Corporate Investment Over the Business Cycle

Thomas Dangl Youchang Wu

Jan 31, 2013

Abstract

We investigate the dynamics of corporate investment at the firm and aggregate levels. We find that the average capital growth rate across firms exhibits negative spikes, despite the positive spikes at the firm level. Furthermore, while individual firms climb up their investment spikes and come down at an equal speed, the average capital growth rate declines to the trough much faster than it recovers. We develop a dynamic model of investment that replicates these patterns. The model features costly reversibility, cyclical macroeconomic shocks, and uncertainty about the true state of the economy. The interaction between the expected future growth and the option value of waiting leads to a convex relation between a firm’s optimal capacity and its posterior belief of being in an expansion. The endogenous distribution of firms relative to their optimal capacities causes firms in aggregate to react more strongly to negative signals during an expansion than to positive signals during a recession.
Capital investment is one of the most important decisions in corporate finance. It not only determines the long-term prospect of a corporation and the welfare of its shareholders, but also have profound macroeconomic implications. Despite extensive research on this subject in both economics and finance, what drives corporate investment remains an elusive question. This paper addresses a particular feature of corporate investment behavior, i.e., the strikingly different patterns of investment at the firm and aggregate levels. It is well-known that investment of individual firms, measured by the real capital growth rate, tend to have big positive spikes occasionally. However, once we average the capital growth rates across firms, we find the resulting time series exhibits strong negative spikes. Furthermore, while individual firms climb up their investment spikes and come down at an equal speed, the average capital growth rate declines to the trough much faster than it recovers.

We uncover these patterns of corporate investment using the quarterly Compustat-CRSP merged database from 1975 through 2011. We first plot the average capital growth rate around the trough quarters dated by the National Bureau of Economic Research. We find that the decline before the trough is much sleeper than the subsequent recovery. It takes sixteen quarters for the investment activity, measured by either net capital growth rate or capital expenditure normalized gross value of capital stock, to get back to the level six quarters prior to the trough. We then test statistically whether the distributions of the average capital growth rate and its slope (i.e., first-order difference) are symmetric, and find both of them are negatively skewed, especially when growth rate is measure over the time intervals of three or quarters. The negative skew of the average capital growth rate suggests troughs are further below the mean than peaks are above (level asymmetry), while the negative skew of its slope suggests that downturns are steeper than upturns (slope asymmetry). By contrast, capital growth rates of individual firms exhibit strong positive skewness in levels and virtually no skew in first-order differences, indicating positive spikes in corporate investment and symmetric upward and downward slopes.

These striking patterns pose interesting research questions. In particular, what are the economic mechanisms that lead to such sharp differences of investment behavior at the firm and aggregate levels? Can we reconcile these patterns in a unified framework of optimal investment? These are the questions we explore in this paper.

\[1\text{See Dixit and Pindyck (1994) and Stein (2003) for two good surveys of the literature.}\]
We develop a dynamic model of costly reversible investment with cyclical macroeconomic shocks and incomplete information. Our model captures three important features of the real world: first, the profitability of an individual firm is strongly influenced by macroeconomic conditions; second, firms often face substantial uncertainty about the true state of the economy; third, disinvestment is more costly than investment. We consider a cross-section of firms, each facing its own business conditions that are summarized by a random demand factor. The expected growth rate of each firm’s demand factor depends on the state of the economy, which shifts between a high growth state (expansion) and a low growth state (recession) at random times. Firms are risk-neutral, and have an infinite time horizon. Capital can be expanded incrementally and instantaneously at any time at a constant marginal cost, but can only be sold at a discount, i.e., the resale price of capital is lower than its purchase price. Furthermore, the true state of the economy is not directly observable. Firms update their beliefs about the state of the economy by observing their own operating profits and a public signal. We model this updating process in a continuous-time Bayesian learning framework.

We first derive the optimal investment/disinvestment policy of a typical firm in such an environment. Under our assumptions, the optimal investment/disinvestment policy is characterized by an upper bound and a lower bound on the firm’s capital stock normalized by its current demand. The lower bound represents the firm’s optimal normalized capacity, while the upper bound represents its maximum tolerated normalized capacity. Both boundaries act as a reflecting boundary. The firm takes no action as long as its normalized capital stock is between these two boundaries, and expands (shrinks) its capital stock instantaneously once it hits the lower (upper) boundary. Since demand grows faster during an expansion period, the firm’s optimal normalized capacity increases in the posterior belief of being in an expansion. Furthermore, this relation is nonlinear. When the posterior probability of an expansion is low, its increase leads to only a modest increase of optimal capacity. However, when this posterior probability is high, its decrease leads to a sharp decline. By contrast, the disinvestment boundary, which represents the maximum tolerated normalized capacity, is concave in the belief.

The convex relation between the optimal capacity and the belief is a result of an interaction between the expected growth rate and the option value of waiting associated with
time-varying uncertainty about the true state of the economy. The intuition is as follows. When the firm is almost sure of being in an expansion, the arrival of a negative signal generates two effects that reinforce each other. First, it reduces the conditional probability of being in an expansion, thus lowering the expected future demand growth rate. Second, it increases the conditional variance of the belief, i.e., it makes the firm less certain about the true state of the economy. The greater uncertainty generates a higher option value of waiting. These two effects together greatly reduce the firm’s willingness to invest. By contrast, when the firm is almost sure of being in a recession, the arrival of a positive signal generates two effects that partially offset one another. While the signal increases the expected demand growth rate, it also brings more uncertainty about the current state of economy, thus increasing the option value of waiting. As a result, the firm adds capacity only modestly.

The intuition for the concavity of the disinvestment boundary is similar. When a firm receives a bad signal in a good time, the lower expected demand growth induces an incentive to disinvest, yet the increased uncertainty induces an incentive to wait. As a result, the firm does not disinvest aggressively. On the other hand, when a firm receives a good signal in a bad time, both the higher expected growth rate and increased uncertainty weaken the firm’s incentive to disinvest, therefore the maximum tolerated capacity increases significantly.

Interestingly, when signals about the state of the economy are less noisy, the nonlinearities of both investment and disinvestment boundary become more pronounced. This is because when the signals are more precise, the benefit of waiting for the next signal, which can resolve a lot of uncertainty, is higher. Thus firms do not rush to invest or disinvest. They behave conservatively with respect to both actions. That is, they do not invest (disinvest) unless they are very confident that the economy is in an expansion (recession). This means both boundaries bend outward.

After deriving the optimal investment/disinvestment policy, we simulate a large number of firms following this strategy. These firms experience the same macroeconomic shocks, observe the same public signal, but also face heterogeneous shocks to their own business conditions. We examine the patterns of capital growth rate at both the firm and aggregate

---

[Veronesi (1999)] uses a similar argument to explain asymmetric responses of stock prices to signals. He does not consider firm investment. The asymmetry arises in his model due to risk aversion rather than option value of waiting.
levels. Our simulated data display patterns that match the empirical patterns remarkably well. The average capital growth rate declines fast and recovers slowly. It exhibits negative skewness both in levels and in first-order differences, while the capital growth rates of individual firms are strongly positively skewed in levels and have virtually no skew in first-order differences.

