the initial beliefs does not change these observations. However, negatively (positively) biased initial beliefs have a more (less) skewed cross-sectional distribution, because the firm's beliefs are more (less) concentrated near the flat area of the parabola.

Figure 13 shows the year 1 and 10 cross-sectional distribution of overinvestment for the case of unbiased initial beliefs.[19] The cross-sectional distribution of overinvestment is still negatively skewed, due to the convexity of the overinvestment to firm's belief relation (right panel of figure 7). However, the skewness decreases with time, following the reduction in parameter uncertainty.

The skewness of the cross-sectional distributions of both the overvaluation and overinvestment

[17]However, if the negative bias were large enough, one would observe larger MB overvaluations in the first years, due to the parabola shape of the MB overvaluation to a_t relation as depicted in the bottom right panel of figure 8.

[18]The cross-section distribution of the firm's belief is not reported, but it can be easily inferred from equation (6).

[19]The pictures for the case of negatively and positively biased initial beliefs are very similar.

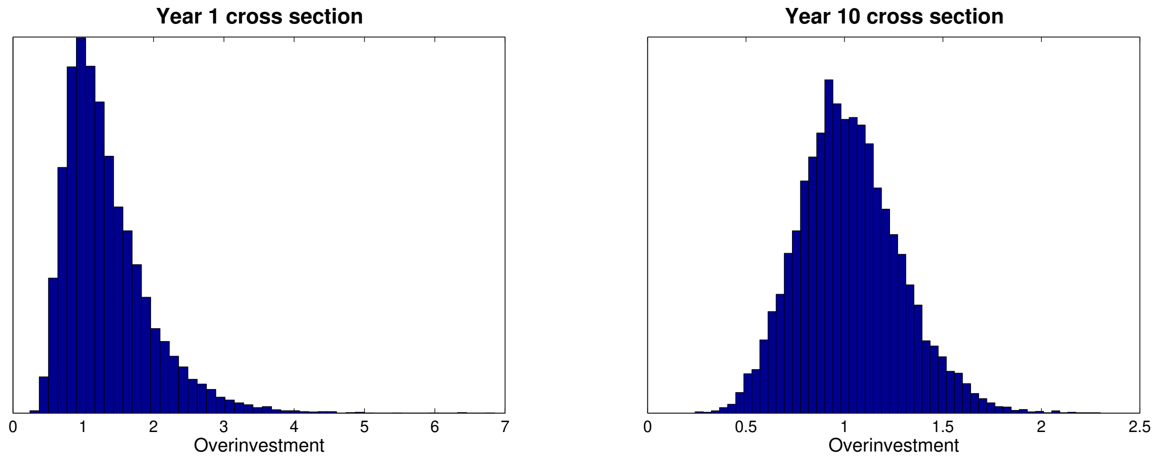


Figure 13: Cross section of market-to-book overvaluation. This figure plots the cross sectional distribution of MB overvaluation ratios, when the initial beliefs are unbiased ($a_0 = \alpha = 1.1$ and $\sqrt{v_0} = 25\%$), based on 10,000 simulations. The left panel shows the year 1 and the right panel the year 10 cross sectional distribution.

Year	1	2	3	5	10	20	39
MB Overvaluation							
$a_0 = 0.9$	56.5%	48.76%	47.0%	45.2%	43.0%	41.7%	42.8%
$a_0 = 1.1$	71.9%	63.4%	59.6%	55.0%	50.8%	48.1%	47.4%
$a_0 = 1.3$	67.6%	74.9%	71.2%	65.6%	60.0%	55.4%	52.3%
Overinvestment							
$a_0 = 0.9$	37.1%	37.9%	38.7%	39.9%	40.2%	41.9%	44.1%
$a_0 = 1.1$	66.8%	59.8%	57.1%	53.9%	51.3%	49.6%	49.4%
$a_0 = 1.3$	81.3%	74.3%	70.5%	65.9%	61.5%	57.6%	54.6%

Table 2: Probability of market-to-book overvaluation and overinvestment. The table shows the probabilities of detecting MB overvaluation (top panel) and overinvestment (bottom panel) against the full information values at a given year, based on 10,000 simulations. The value of α is 1.1, and the initial parameter uncertainty is $\sqrt{v_0} = 25\%$. The initial estimates of α are in the left column.

rises an interesting question: What is the probability of identifying overinvestment and overvaluation at a given year? The answer is provided in table 2.

In the case of unbiased initial beliefs, the odds are favorable to identify MB overvaluation and overinvestment in the first 5-10 years, despite the positive skewness of the cross-sectional distributions. In particular, the probability of detecting MB overvaluation and overinvestment in the first year is almost 3/4. As pointed out in the last subsection, the probability of detecting MB overvaluation is slightly larger than that of detecting overinvestment (this depends on the true value of α). Without surprise, positively (negatively) biased initial beliefs increase (decrease) the odds of detecting both MB overvaluation and overinvestment. However, the bias effect seems to have a stronger impact in the probability of detecting overinvestment. Curiously, in year 1 due to the parabola shape of the overvaluation to firm's belief relation, even with negatively biased initial beliefs the probability of overvaluation is higher than 50%. The complex relation between MB overvaluation, firm's belief and parameter uncertainty produces yet another curious result: positively biased initial beliefs may result in smaller probability of detecting MB overvaluation compared to the case of

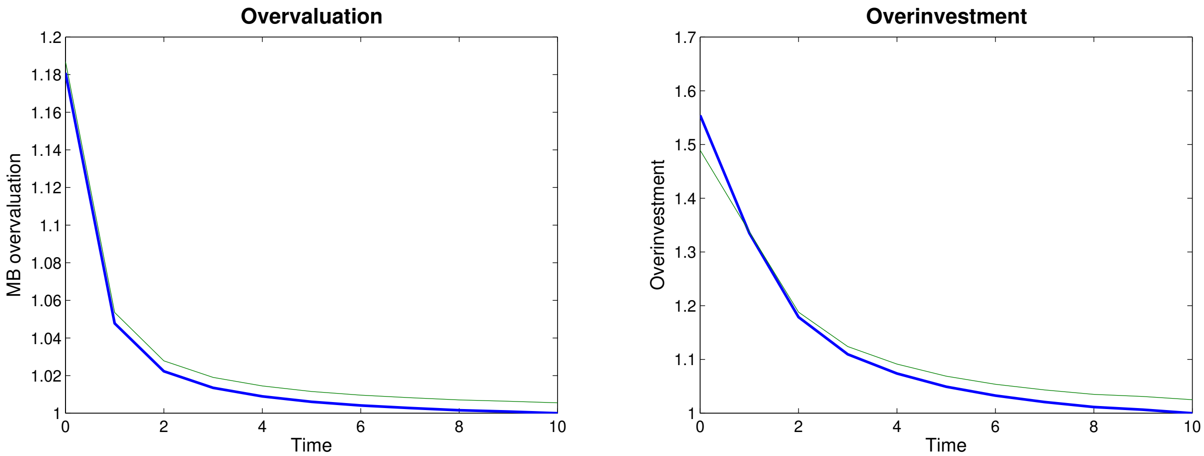


Figure 14: Time series of average market-to-book overvaluation and overinvestment measured at year 10. This figure plots the average MB overvaluation (left panel) and overinvestment (right panel) that is measured ex-post by comparing the past MB and stocks of capital with the year 10 values instead of their full information values, when the initial beliefs are unbiased ($a_0 = \alpha = 1.1$ and $\sqrt{v_0} = 25\%$). The thinner line corresponds to the overvaluation/overinvestment measured against the full information values.

unbiased initial beliefs.

Since even after 40 years the full information state is not achieved (the parameter uncertainty by that time is just under 5%), one may question if the MB overvaluation and overinvestment measured against the full information values are a good representation of what, on average, one would observe after say 10 years of the firm inception. To analyze this I measure, for each of the 10,000 simulations, the MB overvaluation and the overinvestment over the past years against the year 10 MB and stock of capital.

Figure 14 present the results for the unbiased case.[20] For ease of comparison, I also plotted the MB overvaluation and overinvestment against the full information values. The similarity between the actual measure of MB overvaluation and overinvestment and the one using the full information values is striking, even though at year 10 the parameter uncertainty is still around 8%. Moreover, the cross-sectional properties of both are preserved.

Returns The time series of one-period excess returns, plotted in the left panel of figure 15, follows very closely what one would expect from the firm's value plot (figure 10). When the initial beliefs are unbiased the first 7-8 years yield negative excess returns, particularly in the first couple of years. This coincides with the period of time when the parameter uncertainty, and thus also the firm's value, decreases sharply. When the initial beliefs are positively biased, the bias correction implies a further decrease in the firm's value, which contributes to even smaller excess returns. On the contrary, when the initial beliefs are negatively biased, the bias correction which pushes the firm's value up, helps to counteract the opposite effect of the reduction in the parameter uncertainty. If the negative bias is large enough, then the excess returns in the first years become positive. Regardless of the initial bias, after about years 8-10, the excess returns are very close to zero, since after that

[20] Similar results are obtained independently of the initial bias in the beliefs, reason why those results are omitted.