The positive skewness of the capital growth rate at the individual firm level is a natural outcome of costly reversibility, which implies that the expansion of capacity is unconstrained, while the decline of capital stock is costly. The negative skew of the average capital growth rate comes from the fact that expansions last longer than recessions, which implies that the bulk of observations lie to the right of the mean. The sharp-decline-slow-recovery feature, i.e., the slope asymmetry, of the average capital growth rate is mainly due to the endogenous distribution of firms relative to their optimal capacities. During a recession, many firms are farther away from the investment boundary due to the lower demand growth rate and the lower optimal capacity. These firms suffer from excess capacities, and do not increase capital even if their beliefs change significantly after a positive signal. On the contrary, during an expansion more firms are pushed to the investment boundary by the high demand growth, so more firms are affected when a negative signal arrives.

Our model builds on the structure of Guo, Miao, and Morellec (2005), who investigate optimal irreversible investment under random regime shifts. Our paper differs from theirs in several important aspects. Their work focuses entirely on the investment policy of an individual firm, and is purely theoretical. We study a large cross-section of firms facing both common and heterogeneous shocks, present empirical patterns on both the firm and aggregate levels, and calibrate our model to replicate the major features of the data. Furthermore, we extends their model setup in two important dimensions. These extensions not only allow us to better fit the empirical data, but also bring new economic insights. First, while they assume regime shifts to be perfectly observable, we assume that firms can only infer the true regime from noisy signals. This allows us to study how time-varying uncertainty affects firms’ investment policy. As our simulation results suggest, in the absence of uncertainty about the true state of the economy, the average capital growth rate shoots up immediately once the low growth state is over. Second, while investment is completely irreversible in Guo, Miao, and Morellec (2005), we allow for different degrees
of reversibility, encompassing complete irreversibility as a special case. This allows us to investigate the economic impacts of reversibility. Partial reversibility reduces, even though not sufficiently, the positive skewness of capital growth rates at the firm level, thus allowing a better fit with the empirical counterpart.

Our paper contributes to the understanding of corporate investment under cyclical macroeconomic shocks and incomplete information. Several recent papers study optimal investment under incomplete information. For example, Altı (2003) investigates the optimal investment policy when the productivity of capital is unknown and has to be inferred from the realized cash flow. Decamps, Mariotti, and Villeneuve (2005) and Klein (2009) study the timing of a one-shot investment whose value is governed by an unobservable parameter. Unlike us, these authors do not allow for regime shifts. As a result, uncertainty declines monotonically over time, and the models are unable to explain the cyclical features of investment. Grenadier and Malenko (2010) develop a model of optimal investment in which firms are unable to distinguish between temporary and permanent shocks to the cash flow process. They show that augmenting the traditional uncertainty over future shocks with Bayesian uncertainty over the nature of past shocks gives rise to a number of novel implications for firms’ investment behavior.

Our work also contributes to the literature on business cycle asymmetry. There is a long debate about whether economic downturn is generally more abrupt and violent than the upturn. The empirical evidence is somehow mixed. While Neftci (1985), Sichel (1993) find supportive evidence for the asymmetry hypothesis, Falk (1986) and DeLong and Summers (1986) conclude that there is very little evidence of asymmetry. These studies typically examine the macroeconomic data. Our empirical analysis of the firm-level data provides strong evidence of asymmetry in one of the key business cycle variables, i.e., corporate investment. Our microeconomic model of incremental investment reconciles the contrasting patterns of capital growth rates at the individual firm level and in the aggregate of a large cross-section of firms. Treating macroeconomic conditions as an exogenous

---

3One of the most famous statements is made by Keynes (1936): “There is, however, another characteristic of what we call the trade cycle which our explanation must cover; namely, the phenomenon of the crisis – the fact that the substitution of a downward for an upward tendency often takes place suddenly and violently, whereas there is, as a rule no such sharp turning point when an upward is substituted for a downward tendency.”
latent process, we do not aim to give an explanation for the source of such fluctuations. However, our model describes how firms at the micro level optimally respond to changing economic environments, and how such responses translate cyclical shifts of latent regimes into asymmetric time series of capital growth rates.

Previous explanations for the business cycle asymmetry typically rely on some sort of exogenously assumed externality. More recently, Bloom, Bond, and Reenen (2007) show that higher uncertainty reduces the responsiveness of investment to demand shocks, and Bloom (2009) shows that macro uncertainty shocks can generate sharp recessions and recoveries through their impacts on firms’ investment and hiring decisions. Unlike these studies, which take the degree of uncertainty as exogenously given, we model the evolution of uncertainty as an endogenous process resulting from optimal updating of beliefs. While uncertainty is high whenever there is a shift in the underlying economic environment, the interplay between the first and the second moments of conditional beliefs leads to asymmetric reactions to signals at different turning points of the business cycle.

The rest of our paper is organized as follows. Section 1 presents empirical evidence for the asymmetries of capital growth rates at the firm and aggregate levels. In section 2 we present our model, derive the optimal learning process, and derive the partial deferential equation and boundary conditions of the optimal investment policy. In section 3 we calibrate the parameters for the base case of the model. Section 4 characterizes the optimal investment policy. Section 5 compares the simulated and empirical patterns of the capital growth rate at both the firm and aggregate levels. Section 6 presents comparative static results with respect to depreciation and reversibility. Section 7 concludes the paper.

---

4Explanations of the asymmetry based on externalities in the production process include Gale (1996) and Acemoglu and Scott (2003). Gale (1996) shows that the dependence of a firm’s profit on the general level of economic activity leads to strategic delays in investment during a recession, which prolongs the process of a recovery. Acemoglu and Scott (2003) show that the dependence of a firm’s investment costs on its past activity slows down the upturn and amplifies shocks in the trough of a business cycle. Explanations based on informational externalities include Chamley and Gale (1994), Chalkley and Lee (1998), Veldkamp (2005), and Nieuwerburgh and Veldkamp (2006). A general theme of this latter strand of literature is that a firm’s investment activity generates information for other firms or improves the information value of public signals.
1 Asymmetry of Firms’ Capital Growth Rates: Empirical Patterns

1.1 Capital Growth Rates Over the Business Cycle

To explore the empirical patterns of firms’ investment behavior, we examine the quarterly Compustat-CRSP merged database from 1975 through 2011. We exclude all financial firms (SIC codes between 6000 and 6999), utilities (SIC codes between 4900-4999), government entities (SIC codes greater than or equal to 9000). We use a firm’s net property, plant and equipment (Compustat data item PPENT) to measure its net capital stock, and use the growth rate of net capital to measure its investment behavior. The is a normalized measure of net investment, i.e., gross investment minus depreciation. All nominal values of are converted into year 2005 dollars using the quarterly GDP deflator. A firm enters the sample after its PPENT reaches $1 million (in year 2005 dollars), stays in the sample even if its PPENT drops below it later. Firm-quarters with missing PPENT data are excluded.

Since we are interested in capital growth arising from physical investment, which is different from growth through mergers and acquisitions, we excludes firms that are heavily involved in mergers and acquisitions. For this purpose, we match our sample to the SDC Mergers&Acquisition database of Thomson Reuters, which has the most comprehensive coverage of M&A deals since 1980 to 2011. If a firm acquires assets worth more than 20% of its total asset value in a given fiscal year, we exclude that fiscal year from our sample. Our final sample consists of 532,306 firm-quarter observations, with an averages of 3597 firms in each quarter.

We measure a firm’s capital growth by the continuously compounded growth rate of PPENT. Specifically, a firm’s capital growth rate in quarter $t$ is defined as

$$ g_{t,t-1} = \ln(PPENT_t) - \ln(PPENT_{t-1}). $$

A simple percentage growth rate is inherently asymmetric, since it cannot go below -100%.

---

5Before 1975, quarterly data on firms’ capital stock are limited. We have also analyzed empirical capital growth rates using the annual Compustat-CRSP merged database from 1950 through 2011, and the patterns are similar.
due to the nonnegativeness of capital stock. The continuously compounded growth rate does not have this problem. To limit the impact of extreme outliers or potential data errors, we exclude the growth rates that are below the 1st or above the 99th percentile of the whole sample. We average across firms (equally weighted) to get a time series of quarterly capital growth rates.

We first investigate the patterns of the average capital growth rate over the business cycle. We select an event window of 23 quarters around the quarters that are identified by the National Bureau of Economic Research (NBER) as a business cycle trough. According to the NBER, there are six cycles between 1975 and 2011. A trough is a turning point that marks the end of a contraction and the start of an expansion. Our event window goes from six quarters before a trough through sixteen quarters after it.

Panel (a) of Figure 1 shows the average capital growth rate, further averaged across the six cycles, in each quarter of the event window (from -6 to 16). As one can see, the downward slope before trough is very steep, especially starting from quarter -4. The recovery after the trough is slower. It takes 16 quarters for the average capital growth to get back to the level of quarter -6. This pattern of fast decline and slow recovery is consistent with Keynes’s observation of business cycle asymmetry.

Panel (b) of Figure 1 plots the average gross investment rate around the trough. The gross investment rate is defined as

\[ g'_{t, t-1} = \log(PPEGT_{t-1} + CAPX_t) - \log(PPEGT_{t-1}) \]

where \( CAPX \) is capital expenditure minus sale of property, plant and equipment, measured in year 2005 dollars, \( PPEGT \) is the gross value of property, plant and equipment. The pattern of the gross investment rate is very similar to that of the net investment rate. In the rest of the paper we focus on the net capital growth rate because the availability of a longer times series of data.

The six trough quarters are 1975 Q1, 1980 QIII, 1982 QIV, 1991 QI, 2001 QIV, and 2009 QII.

The quarterly CAPX data is only available starting from 1984.
Figure 1: Average net capital growth rate and gross investment rate over the business cycle

Panel (a) shows the average capital growth rates around the business cycle troughs dated by the NBER. Panel (b) shows the average gross investment rates. Capital growth rate and gross investment rate are defined in equations (1) and (2), respectively, and are estimated using the quarterly Compustat-CRSP merged database from 1975 through 2011.
1.2 Skewness of Capital Growth Rates

Figure 1 shows that the upturn and downturn of firms’ capital growth rates are asymmetric. To quantify the asymmetry, we use a standard measure of asymmetry in statistics, i.e., skewness. The skewness of a random variable is defined as its third central moment normalized by the third power of its standard deviation. A negative skewness value indicates that the left tail of a distribution is longer than the right tail and the bulk of the observations lie to the right of the mean. A positive skewness value indicates the opposite. A symmetric distribution has zero skew.

Following Sichel (1993), we distinguish between two types of asymmetry: level asymmetry (deepness), and slope asymmetry (steepness). Level asymmetry refers to the characteristic that troughs are farther below the trend than peaks are above, while slope asymmetry refers to the characteristic that contractions are steeper than expansions. We use the skewness of the capital growth rate to measure level asymmetry, and the skewness of its first-order difference to measure slope asymmetry. We examine these asymmetries both at the individual firm level and in the aggregate.

To capture the asymmetry of capital growth rates over different time intervals, we extend the definition in (1) to account for the capital growth over multiple quarters, following Nieuwerburgh and Veldkamp (2006). Specifically, the $n$-quarter capital growth rate is defined as

$$g_{t,t-n} \equiv \ln(PPENT_t) - \ln(PPENT_{t-n}),$$

(3)

where $n$ varies from 1 to 10. The first-order difference of the $n$-quarter capital growth rate is defined as

$$\Delta g_{t,t-n} \equiv g_{t,t-n} - g_{t-n,t-2n}.$$ 

(4)

For the asymmetry at the firm level, we first calculate the skewnesses of the $n$-quarter capital growth rate and its first-order difference for each individual firm, and then average these measures across all firms with at least 40 observations. For the asymmetry at the aggregate level, we first calculate the average $n$-quarter capital growth rate across firms, for each $n$ from 1 to 10; we then use the skewness of the average $n$-quarter growth rate to

---

8 We exclude the observations that are below the 1st or above the 99th percentile of the whole sample.
measure the level asymmetry, and the skewness of its first-order difference to measure the slope asymmetry.

Panels (a) and (b) show the skewnesses of firm-level capital growth rates and their first-order differences, respectively. Panels (c) and (d) show the skewnesses of average capital growth rates and their first-order differences, respectively. The length of the time interval over which the growth rate is measured varies from 1 to 10 quarters. To test whether the values are statistically different from zero, we also plot the 95% confidence interval of each estimate, which is within 1.96 standard errors of the point estimate. For the skewness estimates at the firm level, the standard error is equal to the standard deviation of the skewnesses across firms divided by the squared root of the number of firms. At the aggregate level, there is only one skewness estimate for each time interval, and its standard error cannot be easily computed due to the strong autocorrelation of capital growth rates. We therefore follow the Monte Carlo procedure of DeLong and Summers (1986). First, a fifth-order autoregressive model for each time series of growth rates is estimated. Second, the estimated model is used to generate 1000 artificial series for the sample period under the assumption that the shocks to the autoregressive process were independent and normally distributed. Third, the standard deviation of the skewnesses across the simulated series is used as the standard error of the skewness estimate under the null hypothesis of no asymmetry.

A striking feature of Figure 2 is that there is a sharp contrast between the patterns at the firm level and those at the aggregate level. Consistent with the sharp decline and slow recovery observed in Figure 1, average capital growth rates exhibit both level asymmetry (Panel (c)) and slope asymmetry (Panel (d)). The estimated values of skewness are below -0.5 for all the ten time intervals we examine. For time intervals up to five quarters, they are significantly negative at the 95% level. This indicates that troughs are deeper below than peaks are above the mean. The first differences of average capital growth rates are also negatively skewed, for time intervals up to eight quarters, indicating that the slope of the downturn is in general steeper than the slope of the upturn, lending support the observation of Keynes (1936) that recessions arrive violently while recoveries only occur slowly. The estimated value of skewness first decreases and then increases with the length of the time interval. At the quarterly interval ($n = 1$), it is -0.22, insignificantly different.
Panels (a) and (b) show the skewnesses of firm-level capital growth rates and their first-order differences, respectively. Panels (c) and (d) show the skewnesses of average capital growth rates and their first-order differences, respectively. The length of the time interval over which the growth rate is measured varies from 1 to 10 quarters. The dashed curves show the upper and lower bounds of the 95% confidence interval of each estimate of skewness. The estimation is based on the quarterly Compustat-CRSP merged database from 1975 through 2011.
from zero. It becomes significantly negative for \( n = 2, 3 \) and 4. This suggests that while the decline of the capital growth rate in a given quarter is not be very severe, the consecutive declines over multiple quarters generate outliers in the left tail of the distribution. As \( n \) increases further, the negative skewness disappears gradually, suggesting that the slope asymmetry is gradually smoothed out as the length of time interval increases.