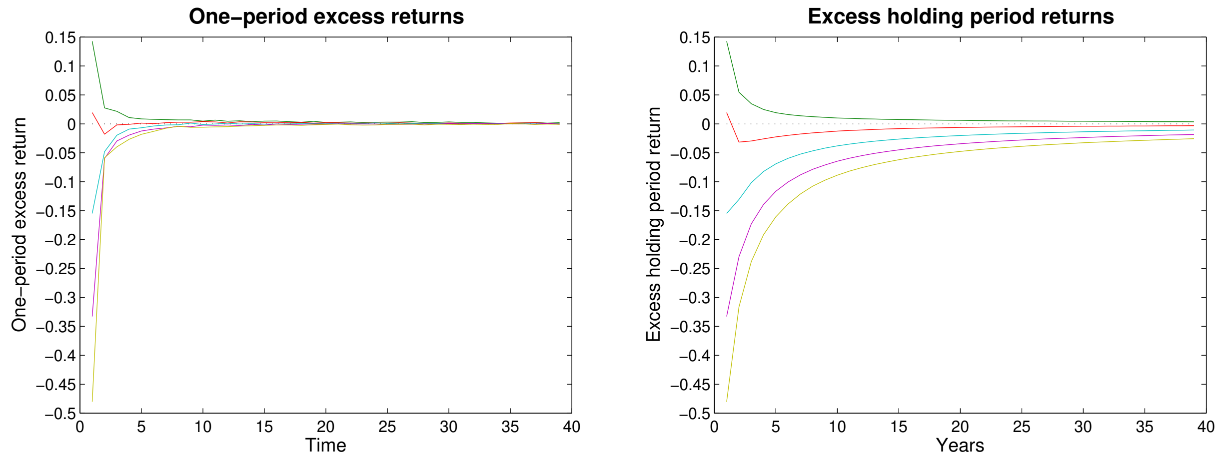


Figure 15: Time series of average excess returns. This figure plots the time series of the average excess one-period returns r_t (left panel) and the average annualized excess holding period returns (right panel) over 39 years for different initial beliefs. The average is taken over 10,000 simulations and the initial beliefs have a_0 ranging from 0.9 to 1.3 in 0.1 increments and $\sqrt{v_0} = 25\%$. The true value of α is 1.1. The outermost lines correspond to smaller initial beliefs.

point in time the firm's value and stock of capital are nearly steady over the years. These results are in line with those of Johnson (2007).

The right panel of figure 15 plots the time series of average annualized excess holding period returns (excess HPR), the effective rate of return that an investor obtains if he invests in the firm at time 0 and reinvests all the cash flows received in the firm. The most interesting finding from this plot is that, although the excess HPR converges to zero, it actually (almost) never reaches zero. This follows because the excess one-period returns converge monotonically to zero (except for the case of initial beliefs slightly negatively biased). Thus, on average, the investor (almost) never obtains positive (negative) excess one-period returns to compensate for early negative (positive) excess one-period returns, and the excess HPR (almost) never reaches zero. This means that if the initial beliefs are positively biased, unbiased or even slightly negatively biased, the investor will never recover from the negative excess returns obtained in the early years.

These results are in line with those of the initial public offerings (IPO) literature, specifically those of Ritter (1991). He finds that in the period of 1975-84 the IPO issuing firms on average underperform matching firms by 7% a year in the first 3 years after the IPO (excluding the first day return), which the model, even without specific calibration, is able to replicate when initial beliefs are unbiased.[21] Ritter (1991) explains the phenomenon with periods of investor's overoptimism in the new firms and consequent wave of IPO's that take advantage of that overoptimism. This model provides an alternative explanation: (i) a technological breakthrough occurs, (ii) as a consequence a wave of new technology firms enters in the market and (iii) the IPO underperformance follows even if the investors are not overoptimistic, because of the uncertainty about the returns to scale of the new technology firms. Although the technological breakthrough part of the story is useful to justify the IPO wave, it is not an essential ingredient for the underperformance. So long as

[21] He does not analyze the performance over longer periods of time.

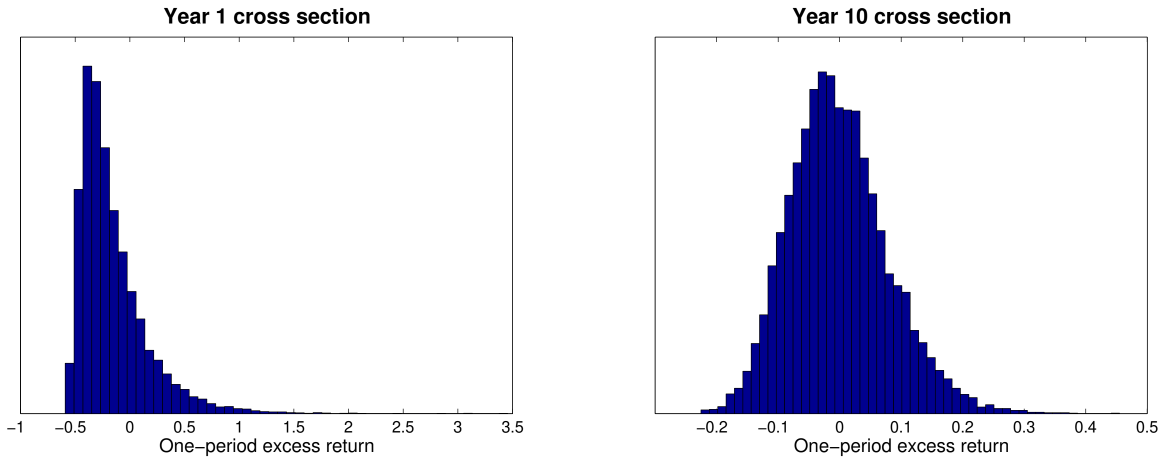


Figure 16: Cross section of one-period excess returns. This figure plots the cross sectional distribution of the one-period excess returns, when the initial beliefs are unbiased ($a_0 = \alpha = 1.1$ and $\sqrt{v_0} = 25\%$), based on 10,000 simulations. The left panel shows the year 1 and the right panel the year 10 cross sectional distribution.

there is uncertainty about the returns to scale of the issuing firm, which should be the common case, the underperformance occurs. In addition, the larger the parameter uncertainty, the larger the underperformance, which is consistent with the findings of Ritter (1991) that the underperformance is negatively correlated with firm's age when it goes public. In this model, the younger the firm, the larger the parameter uncertainty. But even if not all IPO underperformance episodes can be explained in these grounds, the uncertainty about the returns to scale definitely plays an important role.

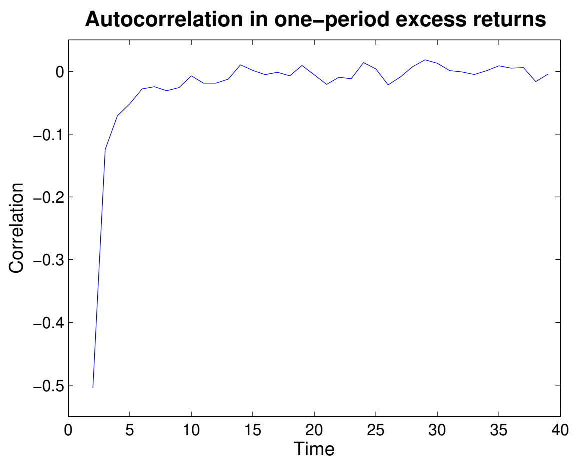
Looking now at the cross-section of one-period excess returns, figure 16 provides the cross-sectional distribution of the year 1 and 10 one-period excess returns, respectively, for the case on unbiased initial beliefs.[22] The cross-sectional distributions are very similar to those of the firm's value (not reported here), following the convexity of the firm's value to the firm's belief relation. The distributions are positively skewed, with the skewness decreasing with time. Also apparent from figure 16 is the progressive decrease in the volatility of the returns as time passes, associated with the decrease in parameter uncertainty.

Once again it is interesting to see the probability of obtaining a negative one-period excess return and a negative holding period excess return. Those probabilities are presented in table 3. Unbiased, positively biased or even slightly negatively biased initial beliefs all imply favorable odds of negative one-period excess returns. In all cases the probability of negative excess returns converge to slightly above 50%, following the convergence of the parameter uncertainty to 0 and consequently of all other variables to their full information values. Obviously, the larger the positive initial bias, the larger the probability of a negative excess return in the first years, due to the bias correction and to the decrease in parameter uncertainty. The former is responsible for the more modest favorable odds of positive excess returns when the initial beliefs are negatively biased. Remarkable is the 3/4

[22]The cross-sectional distributions for the case of biased initial beliefs and for the case of excess holding period returns are very similar to the ones reported here, and so are ommitted.

Year	1	2	3	5	10	20	39
One-period excess returns							
$a_0 = 0.9$	45.6%	50.6%	49.5%	50.7%	51.8%	51.8%	52.6%
$a_0 = 1.1$	77.6%	62.2%	57.3%	55.4%	53.9%	54.0%	54.6%
$a_0 = 1.3$	98.3%	63.3%	63.1%	59.4%	56.2%	53.7%	55.2%
Holding period excess returns							
$a_0 = 0.9$	45.6%	44.2%	41.5%	38.9%	33.2%	25.8%	23.4%
$a_0 = 1.1$	77.6%	85.9%	90.1%	94.2%	97.5%	98.5%	98.0%
$a_0 = 1.3$	98.3%	99.8%	100%	100%	100%	100%	100%

Table 3: Probability of negative excess returns. The table shows the probabilities of detecting a negative excess one-period return (top panel) and a negative excess holding period return (bottom panel) at a given year, based on 10,000 simulations. The value of α is 1.1, and the initial parameter uncertainty is $\sqrt{v_0} = 25\%$. The initial estimates of α are in the left column.



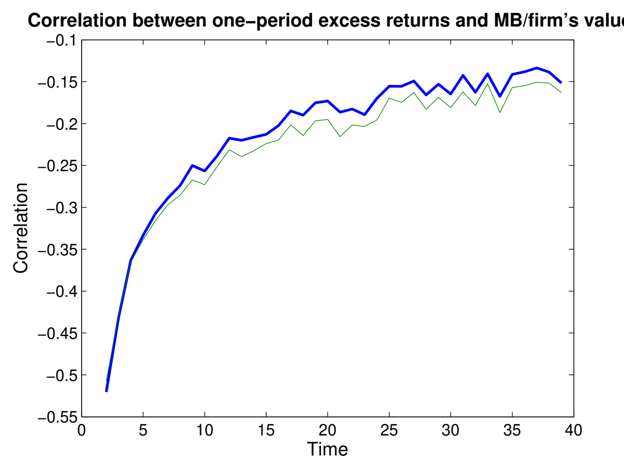


Figure 17: Time series of returns' correlations. This figure plots the first order autocorrelation in one-period excess returns (left panel) and the correlation between the one-period excess returns and the previous period MB ratio, the thicker line, and firm's value (right panel), over the first 39 years when the initial beliefs are unbiased ($a_0 = \alpha = 1.1$ and $\sqrt{v_0} = 25\%$).

chance of a negative excess return in the first year when the beliefs are unbiased, attributable solely to the decrease in the parameter uncertainty.

In turn the probability of a negative holding period excess return converges to either 100% or 0%, depending on whether the long-run average holding period excess return is negative or positive. This reflects the cumulative effect of the bias correction and of the convergence of the parameter uncertainty to zero. In the case of unbiased or positively biased initial beliefs, although large positive productivity shocks may, at an initial stage, generate large one-period returns and positive holding period excess returns, the reduction in parameter uncertainty and/or the bias correction that takes place in the long-run will eventually drive the long-run holding period excess return toward negative values.

Next I investigate the model's ability to reproduce two stylized facts: (i) the negative autocorrelation in long-horizon returns (Fama and French, 1988) and (ii) the book-to-market (or equivalently the market-to-book) and size effects (Fama and French, 1992). A priori the model is expected to reproduce these facts due to its learning feature, and the consequent convergence of the firm's value to its full information value.

The left panel of figure 17 plots the autocorrelation in one-period excess returns when the initial beliefs are unbiased.[23] From it we can see that although the average one-period excess return converges monotonically to zero, the path by path time series of excess returns exhibits significant negative autocorrelation in the first 3-5 years, being about 0 in the subsequent years. To understand how this negative autocorrelation is generated by the model, consider that the current beliefs are positively biased.[24] Large positive (negative) productivity shocks generate firm's beliefs, and thus also firm's values, above (below) the average, which imply higher (smaller) than average returns.[25] Since in that case the next period's firm's value is further away from (closer to) its full information value than average, the convergence to the full information value is weaker (stronger) than average, resulting in smaller (larger) returns. Because the mechanism that generates negative autocorrelation in the returns depends on the impact of the productivity shock on the firm's beliefs and firm's value, it is natural that one observes stronger autocorrelation when the parameter uncertainty is larger, that is, when the signal's precision is larger relative to the belief's precision.

In turn, the right panel of figure 17 shows a negative (positive) correlation between the one-period excess returns and the previous period MB (BM) and firm's value, which is consistent with BM and size effects. This correlation arises because for a given α large (small) MB and firm's value are associated with optimistic (pessimistic) beliefs, as can be deduced from figures 3, 7 and 8.[26] The subsequent correction of the biased beliefs has a negative (positive) impact in the firm's value which contributes to negative (positive) excess returns and the observed pattern of correlation. As time passes the bias correction weakens because (i) the bias is smaller due to past corrections and (ii) because the new information has less impact in the firm's beliefs due to the smaller parameter uncertainty. Thus, the autocorrelation gradually disappears as the firm learns about α. In the limit case of full information there is obviously no MB or size effects.

Also apparent from the figure is the large positive correlation between the MB and the firm's value (above 95%), which is not surprising since in the model both are influenced by the same effects (parameter uncertainty and bias effects) and in similar ways. Thus, in the model there is not really separate MB and size effects. Both are a direct consequence of the way the firm learns about α.

Table 4 gives the more traditional representation of the MB (BM) and size effects, presenting

[23]The plots for the case of biased initial beliefs are similar.

[24]If the current beliefs are negatively biased the opposite of what follows holds.

[25]The impact of a positive (negative) productivity shock on the current cash flow is indeterminate, since both the operational cash flow and the gross investment become above (below) the average. However, even in the worst scenario, the effect on the current cash flow does not, in general, dominate the effect on the continuation value.

[26]Actually, the MB to a_t relation exhibits a parabola shape, which can kill the MB effect. However, when the parameter uncertainty, and thus the standard deviation of a_t is large, the bottom of the parabola is shifted to the left and the probability of a_t lying on the left part of the parabola is small. As the parameter uncertainty reduces, the bottom of the parabola shifts to the right, but due to the bias corretion that takes places, the probability of a_t lying on the left part of the parabola remains small. But keep in mind that the choice of α is important. The choice of a smaller α could kill the MB effect. But such a small α would necessarily imply a full information optimal stock of capital too close to the minimum stock of capital, which is not sensible. Finally, notice that the parabola shape of MB to a_t relation is only an artifact created by the specific production function used, and is not a feature associated to decreasing returns to scale or a feature that is needed for any of the results in the paper.

Year	1	3	5	10	39
Low MB	7.62% (0.91)	4.9% (0.95)	3.5% (0.97)	2.1% (0.99)	1.1% (1.02)
2	4.6% (0.97)	3.0% (1.00)	2.2% (1.01)	1.2% (1.03)	0.5% (1.07)
3	−1.9% (1.09)	0.7% (1.06)	−0.1% (1.07)	0.1% (1.07)	0.0% (1.10)
4	−16.2% (1.22)	−3.1% (1.13)	−1.6% (1.12)	−1.1% (1.12)	−0.3% (1.12)
High MB	−38.8% (1.30)	−10.0% (1.24)	−6.0% (1.21)	−2.8% (1.18)	−1.0% (1.15)
Low Size	6.59% (0.87)	4.22% (0.88)	3.58% (0.90)	2.3% (0.93)	1.19% (1.01)
2	6.44% (0.98)	3.92% (1.01)	2.45% (1.02)	1.4% (1.04)	0.4% (1.07)
3	−2.6% (1.10)	0.8% (1.09)	0.0% (1.09)	0.1% (1.09)	0.1% (1.10)
4	−18.0% (1.22)	−3.5% (1.16)	−2.2% (1.15)	−1.3% (1.14)	−0.3% (1.12)
High Size	−37.2% (1.33)	−9.9% (1.25)	−6.2% (1.23)	−3.0% (1.20)	−1.0% (1.16)

Table 4: Average one-period excess returns of portfolios formed on MB and Size. The table reports the average excess return of portfolios formed on MB (top panel) and firm's value/size (bottom panel). The values in parenthesis correspond to the average belief on α for the firms in that portfolio. The excess returns were obtained through the simulation of 25,000 firms over 40 years. Each firm was created in the same year, with the same $\alpha = 1.1$. The beliefs were drawn from uniform distributions, $a_0 \sim U[0.8, 1.4]$ and $v_0 \sim U\left[0, 0.25^2\right]$. Firms have idiosyncratic productivity shocks. Each column represents the average returns at different points in time for the portfolio.

the average one-period excess returns for portfolios formed on MB or firm's value.[27] The values in the table were computed based on the 25,000 simulations of a firm with $\alpha = 1.1$ and initial beliefs drawn from uniform distributions, $a_0 \sim U[0.8, 1.4]$ and $v_0 \sim U[0, 0.25^2]$.[28] It was considered that each simulation corresponds to a firm, which amounts to say that: (i) all the firms have the same α; (ii) the productivity shock is idiosyncratic. This scenario is admittedly not what one would find in the economy, where firms from different sectors have different α's. Instead, this corresponds to a wave of firms sharing a new technology. The fact that the productivity shocks are entirely idiosyncratic, although not sensible, should not hinder the qualitative results if they were replaced by productive shocks with both systematic and idiosyncratic components.

From table 4 we can clearly see the MB/size effect documented by Fama and French (1992), and confirm that smaller (larger) MB and firm's values are associated with pessimistic (optimistic) beliefs. As time passes and bias correction takes place, the initial wide dispersion of average returns among portfolios narrows down to the magnitudes found by Fama and French (1992). The differences between the MB and size effects are small, as expected.

The interesting aspect of the MB and size effects generated by this model is that they are only apparent rather than true MB and size effects. In this model the firm's expected return is always the riskless rate of return, regardless of its beliefs, MB and value. As such it is impossible to have true MB and size effects. The firm does recognize that if its current beliefs happen to be positively (negatively) biased, the future average returns will be smaller (larger) than the riskless rate of return. However, averaging over the firm's beliefs, the expected return matches the riskless rate of return. In turn, the observer proxies the firm's expected return by averaging over past realized

[27]The use of portfolio formed both on MB and size makes no sense in this model because both effects are not separable.

[28]Similar results are obtained if the initial beliefs are the same for all firms, regardless of whether they are biased or not. This ensures the robustness of the results for different distributional assumptions for the initial beliefs.

returns, thus conditioning on the realization of α. By doing so he captures the bias correction impact on the returns thus failing to find a constant expected rate of return regardless of the MB and firm's value.

But even if the model can generate an apparent MB and size effect within a sector where all firms have the same α, that does not necessarily mean that the model delivers the same results in an economy-like scenario, where firms are clustered in sectors, each with different α's. Firms with smaller than average MB/values compared to their sectors may be included in the same portfolio as firms with larger than average MB/values in their sector. The first case is associated with large returns, whereas the latter is associated with small returns. Thus, the mixture of firms from sectors with different α's may wash out the MB and size effects present when they are considered in isolation. Whether this model is able to produce MB and size effects in an economy like scenario depends crucially on the distribution of α's in the economy.[29]

4 Solutions to the Firm's Problem with Adjustment Costs

4.1 Parameter Values for the Solutions

Except for the adjustment cost parameter, ϕ, all parameters have the same values as in the case without adjustment costs, presented in table 1. The choice of a reasonable value for ϕ was based on empirical estimations of this parameter. Nonetheless, it was difficult to choose one single value of ϕ given the wide range of estimations for this parameter obtained by different authors employing diverse methodologies and data sets: Cooper and Haltiwanger, 2006 present estimates ranging from 0.05 to 0.46, Hall, 2002 and Hall, 2004 a median across industries of 0.91 and 0.36, respectively, and Shapiro, 1986 a value of 2.2. Given the disparity between estimates and the importance of this parameter in the model, I considered 4 different values for ϕ: 0.2, 0.4, 1 and 2.