By contrast, capital growth rates at the firm level exhibit a strong level asymmetry with a positive skewness (Panel (a)) and almost no slope asymmetry (Panel (b)). The estimated skewness at one quarter horizon is 0.71. The estimate decreases steadily as the time interval lengthens, but remains positive even at the 10-quarter interval. This suggests that for an individual firm, it is easier to expand capital than cut it back, consistent with idea of costly reversibility. The skewness of the first-order difference is below 0.03 in absolute value for all time intervals, indicating that at the firm level, changes in capital growth rates are symmetric.

The negative skew of the average capital growth rate, both in the level and in the slope, and the positive skew of the growth rate at the firm level pose an interesting question for theoretical research: can these two almost opposite features be reconciled in a unified model of optimal firm investment? In the next section, we present such a model.

# 2 The Model

## 2.1 Setup

We consider a cross-section of risk-neutral firms with an infinite time horizon. The firms are \textit{ex ante} identical but face heterogeneous demand shocks. A typical firm operates in an environment similar to that of Guo, Miao, and Morellec (2005), except that it does not observe the true state of the economy and that investment is not entirely irreversible. Time is continuous. Investment is incremental. Each firm’s cash flow is driven by a distinct stochastic demand factor. Depending on the state of the macroeconomy, the expected growth rate of the demand factors of all firms shifts between a high level (expansion) and a low level (recession) at random times.

We now describe the model setup and the optimization problem from the perspective
of a typical firm in the cross-section.

**Cash Flows:** The operating income (before depreciation) of the firm is assumed to be given by a linearly homogeneous function \( f : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+ \) satisfying:

\[
f(x_t, k_t) = \frac{1}{1 - \alpha} x_t^\alpha k_t^{1-\alpha},
\]

where \((k_t)_{t \geq 0}\) represents the process of the firm’s capital stock, \((x_t)_{t \geq 0}\) represents the process of a demand factor.\(^9\) Assuming that the firm’s output is nonstorable, equation (5) can be interpreted as the profit of either a price-taking firm with decreasing returns to scale, or a monopolist facing constant returns to scale and a constant elasticity demand curve (see Abel and Eberly (1996) and Morellec (2001)).

**Demand Shocks:** Assume that the demand factor for the firm, \(x_t\), evolves according to the stochastic differential equation:

\[
\frac{dx_t}{x_t} = \mu_t dt + \sigma_x dW_{xt}, \quad x_0 > 0,
\]

where \(W_{xt}\) is a standard Wiener process. The volatility \(\sigma_x\) is a known constant.\(^{10}\) The expected growth rate of the demand factor, \(\mu_t\), is determined by the macroeconomic condition. It is low (\(\mu_t = \mu_l\)) in a recession and high (\(\mu_t = \mu_h > \mu_l\)) in an expansion. Within a given state of the economy, the demand factor follows a standard geometric Brownian motion.

The macroeconomic condition switches between expansions and recessions at random times. Correspondingly, \(\mu_t\) switches randomly between \(\mu_h\) and \(\mu_l\). More specifically, we assume that \(\mu_t\) is driven by a continuous time Markov jump process with the transition probabilities:

\[
P(\Delta t) = \mathbb{I} + \begin{pmatrix} -\lambda_{h,l} & +\lambda_{h,l} \\ +\lambda_{l,h} & -\lambda_{l,h} \end{pmatrix} (\Delta t + o(\Delta t)),
\]

where \(\mathbb{I}\) is the identity matrix and \(\lambda_{h,l}\) and \(\lambda_{l,h}\) are the constant instantaneous intensities.

---

\(^9\)More generally, \(x_t\) can be interpreted as an index of business conditions, which is positively related to the strength of the demand for the firm’s product or the firm’s productivity, and negatively related to the firm’s cost of factors other than capital.

\(^{10}\)Note that \(\sigma_x\) is the same in both states. Since the volatility of a process can be estimated almost instantaneously in continuous time, firms would learn the true state of the economy almost instantaneously if \(\sigma_x\) differs across states.
of transition from $\mu_h$ to $\mu_l$ and vice versa. The magnitudes of $\lambda_{h,l}$ and $\lambda_{l,h}$ determine the persistence of the expansion and the recession, respectively. The lower the transition intensity, the higher the persistence.

**Investment, Disinvestment and Depreciation:** As in [Abel and Eberly (1996)](#), a firm can add capital incrementally and instantaneously at a constant marginal cost, which is normalized to be one; it can also sell capital at a price $b$ lower than one. The wedge between the purchase and sell prices of capital, $1 - b$, captures the feature of costly reversibility. Installed capital depreciates at a constant rate of $\xi$.

**Information:** It is assumed that the firm is able to observe its own realized demand factor $x_t$ through the realized operating profit $f_t$, but not its expected growth rate $\mu_t$. In other words, the true state of the economy that determines $\mu_t$ is a hidden process. This implies that $W_{xt}$ is not observable as well. The firm’s problem is that when observing a certain increase or decrease in the cash flow, it is not immediately clear which portion of this observed change comes from the current growth rate $\mu_t$ and which portion comes from the noise terms $W_{xt}$. Yet the distinction between these possible sources is of core relevance for the firm’s investment decision. Since disinvestment is costly, investment depends not only on the current demand factor, but also on expectations about its future growth rate.

In reality, a firm’s own profit is clearly not the only source of information about the state of the economy. To summarize the other sources of information available to the firm, we assume that the firm can also observe a public signal, $s_t$, which evolves according to the stochastic differential equation:

$$
\frac{d s_t}{s_t} = \mu_t dt + \sigma_s d W_{st}, \ s_0 > 0; 
$$

where $W_{st}$ is a standard Wiener process, $\sigma_s$ is a publicly known constant representing the volatility of the signal process. We also assume the instantaneous correlation between the two Wiener processes, $W_{st}$ and $W_{xt}$, to be a known parameter, $\rho \in [-1, 1]$. Note that the drift terms in equations (6) and (8) are identical. This assumption is made for simplicity. Our main results remain unchanged if we allow for imperfect correlation between

\footnote{The limitation to only two states is not essential from a technical point of view and can easily be relaxed to a finite number of states.}
the expected growth rates of \( x_t \) and \( s_t \).

Allowing the firm to observe a signal other than its own profit not only makes our model more realistic, but also allows us to separate the effect of information quality from the effect of demand volatility. The instantaneous volatility, \( \sigma_s \), measures the noisiness of the public signal. It characterizes (inversely) the accuracy of the publicly available data about the economy. By varying \( \sigma_s \) without changing the volatility of the demand factor, \( \sigma_x \), we can analyze the effect of information quality on the firm’s investment. If the firm’s own cash flow process is the only source of information, then the effects of information quality and demand volatility are coupled, because the information about the current state of the economy is more accurate under lower volatility of the demand factor.

To summarize the information structure of our model, let \( \mathcal{F}_t \) be the canonical non-decreasing filtration jointly created by the firm’s own demand factor process \( x_t \) and the public signal \( s_t \). Our assumptions about observability imply that \( \mu_t, dW_{xt}, \) and \( dW_{st} \) are not measurable with respect to the information set \( \mathcal{F}_t \).

### 2.2 Learning About the State of the Economy

Observing the firm’s demand factor \( x_t \) and the signal \( s_t \) over the time interval \([0, t)\), an agent can continuously update its belief about the state of the economy. To formalize the rational learning rule, we denote by \( \pi_t \) the probability that \( \mu_t = \mu_h \) conditional on \( \mathcal{F}_t \) and a prior \( \pi_0 \). Consequently, the rational belief about the current state of the economy is given by \( \pi_t \mu_h + (1 - \pi_t) \mu_l \). Unexpected changes in \( x_t \) and \( s_t \) give rise to an update of the belief.

Learning under this circumstance is a standard nonlinear filtering problem and can be characterized by the proposition below.\(^{12}\)

**Proposition 1.** The optimal updating of the belief satisfies the stochastic differential equation:

\[
d\pi_t = \left[ -\pi_t \lambda_{h,l} + (1 - \pi_t) \lambda_{l,h} \right] dt + (\mu_h - \mu_l)\pi_t(1 - \pi_t) \Phi^{-1} dW_t, \tag{9}\]

\(^{12}\)David (1997) is one of the first papers in finance that applies this technique to model the learning of unobservable regime shifts.
where $W^F_t$ is a two-dimensional independent Wiener process with respect to $F_t$, defined as

$$dW^F_t \equiv \begin{pmatrix} dW^F_{xt} \\ dW^F_{st} \end{pmatrix} \equiv \Phi^{-1} \left( \begin{pmatrix} \frac{dx_t}{x_t} - E(\mu_t | F_t)dt \\ \frac{ds_t}{s_t} - E(\mu_t | F_t)dt \end{pmatrix} \right),$$

(10)

$\Phi$ is a $2 \times 2$ matrix satisfying

$$\Phi \Phi' = \begin{pmatrix} \sigma_x^2 & \rho \sigma_x \sigma_s \\ \rho \sigma_x \sigma_s & \sigma_s^2 \end{pmatrix},$$

and $1$ is a two-dimensional column vector with both elements equal to 1.

**Proof.** See Theorem 9.1 in Liptser and Shiryaev (2001) for the basic filtering equation, and equation (A1) in Veronesi (2000) for an extension to the vector case. Equation (9) is obtained by applying equation (A1) in Veronesi (2000). The independence between $dW^F_{xt}$ and $dW^F_{st}$ can be established by noting

$$dW^F_{xt} dW^F_{st} = (1,0) \Phi^{-1}(\Phi \Phi' dt)(\Phi^{-1})'(0,1)' = 0.$$

The diffusion process $\pi_t$ is bounded between 0 ($\mu_t = \mu$, almost sure) and 1 ($\mu = \mu_h$, almost sure). The drift term in equation (9) indicates that in the absence of information shocks, there is a tendency for the belief to revert toward the unconditional mean:

$$\bar{\pi} = \frac{\lambda_{l,h}}{\lambda_{h,l} + \lambda_{l,h}},$$

(11)

which satisfies $[-\pi_t \lambda_{h,l} + (1 - \pi_t) \lambda_{l,h}] = 0$. Therefore, the impact of any particular information shock decays gradually over time.

The diffusion term in equation (9) characterizes the response of the belief to unexpected changes in the realized demand factor $x_t$ and the signal $s_t$. $E(\mu_t | F_t)dt$ represents the best forecast of $\frac{dx_t}{x_t}$ and $\frac{ds_t}{s_t}$ conditional on the information set $F_t$, and $dW^F_t$ represents the standardized forecast errors. It is straightforward to see from the equation that the belief is more sensitive to the forecast errors if (i) the difference between the two possible scenarios,
(μh − μl), is greater; or (ii) the uncertainty about the state of the economy, captured by the conditional variance of the belief, πt(1 − πt), is higher. These results are quite intuitive. When the growth rates in the two states do not differ much, or when the firm is very sure that it is in one of these two states (i.e., πt is close to one or zero), unexpected changes in the signals do not have much impact on the belief.

An alternative formulation of the optimal updating rule, equation (A.1) in Appendix A.1, provides some further insights into the learning process. It suggests that the optimal learning from these two jointly normally distributed signals can proceed in two steps. One first forms a minimum variance “portfolio” of both signals, and then update the belief based on this compound signal. When the realized value of this compound signal is higher (lower) than expected, the posterior belief πt is adjusted upward (downward). The information quality of this compounded signal is measured by the inverse of its variance, \( \frac{1}{\sigma^2} = 1'(\Phi\Phi')^{-1}1 \). When this measure of quality is high, the firm’s belief responds to the compound signal more strongly. Furthermore, by the nature of the minimum variance portfolio, the optimal weighting vector \( w \) assigns more weight to the signal with a lower variance, indicating that the firm pays more attention to the less noisy signal.

The fact that \( W_{t|t}^F \) is a Wiener process with respect to the filtration \( F_t \) means that all information about the state of the economy available at time \( t \) is contained in the current belief \( π_t \). In other words, if there were some information about the future, this information is used immediately to update the current belief. Therefore, the belief \( π \) follows a \( F_t \)-Markov process. Alternatively, one can view \( dW_{t|t}^F \) as the perceived innovation of the informative signals, i.e., the disturbance that drives the deviation from the expected drift.

The formulation of the dynamics of \( x_t \) and \( s_t \) in equations (8) and (6) is not suitable for further analysis because the drift and innovation terms are not measurable with respect to \( F_t \). A reformulation of these dynamics in terms of unexpected changes with respect to \( F_t \) solves this problem. Using \( dW_{x|t}^F \) and \( dW_{s|t}^F \) defined by equation (10), we can rewrite the joint dynamics of \( x_t \) and \( s_t \) as

\[
\begin{pmatrix}
\frac{dx_t}{x_t} \\
\frac{ds_t}{s_t}
\end{pmatrix}
= \begin{pmatrix}
π_tμ_h + (1 - π_t)μ_l
\end{pmatrix}
\begin{pmatrix}
1 \\
1
\end{pmatrix}
dt + \Phi
\begin{pmatrix}
\frac{dW_{x|t}^F}{dW_{s|t}^F}
\end{pmatrix},
\]

(12)
where \( \pi_t \) is updated as stated in equation (9).

2.3 Firm Value Dynamics

Since there is no fixed adjustment cost, and the marginal cost of capital is constant, the firm’s optimal investment policy can be characterized by two reflecting boundaries, which splits the state space into an investment region, a non-investment region and a disinvestment region. The firm remains inactive in the interior area of the non-investment region, and increases or decreases capital instantaneously by an infinitesimal amount \( dk \) whenever it hits the boundaries. In this section, we first derive the firm value dynamics in the non-investment region, and then specify the boundary conditions that characterize the optimal investment/disinvestment policy.

From the previous section we know that the Bayesian belief \( \pi_t \) follows a \( \mathcal{F}_t \)-Markov process. Thus, starting from a certain prior, all information about future demand growth is incorporated in the current belief \( \pi_t \). Therefore, the firm’s value is fully determined by the current capital stock \( k_t \), the current demand factor \( x_t \), and the current belief \( \pi_t \), and the value function \( V \) can be written as \( V(k_t, x_t, \pi_t) \). This value function represents the present value of operating profit flow under the optimal investment policy.