4.2 The Full Information Case

The solution to the firm's problem with adjustment costs is obtained adding the adjustment cost component to the cash flow and using the same numerical procedure used for the case without adjustment costs.

The addition of the adjustment costs does not change most of the features of the solution for the case without adjustment costs obtained before: the firm's value and optimal stock of capital are still increasing and convex in α. In fact, the long-run optimal stock of capital and firm's value is the same both with and without adjustment costs and the only two effects of the adjustment costs are the sluggish adjustment of the capital stock and a smaller firm's value when the stock of capital is away from its optimal level. First, given that the marginal value of capital, $\rho\frac{\partial V_{t+1}}{\partial K_{t+1}}$, is decreasing in K_{t+1}, due to decreasing returns to scale, and that the marginal cost of capital, $1+\phi\left(\frac{K_{t+1}-K_t}{K_t}\right)$, is increasing in K_{t+1}, if it is optimal to adjust the stock of capital in one direction when there

[29]For instance, in the case were all firms have the same beliefs and α is drawn from their subjective distribution, there is no MB and size effect at all.

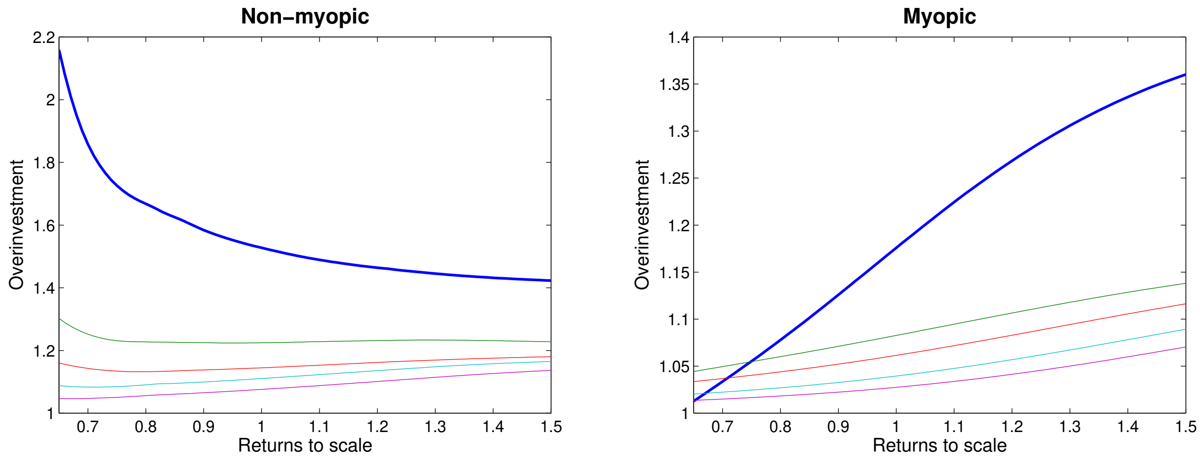


Figure 18: Overinvestment against the full information optimal stock of capital in the same state. This figure plots the ratio of the optimal stock of capital (K_{t+1}) with paramer uncetainty to its full information value in the same state, for a_t/α ranging from 0.65 to 1.5,$\sqrt{v_t} = 25\%$ and $K_t = 4.1438$. Each line corresponds to a different level of adjustment costs, $\phi = \{0, 0.2, 0.4, 1, 2\}$. In both panels the uppermost lines correspond to smaller adjustment costs. The thicker line corresponds to the case of no adjustment costs. In the right (left) panel it is considered a (non-)myopic firm.

are no adjustment costs, then the same is true in the presence of adjustment costs. Moreover, the larger the ϕ, the smaller the adjustment in the stock of capital. This implies that the long-run stock of capital is the same regardless of the value of ϕ, but that the larger the ϕ, the longer it takes to attain its long-run value. Second, once the long-run stock of capital is attained, no further changes to the stock of capital take place, which implies that the adjustment cost component of the cash flow becomes null and so the long-run firm's value is independent of the value of ϕ. However, when the stock of capital is not at its long-run optimal level, the firm pays adjustment costs which, in addition to the sluggish adjustment of the stock of capital, contributes to a smaller firm value compared to the case without adjustment costs. The larger the deviation of the stock of capital from its long-run optimum, the larger the difference in the firm value.

4.3 The Case of Parameter Uncertainty

4.3.1 Static Analysis

The natural question to pose in this section is "What is the impact that adjustment costs have on the overvaluation and overinvestment obtained in their absence?". Johnson (2007) conjectures that one should still observe overinvestment in the presence of symmetric adjustment costs since the incentive to learn is not affected by the existence of symmetric adjustment costs. However, he abstains from commenting on the effect of the adjustment costs on the magnitude of the overinvestment and overvaluation.

Figure 18 provides the comparison of the overinvestment for different levels of adjustment costs, including the case of no adjustment costs (thicker line). The overinvestment is measured against the full information optimal stock of capital in the same state, like in figure 6, and I considered

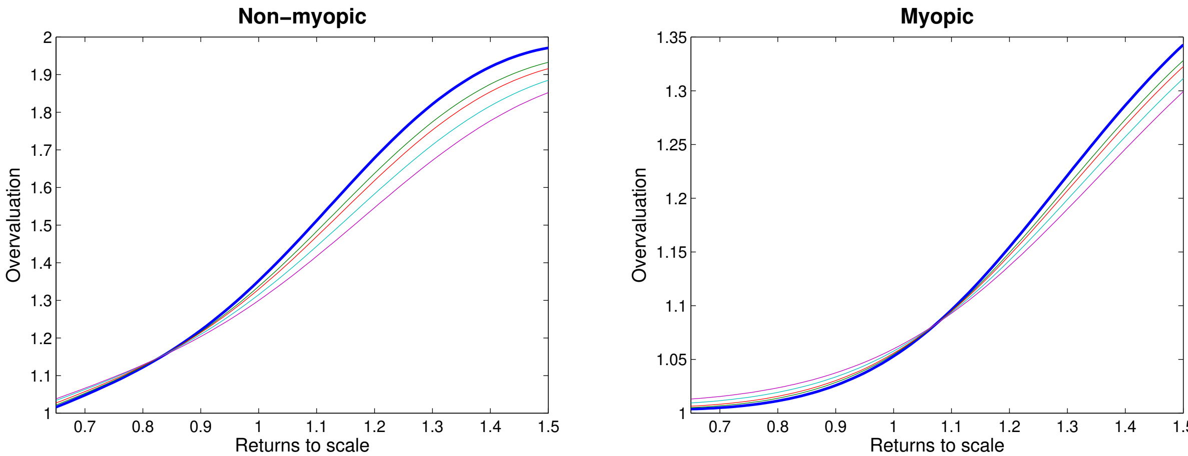


Figure 19: Overvaluation against the full information firm's value in the same state. This figure plots the ratio of the firm's value with paramer uncetainty to its full information value in the same state, for a_t/α ranging from 0.65 to 1.5,$\sqrt{v_t} = 25\%$ and $K_t = 4.1438$. Each line corresponds to a different level of adjustment costs, $\phi = \{0, 0.2, 0.4, 1, 2\}$. . The thicker line corresponds to the case of no adjustment costs. In the right (left) panel it is considered a (non-)myopic firm.

the case of $\sqrt{v_t} = 25\%$ and $K_t = 4.1438$.[30] It stands out immediately that adjustment costs have a large negative impact on the overinvestment, reducing the overinvestment to less than a half of its value in the absence of adjustment costs, even for the smallest adjustment cost considered here. Comparing both panels, we can see that this is mostly due to a smaller "convexity" effect. In practice, the convexity of the marginal value of capital remais the same, but since the marginal cost of capital is increasing in K_{t+1} (due to adjustment costs), the incentive for a myopic firm to overinvest is smaller. The multiplicative learning effect seems to be just as strong as in the case without adjustment costs. This partially backs up Johnson's (2007) conjecture. The learning effect does remain intact, but the smaller "convexity" effect can reduce the overinvestment to insignificant values in the presence of large adjusment costs. Regarding the strategic effect, we can see that it becomes weaker as the adjustment cost increases, since the adjustment cost already contributes to smaller overinvestments and thus prevents a too fast learning.

Figure 19 plots the overvaluation in the same scenario of figure 18. Perhaps surprisingly, it shows that the impact of adjustment costs on the overvaluation, both in the myopic and non-myopic cases, is much smaller than that on the overinvestment. However, there is nothing surprising about the result. Think of the firm's value with parameter uncertainty, as the sum of two components: (i) the expected value of its long-run value, which is independent of adjustment costs as discussed before and (ii) the expected value of intermediate cash flows. The adjustment costs have a direct negative impact on the latter through reduced cash flows, and an indirect negative impact through suboptimal stocks of capital, which implies that they have a negative impact on both the parameter uncertainty firm's value and full information firm's value. Whether this impact is stronger on the

[30] For different values of $\sqrt{v_t}$ the pictures are similiar. Larger (smaler) values of K_t have the same effect of larger (smaller) adjustment costs, as is obvious from the inspection of equation (2). However, the qualitative results remain unchanged when using reasonable K_t values.

former or the latter depends on the initial stock of capital and the (belief on) α. For a given α, if the initial stock of capital is small (large) enough, the adjustment costs to pay in the full information case are smaller (larger) than the expected value of the adjustment costs to pay in the parameter uncertainty case, due to the convexity of the relation between the optimal stock of capital and α (see figure 3). Therefore, the overinvestment is smaller (larger) the larger the level of the adjustment cost. Either way, what figure 19 shows us is that the component of the firm's value that is influenced by adjustment costs is relatively small compared to the expected value of the long-run firm's value, which is why the impact of the adjustment costs on overvaluation is relatively small.

The analysis of the overvaluation measured against the long-run MB is trickier in the presence of adjustment costs, since the MB in the parameter uncertainty case heavily depends on the previous period stock of capital, which may assume any value depending on the history of productivity shocks. The larger the last period's stock of capital, the smaller the current period's MB because the larger will be the current period's stock of capital. For this reason this analysis will be skipped here. In the next subsection it will be analysed in the context of the time series simulations.

4.3.2 Dynamic Analysis

To simulate the time series evolution of the variables of interest I used basically the same procedure as in the case of no adjustment costs. The only difference is that the adjustment cost in the initial period ($t = 0$) is computed based on the assumption that $K_{-1} = 1$, the minimum stock of capital, instead of 0. This assumption is necessary because assuming that $K_{-1} = 0$, as is actually the case, the adjustment cost would be infinite and the firm would never become active. Alternatively one could have simply ignored the adjustment cost in the initial period, or assumed a somewhat ad hoc cost of initial investment, as in Alti (2003). However, the maintained assumption seems to be more reasonable and consistent with what happens in the subsequent periods.

In the remaining of this subsection, the focus will be on the case of unbiased initial beliefs. The results for the case of biased initial beliefs can be infered from the unbiased case analysed here and from the analysis to the case without adjustment costs analysed in the previous section.

Firm's Beliefs, Overvaluation and Overinvestment Figure 20 plots the overinvestment and MB overvaluation measured against the long-run stock of capital and MB and reflects what a less sophisticated observer would find once α is known. Starting with the time series of average overinvestment, depicted in the left panel of figure 20, we can see that the adjustment costs are responsible for a remarkably different pattern of overinvestment along time. The initial periods are characterized by increasing stocks of capital and underinvestment, exactly the opposite of what happened in the absence of adjustment costs, followed by periods of gradually vanishing overinvestment. Moreover, the peak overinvestment becomes quite small, compared to the case without adjustment costs .

These patterns of overinvestment are a result of the sluggish convergence of the capital stock to a threshold, defined as the stock of capital that the firm would choose if the next adjustment cost

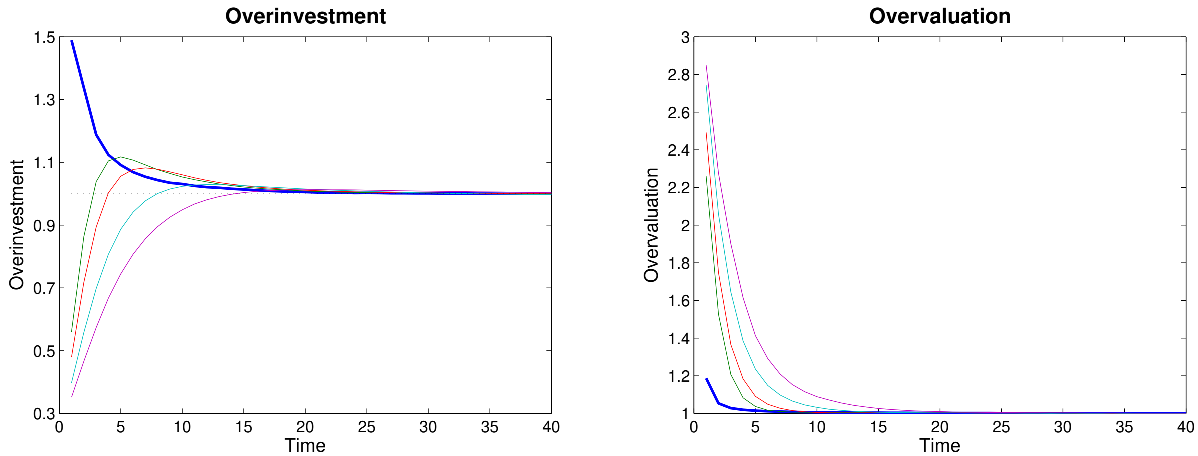


Figure 20: Time series of average overinvestment and market-to-book overvaluation against the long run values. This figure plots the time series of average overinvestment (left panel) and MB overvaluation (right panel) as measured against the respective long run values over 39 years for different levels of adjustment costs, $\phi = \{0, 0.2, 0.4, 1, 2\}$. The average is taken over 10,000 simulations with initial beliefs are $a_0 = 1.1$ and $\sqrt{v_0} = 25\%$. The true value of α is 1.1. In the left (right) panel the outermost lines at $t = 1$ correspond to smaller (larger) adjustment costs. In both panels the thicker line corresponds to the case of no adjustment costs.

were waived. This threshold converges to the long-run stock of capital as the parameter uncertainty converges to zero, and is close to the stock of capital chosen by a firm without adjustment costs. In the limiting case of no adjustment costs, the firm always chooses a stock of capital equal to the threshold, reason why the stock of capital monotonically decreases from the very beginning. However, in the presence of adjustment costs the initial stock of capital chosen by the firm will always fall short of that threshold. If the adjustment costs are large enough, then the stock of capital in the initial periods will fall below the next period's thresholds as well, even though the latter decreases as time passes and parameter uncertainty reduces. As a result we obtain an increasing stock of capital in the initial periods. In addition, if the adjustment costs are large enough, the stock of capital in the initial periods will be smaller than its long-run value, and underinvestment is obtained. Once the stock of capital becomes above the threshold, the stock of capital starts to decrease, trailing the threshold and converging to its long-run value.

How large the overinvestment is and how soon it occurs depends on how large the adjustment costs are. The smaller the adjustment costs, the faster the convergence to the threshold. Since the threshold is always larger than the long-run stock of capital, this implies that the smaller the adjustment costs the sooner overinvestment is obtained. This in turn implies that the larger the overinvestment is, since the threshold decreases as time and parameter uncertainty decrease. If the adjustment costs are large enough, it is possible that the convergence be so slow that overinvestment never obtains. This is almost the case when $\phi = 2$.

On the contrary, the pattern of the MB overvaluation, depicted in the right panel of figure 20, is quite similar to what was obtained in the absence of adjustment costs. However, the introduction of adjustment costs makes the MB overvaluation dramatically larger and more persistent. This is basically the result of a denominator effect. As we have just seen, adjustment costs dramatically

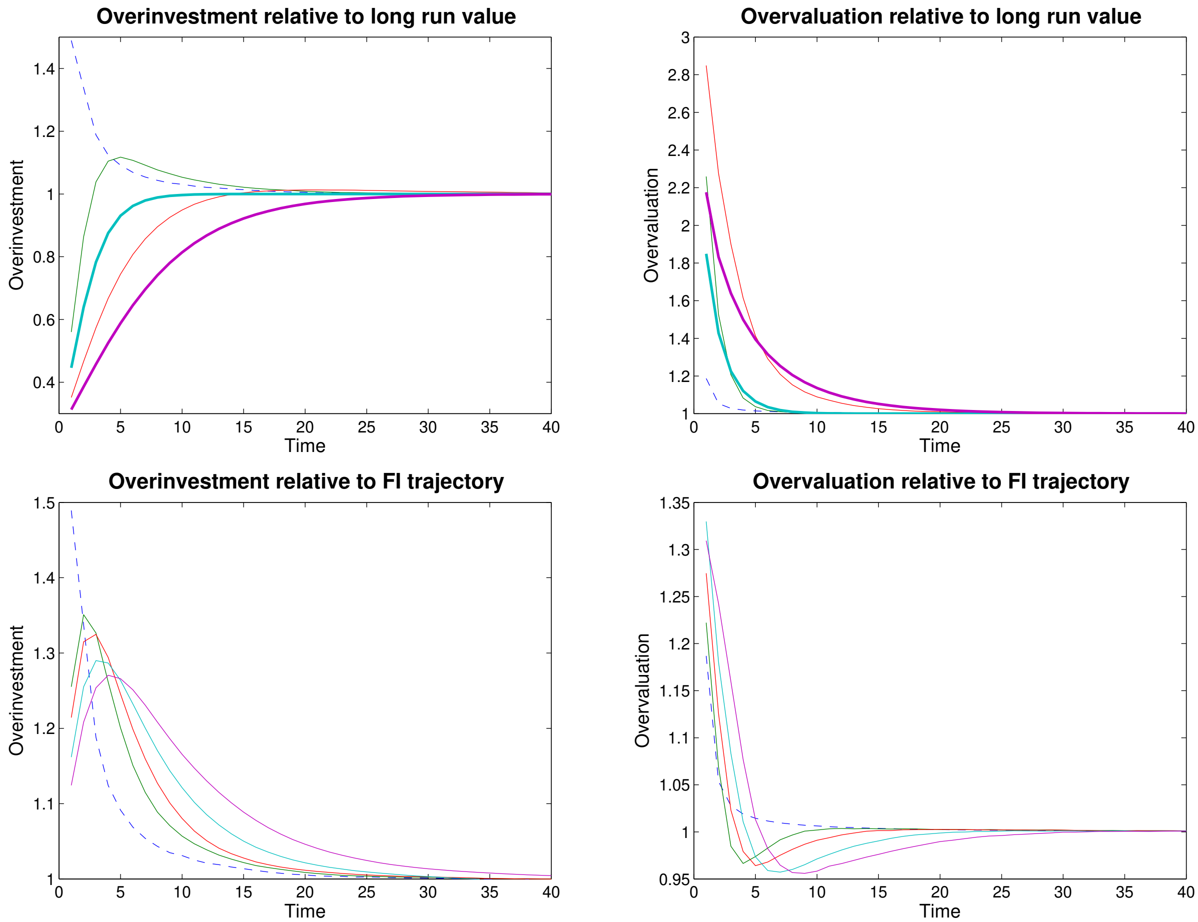


Figure 21: Time series of average overinvestment and market-to-book overvaluation against the full information values. In the top panels it is plotted the time series of average overinvestment (top left panel) and MB overvaluation (top right panel) against the long run values for different levels of adjustment costs, $\phi = \{0.2, 2\}$. The thicker lines correspond to the full information case and the thiner ones to the case of parameter uncertainty ($a_0 = 1.1$ and $\sqrt{v_0} = 25\%$). The average is taken over 10,000 simulations and the true value of α is 1.1. In the top left (right) panel the outermost lines correspond to the smaller (larger) adjustment costs. In the the bottom panel it is plotted the time series of average overvinvestment (bottom left panel) and average MB overvaluation (bottom right panel) measured against the full information values at the same point in time, for different levels of adjustment costs, $\phi = \{0, 0.2, 0.4, 1, 2\}$. In the bottom left (right) panel the outermost lines at $t = 1$ correspond to smaller (larger) adjustment costs. In all panels, the dashed line corresponds to the case of parameter uncertainty with no adjustment costs.

reduce the stocks of capital in the initial periods and for a considerable lengthy period of time. However, the impact on the firm's value with parameter uncertainty is much smaller: the assets in-place component of the firm's value decreases, but it is partially compensated by an increase in the growth options component. This explains why the MB overvaluation increases with the level of adjustment costs. In turn, the slower adjustment in the stock of capital that results from larger adjustment costs is responsible for a more persistent growth option component of the firm's value, and thus more persistent MB overvaluations.