Let \( r \) denote the instantaneous riskless rate of interest. We have the following proposition about the dynamics of the firm value \( V \):

**Proposition 2.** The firm value can be written as \( V(k, x, \pi) = xv(h, \pi) \), where \( h \equiv \frac{k}{x} \) and \( v \equiv \frac{V}{x} \) represent the capital stock and firm value, respectively, normalized by the demand factor. Furthermore, in the non-investment region, the normalized firm value, \( v \), has to satisfy the partial differential equation (Hamilton-Jacobi-Bellman equation):

\[
[r + \lambda_q - (\pi \mu_h + (1 - \pi) \mu_l)]v = \frac{h^{1-\alpha}}{1-\alpha} - h[\pi \mu_h + (1 - \pi) \mu_l + \xi] \frac{\partial v}{\partial h} + \frac{1}{2} \sigma^2 h^2 \frac{\partial^2 v}{\partial h^2} + [\pi \lambda_{hl} + (1 - \pi) \lambda_{lh} + \pi(1 - \pi)(\mu_h - \mu_l)] \frac{\partial v}{\partial \pi} + \frac{\pi(1 - \pi)(\mu_h - \mu_l)}{2 \sigma^2} \frac{\partial^2 v}{\partial \pi^2} - h \pi (1 - \pi)(\mu_h - \mu_l) \frac{\partial^2 v}{\partial h \partial \pi},
\]  

(13)
with
\[\sigma^2 = \frac{1}{\Phi'(\Phi\Phi')^{-1}}1^1,\]
as defined in A.1.

Proof. A proof is provided in Appendix A.2.

From Proposition 2 it follows that the normalized firm value, \(v\), depends only on the ratio of installed capital to the demand factor and the belief about the state of the economy.\(^{13}\)

### 2.4 Boundary Conditions

The partial differential equation (13) has to be solved under proper boundary conditions. We now specify these conditions.

As the marginal value of capital is decreasing in installed capital and increasing in the demand factor, the inaction region where the firm neither invests nor disinvests is associated with intermediate values of \(h = k/x\). When installed capital stock is low relative to demand, the marginal value of installed capital increases. Therefore, at a certain lower boundary of \(h\), the firm will have an incentive to invest. This critical threshold \(h^i\) forms the investment boundary. It specifies the optimal capital stock relative to the demand factor, and therefore can be interpreted as the optimal normalized capacity. At this boundary, every positive demand shock (which represents a negative shock to \(h\)) is offset by an appropriate increase in capital. As a result, the firm never enters the interior area of the investment region. That is why this threshold is called a reflecting boundary. This boundary is a function of the belief \(\pi\), because the marginal value of invested capital depends on growth prospects for the firm's operating profit.\(^{14}\)

---

\(^{13}\)This homogeneity property simplifies the solution of the valuation equation because it reduces the dimensionality of the problem. It follows from the homogeneity property of the operating profit \(f(x_t, k_t) = xf(1, k_t/x_t))\), and the absence of fixed costs of investment. Note that the homogeneity property also allows us to solve the problem in terms of Tobin’s average \(Q\), which can be written as \(Q(x_t/k_t, \pi_t) = V(k_t, x_t, \pi_t)/k_t\) and has to satisfy a partial differential equation similar to (13) in \(x_t/k_t\) and \(\pi_t\).

\(^{14}\)If investment is perfectly reversible, and the firm can adjust capital stock instantaneously to any level that is appropriate for the realized demand factor without adjustment cost, the firm’s optimal capacity will depend only on the current value of the demand factor, not on the belief about its future growth rate, and the capital stock will always be at the optimal level.
When installed capital is high relative to current demand and, thus, the marginal value of capital is low, the firm has an incentive to sell capital at a price $b$ per unit. Due to persistence of demand, over-capacities tend to persist. So at some upper threshold $h^*_d$ firms find it optimal to incrementally divest excess capital at a discount of $1 - b$. At $h^*_d$ (which is again a function of the belief $\pi$) every negative shock in demand will be rebalanced by proper disinvestment such that the region above $h^*_d(\pi)$ will never be entered.

Since the marginal cost of capital is normalized to be one, a rational valuation of the firm implies that at the investment boundary, the firm value has to satisfy the value matching condition:

$$V(x, k, \pi) = V(x, k + dk, \pi) - dk.$$ 

At the disinvestment threshold the respective boundary condition is

$$V(x, k, \pi) = V(x, k - dk, \pi) + b dk.$$ 

These conditions can be written in derivative form as

$$\lim_{k \rightarrow x_{h^*_i}} \frac{\partial V}{\partial k} = 1, \quad \lim_{k \rightarrow x_{h^*_d}} \frac{\partial V}{\partial k} = b.$$ 

This condition follows from the fact that firm value today fully reflects future investment/disinvestment activity. It characterizes an important feature of the boundaries: at the investment boundary, the marginal value of capital, i.e., Tobin’s marginal $Q$, is always equal to the marginal cost of adding capital, which is constant in our model. Analogously, at the disinvestment boundary, Tobin’s marginal $Q$ equals $b$. Using the homogeneity feature of the value function, we can then rewrite these boundary conditions as

$$\lim_{h \rightarrow h^*_i} \frac{\partial v(h, \pi)}{\partial h} = 1, \quad \lim_{h \rightarrow h^*_d} \frac{\partial v(h, \pi)}{\partial h} = b.$$ 

(14)

To ensure the optimality of the endogenously determined boundary, we also require smoothness of the marginal value of capital at the boundaries. This implies the following super contact (or smooth pasting) conditions (see Dumas (2001) at the investment and
disinvestment boundaries:
\[
\frac{\partial^2 V}{\partial k \partial x} = 0, \quad \frac{\partial^2 V}{\partial k^2} = 0, \quad \frac{\partial^2 V}{\partial k \partial \pi} = 0.
\]

These translate into the boundary conditions below for \(v(h, \pi)\), both at the investment as well as at the disinvestment boundary:
\[
\lim_{h \to h^*} \frac{\partial^2 v(h, \pi)}{\partial h^2} = 0, \quad \lim_{h \to h^*} \frac{\partial^2 v(h, \pi)}{\partial h \partial \pi} = 0.
\] (15)

3 Calibration

Our model does not have an analytical solution, so we solve it numerically. We first describe our procedure to calibrate the model parameters, and then discuss the main properties of the numerically derived optimal investment policy.

There are eleven parameters in our model: \(\alpha\) in the operating profit function; regime switching intensities \(\lambda_{h,l}\) and \(\lambda_{l,h}\); expected demand growth rates in the good and bad states \(\mu_h\) and \(\mu_l\); instantaneous volatilities of demand and of the public signal, \(\sigma_x\) and \(\sigma_s\); depreciation rate \(\xi\); risk-free rate \(r\); capital resale price \(b\), and correlation between the Wiener processes driving the demand factor and the public signal, \(\rho\).