However, adjustment costs also have the effect of delaying the convergence of the firm's value and stock of capital to their long-run values even when there is no parameter uncertainty, which in the absence of adjustment costs is immediate. The more attent reader should by now be wondering about what MB overvaluation and overinvestment, a less sophisticated observer would ex-post find in the full information case. The top panels of figure 21 compare the overinvestment and MB overvaluation, measured against the long-run values, that one would observe with and without (thicker lines) parameter uncertainty when $\phi = \{0.2, 2\}$ and all the rest as in figure 20. Without parameter uncertainty the stock of capital gradually converges to its long-run value from below, generating a pattern of vanishing underinvestment as one would expect. Surprising is the large and persistent overvaluation relative to the long-run MB, due only to the friction in the adjustment of the capital stock and value of the growth options introduced by the adjustment costs. The implication is that in the case of parameter uncertainty, the MB overvaluation attributable to parameter uncertainty only is much smaller than what one may have thought by looking to figure 20. This shows how difficult it is in practice to identify a bubble, since the fundamental value may itself exhibit a bubble like pattern (see e.g. Hamilton and Whiteman, 1985).

To give a picture of the true MB overvaluation and overinvestment, attributable to the parameter uncertainty only, the figures in the bottom panel of figure 21 plot the overinvestment and MB overvaluation measured against the values that one would observe in the full information case at the same point in time, the fundamental MB and stock of capital. In the context of the model, a sophisticated observer is able to determine these fundamental values after knowing the value of α. In practice, however, those values would need to be estimated, which makes it hard to identify the bubble.

The bottom left panel shows us that parameter uncertainty always generates overinvestment relatively to the case of no parameter uncertainty, which follows naturally from the findings of the previous subsection. The larger the adjustment costs the smaller the peak overinvestment but the more persistent the overinvestment is, since the longer it takes for the stock of capital to converge to its long-run value (both in the full information and parameter uncertainty cases).

In the bottom right panel we can see that parameter uncertainty also generates a modest (yet larger than in the absence of adjustment costs) MB overvaluation in the first few years, followed by a slight but persistent MB undervaluation that gradually vanishes. The overcorrection to the bubble observed in the model is a common feature of many bubble and crash episodes (see e.g. Shiller (2000)). As we know from previous sections, parameter uncertainty has a positive impact on both the firm's value and the stock of capital. However, the friction in the stock of capital adjustment introduced by the adjustment cost prevents the stock of capital in the full information and parameter uncertainty cases to be too far apart from each other in the initial periods. This accounts for the initial MB overvaluation. However, as time passes the gap between the stocks of capital in both cases increases and the parameter uncertainty decreases, which narrows down the difference in the firm's value in both cases. As a result the initially larger parameter uncertainty MB eventually drops below the full information MB and undervaluation is obtained.

Notice how different mechanisms behind the MB overvaluation in the case with and without adjustment costs are. Without adjustment costs the MB overvaluation was mainly a result of

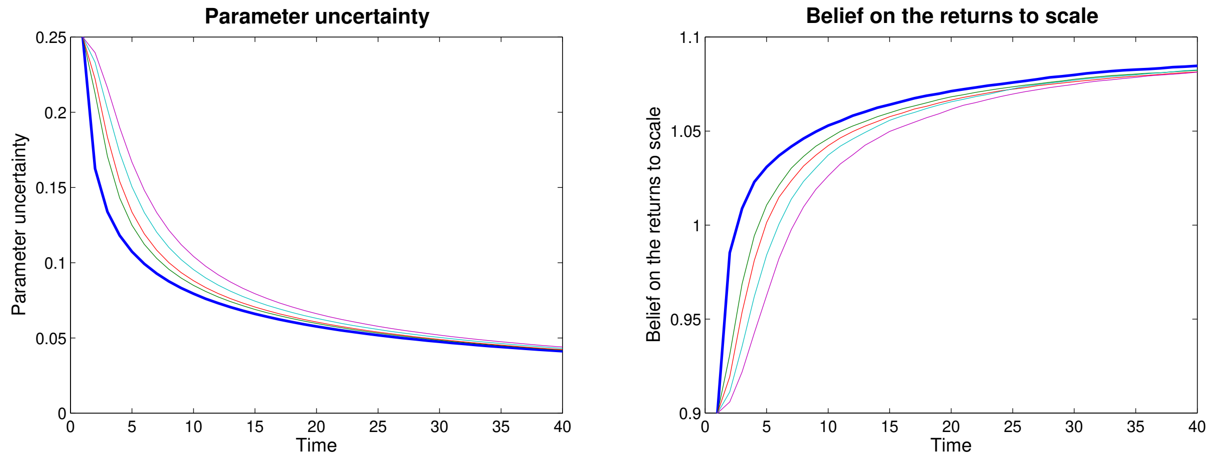


Figure 22: Time series of firm's beliefs. This figure plots the times series of average a_t (right panel) and the average $\sqrt{v_t}$ (right panel) over 39 years and for different levels of adjustment costs, $\phi = \{0, 0.2, 0.4, 1, 2\}$. The average is taken over 10,000 simulations, $\sqrt{v_0} = 25\%$ and the true value of α is 1.1. In the left panel $a_0 = 1.1$ and in the righ panel $a_0 = 0.9$. In the left (right) panel the outermost lines correspond to larger (smaller) adjustment costs. The thicker line corresponds to the case of no adjusment costs.

the strategic effect, which restrained the overinvestment. The MB overvaluation was ephemeral since the strategic effect weakened very fast as parameter uncertainty decreased, and there were no growth options left to exercise. With adjustment costs, the initial MB overvaluation is mainly attributable to growth options generated by the adjustment costs. The strategic effect seems to play a minor effect since the stocks of capital in the initial periods are already restricted by the adjustment costs. The subsequent MB undervaluation is mainly a result of overinvestment.[31] Due to the friction imposed by the adjustment costs, the correction of such overinvestment takes a long time, explaining the persistence of the MB overvaluation.

Regarding the learning process, it is not a surprise to find that adjustment costs slow down the learning process and the correction of initially unbiased beliefs, as can be seen from figure 22. This follows from the negative impact of the adjustment costs in the time series of the stock of capital. The effect is specially strong in the first 5-10 years, when the effect of the smaller initial stocks of capital is more present.

Returns Looking now at the time series of average one-period excess returns in the presence of adjustment costs, plotted in the left panel of figure 23, we can observe the same basic pattern of negative excess returns gradually disappearing, as in the case of no adjustment costs. However, the larger the adjustment cost, the smaller the negative excess returns in the first couple of years, and the more persistent they are. This happens because adjustment costs slow down the reduction in parameter uncertainty, which is the mechanism behind the observed negative excess returns. As the parameter uncertainty reduces, the associated overvaluation reduces and the firm's value tends to go down, which generated negative excess returns. Moreover, this downward pressure on the firm's

[31] Overinvestment relative to the full information case, which may even be underinvestment relative to the long run stock of capital.

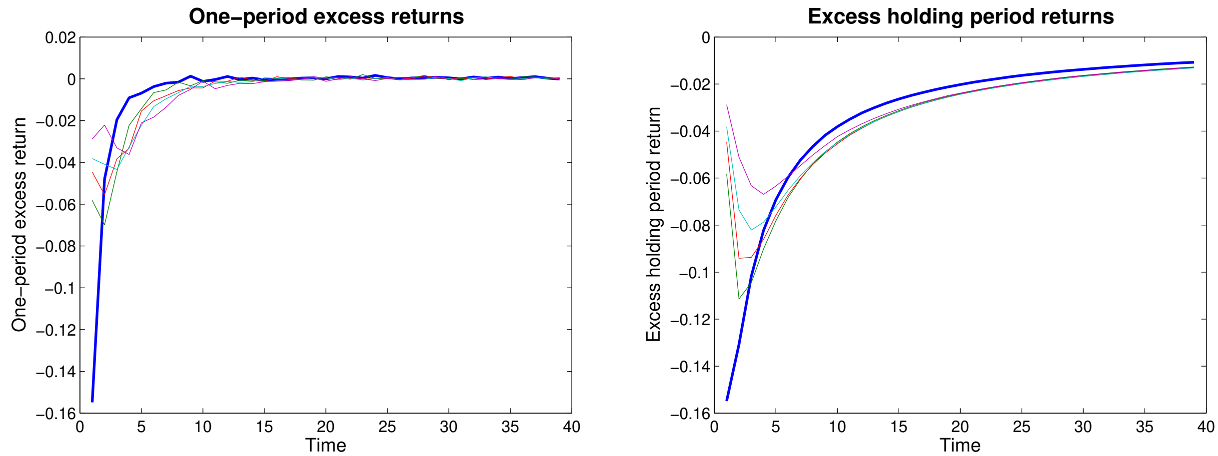


Figure 23: Time series of average excess returns. This figure plots the time series of average one-period excess returns (left panel) and average excess holding period returns (right panel) over 39 years for different levels of adjustment costs, $\phi = \{0, 0.2, 0.4, 1, 2\}$. The average is talen over 10,000 simulations woth initial beliefs $a_0 = 1.1$ and $\sqrt{v_0} = 25\%$. The true value of α is 1.1. In both panels the outermost lines at $t = 1$ correspond to larger adjustment costs. The thicker line corresponds to the case of no adjusment costs.

value is partially compensated for the adjustment in the stock of capital that brings it closer to the threshold defined previously and thus tends to increase the firm's value. This is actually the reason why excess returns drop from year 1 to year 2. The same reasoning explains the patterns of the time series of average excess holding period returns, pictured in the right panel.

The adjustment costs were found to have a minor impact on the results concerning the auto-correlation in the excess returns, the correlation between excess returns and MB and firm's value and the MB and size effects, reason why those results are ommited here.

5 Conclusion

This paper provides a model that explains overvaluation of and overinvestment by innovating firms usually associated to periods of technological revolution without recurring to irrationality or market imperfections. The central feature of the model is that the firm does not know a priori the returns to scale (α) of its production function and has to learn it from the observation of the cash flows it generates. Overvaluation and overinvestment relative to the full information case are shown to be a direct result of the convexity of the firm's value in α and of the convexity of the marginal value of capital in α and decreasing returns to scale, respectively. Even a myopic firm which does not anticipates the effects of its decisions over the stock of capital on the learning of α would obtain overvaluation and overinvestment. A non-myopic firm who manages the stock of capital in order to achieve optimal learning obtains even larger overvaluation and in general also overinvestment.

The model is able to generate convincing time series patterns of overvaluation, overinvestment and excess returns, with initial periods characterized by overvaluation, overinvestment and negative excess returns, that gradually disappear as the firm learns the value of α. The time series pattern of excess returns generated by the model is consistent to what is documented in the IPO literature. In particular, without being specifically calibrated for that, the model produces magnitudes remarkably similar to those reported in Ritter, 1991. The model is also able to reproduce two other empirical regularities: negative autocorrelation in the excess returns (Fama and French, 1988) and what an observer would identify as a market-to-book (MB) and size effect (Fama and French, 1992). Both effects follow from the correction of inferencial errors, and the latter also from the positive correlation between α and the firm's value and MB.

The introduction of adjustment costs is shown to have little impact on the overvaluation but a considerably negative impact on overinvestment. In the time series, the sluggishness of the adjustment in the stock of capital produces what a naive observer would ex-post identify as dramatic and more persistent overvaluations. A sophisticated observer who is able to recognize that fundamental (full information) values change over time, sees a different picture. Initial moderate overvaluations, yet larger and more persistent than in the absence of adjustment costs, are followed by slight overcorrection that graudally disappears. The overcorrection is mainly a result of the excess capacity accumulated in the initial years. The other results are found to be robust to the introduction of adjustment costs.

Given its simplicity, the model excludes important features. In particular, the model ignores any type of interaction between firms of the new economy. Free-riding and first-mover advantage have the potential to provide new interesting insights. The introduction of these feature would, however, require a significantly different model.

In the realms of the possible extensions to this model we have a more realistic financing side of the firm's problem and asymmetric adjustment costs. The introduction of costs of obtaining external funding should have a similar effect in the results to the introduction of adjustment costs. The limited autofinancing and the costs of external funding are expected to generate a friction to increases of the stock of capital, specially in the first years. This in turn contributes to a smaller overinvestment and more persistent overvaluation, through more persistent growth options. In addition, the history of productivity shocks becomes more important, given its impact in the firm's autofinancing. But, above all, the introduction of a more realistic financing side to the firm's problem will allow the analysis of dividend policies and capital structures. The drawback is that the added complexity may make infeasible to obtain numerical solutions to the model.

By the contrary, the introduction of asymmetric adjustment costs, with different ϕ for decreases (ϕ_D) and increases (ϕ_I) in the stock of capital, does not complexify the model much. The overvaluation and overinvestment obtained in a given state of nature are expected to be approximately an average of what they would be in the case of symmetric costs with ϕ_D and ϕ_I. The interesting implications of the asymmetric adjustment costs should be on the patterns of overvaluation and overinvestment.

References

Abreu, D. and Brunnermeier, M. (2003). Bubbles and crashes. *Econometrica*, 71(1):173–204.

Allen, F. and Gale, D. (2000). Bubbles and crises. *The Economic Journal*, 110:236–255.

Allen, F. and Gorton, G. (1993). Churning Bubbles. *The Review of Economic Studies*, 60(4):813–836.

Allen, F., Morris, S., and Postlewaite, A. (1993). Finite bubbles with short sale constraints and asymmetric information. *Journal of Economic Theory*, 61(2):206–229.

Alti, A. (2003). How Sensitive Is Investment to Cash Flow When Financing Is Frictionless? *The Journal of Finance*, 58(2):707–722.

Bertocchi, G. and Spagat, M. (1998). Growth under uncertainty with experimentation. *Journal of Economic Dynamics and Control*, 23(2):209–231.

Cooper, R. and Haltiwanger, J. (2006). On the Nature of Capital Adjustment Costs. *Review of Economic Studies*, 73(3):611–633.

Fama, E. and French, K. (1988). Permanent and Temporary Components of Stock Prices. *Journal of Political Economy*, 96(2):246–273.

Fama, E. and French, K. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2):427–465.

Gill, J. (2002). *Bayesian Methods: A Social and Behavioral Sciences Approach*. Chapman and Hall/CRC, Boca Raton.

Hall, R. (2002). Industry Dynamics with Adjustment Costs.

Hall, R. (2004). Measuring Factor Adjustment Costs. *Quarterly Journal of Economics*, 119(3):899–927.

Hamilton, J. and Whiteman, C. (1985). The Observable Implications of Self-Fulfilling Expectations. *Journal of Monetary Economics*, 16:353–373.

Johnson, T. (2007). Optimal learning and new technology bubbles. *Journal of Monetary Economics*, 54(8):2486–2511.

Liptser, R. and Shiryaev, A. (2001). *Statistics of Random Processes I & {II}*. Springer, New York, second edition.

Malkiel, B. (1996). *A Random Walk Down Wall Street*. W. W. Norton & Co., New York.

Miranda, M. and Fackler, P. (2004). *Applied Computational Economics and Finance*. MIT Press, Cambridge.

Pástor, Ä. and Veronesi, P. (2009). Technological Revolutions and Stock Prices. *American Economic Review*, 99(4):1451–1483.

Perez, C. (2003). *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages.* Edward Elgar Publishing, Cheltenham.

Rappoport, P. and White, E. (1993). Was There a Bubble in the 1929 Stock Market? *Journal of Economic History*, 53(3):549–574.

Ritter, J. (1991). The Long-Run Performance of Initial Public Offerings. *Journal of Finance*, 46(1):3–27.

Shapiro, M. (1986). The Dynamic Demand for Capital and Labor. *The Quarterly Journal of Economics*, 101(3):513–542.

Shiller, R. (2000). *Irrational Exuberance.* Princeton University Press, Princeton.

Appendix

A Algorithm Used in the Numerical Approximations to the Solution of the Firm's Problem

The solutions to the functional fixed point problems associated to each variant of the firm's problem analyzed in the paper, were obtained by iterating over approximations of the value function candidates. To approximate the value functions it was used Chebyshev polynomials, fitted to Chebyshev collocation nodes using the collocation method. The algorithm, whose extensive details can be found in Miranda and Fackler (2004), can be summarized in the following steps: (i) define the collocation nodes, over a range of values, for each of the state variables (these are the points in the state space where the Chebyshev polynomial approximation exactly matches the value function); (ii) provide an initial guess of the value function at each of the collocation nodes, and use it find the coefficients of the Chebyshev polynomial that matches those values or, alternatively, provide an initial guess for the Chebyshev polynomial coefficients; (iii) find the optimal policy at each of the collocation nodes and respective optimal value function, using the current Chebyshev polynomial approximation to the value function; (iv) based on the optimal values at each of the collocation nodes obtained in step (iii), find the coefficients of the Chebyshev polynomial that match those values; (v) repeat steps 3 and 4 until the optimal values at each of the collocation nodes converge.

By definition, the Chebyshev approximation to the value function thus obtained ensures that the value at the collocation nodes is a fixed point. However, in other points of the state space, that may not be the case. To access the accuracy of the value function approximation one can compare the difference between the left hand side and the right hand side of the Bellman equation at other points of the state space, using the Chebyshev approximation in place of the value function.

B Details of the Solution in the Full Information and Myopic Cases

In the full information case the firm's problem, described by equations (9) to (11), reduces to equation (12), thus eliminating the firm's beliefs (a_t and v_t) as state variables. It is also possible to eliminate the productivity shock as a state variable by using the certainty equivalent version of the problem (12). The stochastic operational cash flow is then substituted by its expected value, which amounts to fix the productivity shock at $\frac{\sigma^2}{2}$.[32] This modification is innocuous since the optimal policy is independent of the productivity shock and the continuation value in the original and modified versions of the problem is the same. In the modified model the value function is already the expected value, over the productivity shock, of the original model's value function. Thus, the simplified model has the added benefit of being a deterministic dynamic programming problem instead of a stochastic one, like the original problem. The only thing to take in consideration, is that the firm's values obtained in the simplified model are for the case of $z_t = \frac{\sigma^2}{2}$. For other values of the productivity shock one just has to remove the expected value of the operational cash flow and substitute it by its appropriate value.