We first estimate the business cycle parameters, \(\mu_h, \mu_l, \lambda_{h,l},\) and \(\lambda_{l,h}\), using the average annual growth rate of the demand factors of firms in the Compustat-CRSP merged database over 1950-2011. Note that from equation (5) we can back out a firm’s demand factor \(x_t\) using its operating profit \(f(x_t, k_t)\) and capital stock \(k_t\):
\[
x_t = \left[ \frac{(1-\alpha)f(x_t, k_t)}{k_t^{1-\alpha}} \right]^{1/\alpha}.
\] (16)

Using this formula, we compute \(x_t\) for each individual firm in each year. We measure \(f(x_t, k_t)\) by operating income before depreciation (OIBDP), and \(k_t\) by operating assets (defined as total asset (AT) minus cash and short-term investment (CHE)). To account for inflation, we convert the values of both operating income and operating assets in each year into year 2005 dollars using the annual GDP deflator obtained from the U.S. Bureau of Economic Analysis (BEA). We exclude firm/year observations where the operating asset
value is less than $5 million (in year 2005 dollars). We set the operating profit parameter \( \alpha \) equal to 0.74, following an argument by Guo, Miao, and Morellec (2005).\(^{15}\)

After estimating the \( x_t \) series for each firm, we calculate its continuously compounded annual growth rate. We calculate the average growth rate in each year after excluding observations below the 1st or above the 99th quantile of the sample. This gives us a time series of 60 annual observations. We apply the Expectation-Maximization (EM) algorithm developed by Dempster, Laird, and Rubin (1977) to estimate the parameters of the hidden Markov chain: \( \lambda_{h,l}, \lambda_{l,h}, \mu_h, \mu_l \). The results are summarized in Table 1 together with the values of other parameters.

Our estimates of the transition intensities imply an expected length of 3.99 years (=1/0.2504) for an expansion and 1.58 years (=1/0.6345) for a recession. This is well in line with the estimates by Hamilton (1989) using quarterly GDP data. According to his estimation, expansions last 10.5 quarters on average and recessions 4.1 quarters. We plot the time series of estimated demand growth rate in Panel (A) and the posterior beliefs \( \pi_t \) in Panel (B) of Figure 3 along with historical expansion and recession periods as identified by the National Bureau of Economic Research (NBER). One can see that the \( \pi_t \) series matches the NBER-designated business cycles very well. Recession periods are always associated with \( \pi_t \) less than 0.5, while expansion periods are associated with \( \pi_t \) higher than 0.5 except for the year 1952. Our results also indicate that expected growth rates in the good and bad states are quite different: 0.0938 vs. -0.0629. This highlights the importance of the business cycle for firms’ profitability and investment decisions.

The other parameter values for the base case are estimated or chosen as follows. The depreciation rate \( \xi \) is set to be 0.1211. This is also estimated using the annual Compustat-CRSP merged database. We divide the Depreciation of Tangible Fixed Assets (DFXA) by the PPENT at the end of the previous year, and take the median value of the sample as our estimate. The instantaneous volatility of the demand factor, \( \sigma_x \), is set to be 0.3036.

\(^{15}\)As explained in footnote 7 of Guo, Miao, and Morellec (2005), the operating profit function (5) approximates the following specification: (1) Constant returns to scale Cobb-Douglas production function with labor and capital: \( q = \lambda L^\phi K^{1-\phi} \); (2) Isoelastic demand function given by the inverse demand curve: \( p = x^{1-\theta} q^{\theta-1} \), where \( 0 < \theta < 1 \). It follows from this specification that the share of profits going to capital depends on \( \theta \) and \( \phi \) through the following relation: \( 1 - \alpha = (1 - \phi)\theta/(1 - \theta\phi) \). Labor’s share of national income in U.S. since world war II is relatively stable at \( \phi = 0.64 \). Assuming \( \theta = 0.5 \), we get \( \alpha \approx 0.74 \). Guo, Miao, and Morellec (2005) obtain \( \alpha \approx 0.53 \) due to a typo in their expression for \( 1 - \alpha \).
Panel (a) shows the average annual growth rate (continuously-compounded) of the demand factor \( x \), \( \Delta \ln(x) \), across firms in the Compustat-CRSP merged database. Panel (b) shows the time series of the posterior belief \( \pi \) estimated from the time series of \( \Delta \ln(x) \) using the EM algorithm. The narrow bars correspond to the recession periods dated by the NBER.

Figure 3: Posterior beliefs of being in an expansion
Table 1: Parameter values

This table summarizes the parameter values for the base case scenario. Parameters with four digits after the decimal point are estimated using the annual Compustat-CRSP merged database over 1950-2011.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Economic meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>operating profit parameter</td>
<td>0.74</td>
</tr>
<tr>
<td>$\mu_h$</td>
<td>expected growth rate in good state</td>
<td>0.0938</td>
</tr>
<tr>
<td>$\mu_l$</td>
<td>expected growth rate in bad state</td>
<td>-0.0629</td>
</tr>
<tr>
<td>$\lambda_{h,l}$</td>
<td>transition intensity from good state to bad state</td>
<td>0.2504</td>
</tr>
<tr>
<td>$\lambda_{l,h}$</td>
<td>transition intensity from bad state to good state</td>
<td>0.6345</td>
</tr>
<tr>
<td>$\xi$</td>
<td>depreciation rate</td>
<td>0.1211</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>instantaneous volatility of demand factor $x_t$</td>
<td>0.3036</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>instantaneous volatility of signal $s_t$</td>
<td>0.25</td>
</tr>
<tr>
<td>$b$</td>
<td>resale price of one unit of capital</td>
<td>0.80</td>
</tr>
<tr>
<td>$r$</td>
<td>risk-free rate</td>
<td>0.05</td>
</tr>
<tr>
<td>$\rho$</td>
<td>instantaneous correlation between $dW_{st}$ and $dW_{xt}$</td>
<td>0.05</td>
</tr>
</tbody>
</table>

We regress the annual demand growth rate on an NBER recession dummy firm by firm, and take a simple average of the root mean square error of each regression to obtain this estimate. The risk-free rate is 0.05. The resale price of capital, $b$, is 0.80. The instantaneous volatility, $\sigma_s$, of the public signal is 0.25, while the instantaneous correlation, $\rho$, between the public signal and the firm’s demand factor is 0.05.

4 Optimal Investment Policy

We now derive investment and disinvestment boundaries $h^*_i(\pi)$ and $h^*_d(\pi)$ at which the firm finds it optimal to add or reduce capacity. To solve the Hamilton-Jacobi-Bellman equation (13) along with the boundary conditions (14) and (15) numerically, we apply the approach derived in Nelson and Ramaswamy (1990) to map the dynamics of the belief $\pi$ onto a recombining tree, and then use a two-dimensional tree (as outlined in Boyle, Evnine, and Gibbs (1989)) to jointly determine the firm value and the investment boundary. A detailed description of the numerical procedure is provided in Appendix A.3.

Panel (b) of Figure 4 plots the optimal investment boundary as a function of the current belief $\pi$. The solid upward-sloping curve represents the boundary for the baseline.

---

16 We use firms with at least 30 annual observations for this estimation. We do not apply the EM algorithm to individual firms because the demand factors of individual firms are too noisy.
This figure depicts the optimal investment boundary (Panel (b)) and disinvestment boundary (Panel (b)) as a function of the belief $\pi$ for different levels of information quality. The solid curve depicts the investment boundary for the base case ($\sigma_s = 25\%$). The two dashed curves in each panel are boundaries under higher information quality ($\sigma_s$ equal to 1% and 0.5%, respectively). All other parameter values are fixed at the base case level specified in Table 1.