The one-dimensional problem described above needs to be solved for each α considered. To avoid this and to serve as a basis for the general case with parameter uncertainty, I add α as a state variable. Obviously,

[32] The operational cash flow, and thus the productivity shock, could have been replaced by any deterministic value, so long as the proper corrections were made when computing the continuation value.

each state of α is absorbing, that is, it never changes. In effect, the two-dimensional problem is only an aggregation of the continuum of one-dimensional problems, with different values of α. The solution to the two-dimensional problem is then approximated by a two-dimensional Chebyshev polynomial.

The range of interpolation (value function approximation) used for the state variables, was $K \in [1, 36]$ and $\alpha \in [0.65, 1.5]$. Since the full information solution is going to be used mainly as a basis for the solution with parameter uncertainty, the choices of these interpolation ranges are mostly related to the parameter uncertainty problem. The lower limit of K is the minimum stock of capital, and the upper limit is a value high enough to ensure that optimal stock of capital (control) in the parameter uncertainty case is always inside the range of interpolation for K, even in the worst case scenario when α/a and v are at their upper limits. The lower limit of α was chosen so that the optimal stock of capital is slightly above the minimum of 1.[33] In this way it is avoided the kink in the value function originated by the minimum stock of capital constraint, which would pose severe problems in the precision of the Chebyshev polynomial approximation.[34] The upper limit of α is dictated by precision and computation time concerns in the solution of the parameter uncertainty problem. The larger the α, the larger the upper limit of K must be, which asks for more collocation nodes in order to maintain the solution's precision and increases the computation time exponentially.

For Chebyshev polynomial approximations, there is very little degrees of freedom in the choice of the collocation nodes (unlike in splines approximations), under the penalty of severe lack of precision and or difficulties in the convergence of the value function. Thus it was used Chebyshev collocation nodes, which concentrate more nodes in the extremes of the interpolation region, where the approximation is usually more unstable.[35] The number of collocation nodes for K and α was 27 and 10, respectively. The number of collocations nodes was determined so that the max discrepancy between the left and the right hand side of the Bellman equation was in the order of magnitude of 10^{-7} (the actual value is 2.5×10^{-7}). After this point further increments in the precision required large increases in the number of collocation nodes and thus in the computation time. In the case of the collocation nodes for K, they were shifted to the left so that the smaller collocation node coincided with the lower limit of the interpolation range for K, in order to help in the abnormal lack of precision in that region.

To find the optimal policy and respective optimal values for a given value function approximation, it was used Newton's method, with the derivatives of the optimand with respect to the control computed analytically. The computation of the analytical derivative of the value function approximation with respect to any of the state variables is straightforward, since it is a polynomial on the state variables, and the derivatives are itself polynomials.[36] The criterion to stop the Newton's iteration was that the maximum absolute value of the Newton's update on the control was less than 10^{-7}.

Finally, the criterion to stop the value function iteration was that the norm of difference between the value function at the collocation nodes in two consecutive steps were less than 10^{-7}.

The solution to the myopic firm's problem is actually quite similar to that of the full information problem, despite the existence of parameter uncertainty. In the myopic case $a_{t+1} = a_t$ and $v_{t+1} = v_t$ and so the only difference for the full information case is that instead of knowing α, the firm considers $\alpha \sim N(a_t, v_t)$. The solution is obtained exactly in the same way, substituting the stochastic operational cash flow by its average over the independent random variables z_t and α.

[33] The full information optimal stock of capital is 1 somewhere between 0.62 and 0.63.

[34] Chebyshev polynomial approximations perform very poorly in the presence of kinks in the value function.

[35] See Miranda and Fackler (2004) for details on how Chebyshev nodes are computed. Other collocation nodes schemes, like equidistant nodes, were tried without much success (less precision and difficulties in the convergence of the value function).

[36] See Miranda and Fackler (2004) for the details on how to map the coefficients of the Chebyshev polynomial into the coefficients of its derivatives.

C Details of the Solution with Parameter Uncertainty

In order to eliminate the productivity shock as a state variable, the firm's problem with parameter uncertainty is simplified in the same way as the full information problem. Thus the stochastic operational cash flow is substituted by its expected value. However, unlike in the full information case, one must deal with the effect of the next period's productivity shock in the observed signal (S_{t+1}) and thus on next period's state ($K_{t+1}, a_{t+1}, v_{t+1}$), through a_{t+1}.[37] For a given current state, each of the possible values of a_{t+1} is going to be associated to a specific value of the productivity shock. Therefore, when computing the expectation of next period firm's value, we have to subtract from each $V(K_{t+1}, a_{t+1}, v_{t+1})$ the expected value of the operational cash flow and add back the operational cash flow resulting from the observation of the signal S_{t+1} associated to the respective state a_{t+1}. The same kind of adjustment to the modified problem's solution is made to obtain the firm's value when the current operation cash flow is other than its mean value.

The range of interpolation for the state variables K and a are the same as those for K and α in the full information case, and for the parameter uncertainty is $v \in [0, 0.25^2]$. Like in the full information case, it was used Chebyshev collocation nodes, with the nodes for K as well as v shifted to the left limit of their interpolation range (the latter to ensure a good precision for comparison purposes with the full information solution).[38] The number of collocation nodes for K, a and v is 27, 10 and 15, respectively.

To obtain the optimal policies and respective optimal firm values for a given value function approximation it was used the Newton's method. However, unlike in the full information case, the analytical derivatives were too complex to be used, and so the choice was to use two-sided numerical derivatives, with perturbation in the control equal to 10^{-4}.[39] The criterion to stop the Newton's iteration was that the maximum absolute value of the Newton's update on the control was less than 10^{-5}.

To compute the expectation in the continuation value, it was used Gaussian integration with 5 points, which works quite well due to the smoothness of the integrand. The addition of a sixth integration point changes the expectation by no more than 10^{-4}, which is small enough to be worth the additional computation time.[40]

Although the three-dimensionality of the problem makes the numerical procedure (two orders of magnitude) more computationally intensive than the full information problem, it is the unboudedness of the state variable a_t transitions that makes the solution to the problem challenging. When the firm is in a specific state, defined by the triple (K_t, a_t, v_t), the transition to the next period's state is deterministic along the K and v dimensions, but stochastic in the a dimension. K_{t+1} is directly chosen by the firm and v_{t+1} is completely determined by its current value v_t and by the control K_{t+1}. However, a_{t+1} depends on the stochastic signal S_{t+1}, and it can assume any real value. As a consequence, no matter how large the interpolation range for the a state variable is, it will always be necessary to evaluate the value function approximation outside the interpolation range. This constitutes a problem, because in general polynomial interpolants provide awful approximations to the value function outside the interpolation range. As a result, one needs to use a scheme to extrapolate the value function outside the interpolation range in the a dimension. In the v dimension such a problem does not exist and in the K dimension it is easy to avoid it by setting the upper limit of the

[37] This makes impossible to reduce the stochastic dynamic programming problem into a deterministic one and further simplify the problem.

[38] In the case of parameter uncertainty there was the additional option of defining the collocation nodes based on $\sqrt{v}$. However, the precision and the convergence of the approximation is considerably worse than when using collocation nodes based on v.

[39] Approximately the quartic root of the machine precision, as suggested in Miranda and Fackler (2004).

[40] Moreover, the addition of several integration points may even deteriorate the accuracy of the expectation, since the larger the number of integration points, the more the extreme points fall way into the extrapolation region (to be discussed next), where the accuracy is inevitably not very good.

interpolation range high enough. Given that the computation of the expectation uses a limited number of points, the unboudedness of a is not as big of a problem as it could be, since the value function only needs to be extrapolated in the neighborhood of the interpolation range.

In simple terms, to extrapolate the value function one starts with the respective full information value (the limiting firm's value when the parameter uncertainty converges to zero), which is easily obtained, and then add the effect of parameter uncertainty. To obtain an approximation for the value added by the parameter uncertainty it was used the following observations: (i) the value added by the parameter uncertainty is positive and a direct consequence of the convexity in the full information firm's value to α relation; (ii) the value added by the parameter uncertainty is increasing in that convexity; and (iii) the convexity is increasing with α.[41] This means that the information about the local curvature of the full information value function at $\alpha = a$ can be used to assess how much value is added by parameter uncertainty.

The actual procedure used to extrapolate the value functions is as follows. (i) Run OLS regression of the difference between the parameter uncertainty firm's value and the corresponding full information values into the second derivative of the full information values with respect to α. In order to improve the R^2 of the regression, the regression used only values of α at most 0.01 apart from the respective interpolation range limit. The regression R^2's are above 99.99%. (ii) Using the OLS coefficients estimated in step 1 and the second derivative of the full information value function at α equal to the value of a for which the value function is being extrapolated, estimate the value added by parameter uncertainty. It should be noted that at the limits of the interpolation range for a, the actual value added by the parameter uncertainty is marginally above the fitted value. Thus, the estimates obtained from this procedure are slightly conservative. (iii) Add the parameter uncertainty to the respective full information value to obtain the extrapolation.

The above procedure was only used for the extrapolation of a above the interpolation range. The same procedure could have been just as well used for the extrapolation of a below the interpolation range if it wasn't for the discontinuity in the derivatives produced by the minimum stock of capital constraint. Therefore the value added by parameter uncertainty was estimated by linear extrapolation of those values in the lower region of the interpolation range. Although the relationship between the value added by parameter uncertainty and α is convex, this approximation is suitable since at the lower end of the interpolation region this relation is almost linear. To minimize the effects of the convexity, only values of α at most 0.01 apart from the lower endpoint were used to perform the linear extrapolation.

Finally, the criterion to stop the value function iteration was that the norm of difference between the value function at the collocation nodes in two consecutive steps were less than 10^{-5}.

[41] These observations were obtained by solving the firm's problem with a very small level of parameter uncertainty. Since the larger Chebyshev collocation node is smaller than the upper limit of the interpolation range, if the level of parameter uncertainty is small enough, a_{t+1} will not fall outside the interpolation region even when a_t is equal to the larger Chebyshev node. The intuition behind these observations is provided in the main text.