Figure 4: Optimal investment and disinvestment boundaries
parameterization in Table 1. It is a lower bound for the firm’s normalized capital stock. The area above this boundary is the non-investment region, in which the firm remains inactive due to excess capacity. The firm increases its capital whenever $k/x$ hits the boundary from above (as capital depreciates or demand increases) or from the left (as $\pi$ increases). Not surprisingly, when the firm believes that the economy is in an expansion ($\pi \to 1$), it invests earlier, i.e., at a higher boundary, than when it believes the economy is in a recession ($\pi \to 0$).

A notable feature of the optimal investment boundary is that it is convex in the belief $\pi$, indicating that the firm’s investment decision is relatively insensitive to changes in the belief when $\pi$ is low. This results from the interaction of the expected growth rate and uncertainty about the growth rate. Starting from $\pi$ close to zero, i.e., when the firm is almost sure of being in the low growth state, a positive signal increases the expected growth rate of future demand. At the same time, it also increases uncertainty about the current state (captured by a higher value of conditional variance, $\pi(1-\pi)$), thus increasing the option value of waiting. As a result, the firm is reluctant to invest. If, however, a negative signal is received in a high growth state (i.e., when $\pi$ is high), both the resulting lower expected growth rate and greater uncertainty diminish the firm’s incentive to invest. Therefore the investment boundary drops sharply.

The dashed curves in Figure 4 illustrate investment boundaries for better information quality, i.e., lower $\sigma_s$ of the public signal $s_t$. They indicate that higher information quality makes the boundary more convex. Other things equal, the more precise the information, the stronger the response of the belief to signals, as we have noted in Section 2.2. This leads to more volatile beliefs, which in turn generate higher incentives to wait. Higher information quality thus amplifies the convexity of the boundary. For $\sigma_s$ close to zero, the option value of waiting is so high that the firm always behaves as if it were in a recession except when it is almost sure of being in an expansion. As a result, the investment boundary is almost flat for $\pi$ below one and increase sharply as $\pi$ approaches one.

Panel (a) of Figure 4 shows the disinvestment boundaries for different information quality $\sigma_s$. When a firm experiences negative demand shocks such that invested capital is high relative to demand, demand persistence creates an incentive for the firm to disinvest capital even at a discount $b$. The disinvestment boundary characterizes the maximum toler-
ated normalized capital, $k/x$. Thus, if a firm’s capital is hits this boundary, disinvestment takes place, which pushes $k/x$ back towards the non-investment region. This boundary therefore constitutes another reflecting barrier to the firm’s normalized capital. The range below this line is the inaction range where the firm experiences excess capacity but stays inactive, i.e., does not actively adjust invested capital downwards.

The disinvestment boundaries are concave, and the concavity becomes more pronounced as the information quality increases. The reason for this is again closely related to the belief volatility, which is increasing with increasing information quality. When information quality of the signal is high, the option value of waiting is high, and the disinvestment boundary is almost flat for high values of $\pi$. The firm only starts to sell capital aggressively (i.e., at relatively low values of $k/x$) when $\pi$ is close to zero, i.e., when it is almost sure that the current state is the low-growth state.

5 Dynamics of Capital Growth Rates: Simulated versus Empirical Data

While the shape of the investment boundary determines how sensitive a firm’s optimal capacity is to changes in its belief, it alone does not determine the speed of capital adjustment in response to a regime shift, because the actual adjustment speed also depends on how quickly the firm learns about the true state of the economy, i.e., how quickly $\pi$ converges to zero or one. When the information is more precise, the investment boundary is more convex, but the firm also learns faster, therefore its capital increase after a favorable regime shift is not necessarily slow. Furthermore, at the aggregate level, the speed of capital adjustment also depends crucially on the distribution of firms’ normalized capacities relative to their investment boundaries. For example, when a large fraction of firms are farther away from their boundaries, the aggregate response to signals is weak. This distribution is endogenously determined by the history of demand shocks and firms’ investment activities.

To examine the dynamics of the capital growth rate in our model, we use Monte Carlo simulation. We simulate a sample of 1000 firms over a time horizon of 100 years.\(^{17}\) The

\(^{17}\)We drop the first 20 years of this generated data set so that the investment patterns are independent of initial conditions. Increasing the number of firms in our simulation leads to very similar results.
This graph shows the simulated and the empirically estimated equal-weighted average capital growth rate over a full business cycle. The simulated firms are assumed to follow the optimal investment policy derived numerically in Section 4. Capital growth rates are first averaged across firms and then averaged across simulated business cycles over an 80-year period. The black and red dashed lines represent the simulated average quarterly capital growth rate with $\sigma_s = 25\%$ and $\sigma_s = 0.5\%$, respectively. Other parameter values are given in Table 1. The solid line represents the equal-weighted quarterly real capital growth rate around the business cycle troughs dated by the NBER, calculated using the Compustat-CRSP merged database from 1975 through 2011.

common macroeconomic regimes follow the Markov process (7). Each firm observes its own demand factor $x_t$ and a common public signal process $s_t$, and invest according to the optimal policy derived in Section 4. The demand factors are correlated across firms due to their correlations with the common signal. The parameter values are taken from Table 1 unless otherwise noted. We analyze the capital growth rates of individual firms as well as the average capital growth rate across firms, and compare them to the empirical data.
5.1 Average Capital Growth Rate Over the Business Cycle

Figure 5 shows the equal-weighted average capital growth rates over a full business cycle (starting with a recession). The solid line is empirical average capital growth rate reproduced from Panel (a) of Figure I. The black dashed line is the simulated average growth rate under the base case parameterization specified in Table I (with $\sigma_s = 25\%$). The red dashed line is the simulated average growth rate under almost complete information ($\sigma_s = 0.5\%$), with other parameter values unchanged. For the empirical curve, time zero is the troughs of business cycles as dated by the NBER. For the two simulated curves, it is the beginning of an expansion period.

Like the empirical capital growth rate, our simulated capital growth rate under the baseline parameterization ($\sigma_s = 25\%$) features a sharp decline and a slow recovery, although the recovery occurs sooner after time zero. The simulation under almost complete information ($\sigma_s = 0.5\%$) shows a more dramatic decline at the beginning of a recession. At the same time, it also exhibits an almost equally sharp and immediate rebound once the recession is over. When informations is (almost) complete, firms jump immediately from one end of the investment boundary to the other. The sudden change of the optimal capacity leads to a dramatic change in the average capital growth rate at both regime switching points. These dramatic changes do not match well the empirical data.

The main reason for the asymmetry of decline and recovery in our baseline model is the endogenous distribution of firms relative to their optimal capacities. Since optimal capacity is higher, and demand grows faster in an expansion, many firms are pushed to the investment boundary at the end of an expansion. These are the marginal firms that react to changes in beliefs. When a negative signal comes, they stop investment, generating a sharp decline in the average capital growth rate. By contrast, at the end of a recession, many firms are far away from the investment boundary because they have experienced lower demand growth. These firms may not invest even if $\pi$ goes from zero to one, as their capital stock is too high even compared to the boundary for an expansion period.

---

18Note that the capital growth rate in this case declines slightly after its initial rebound at the beginning of the expansion. This is because the initial rebound reflects a discrete jump to a higher optimal capacity by many firms, while the subsequent capital growth is only driven by the continuous demand growth that pushes firms toward a given boundary.

19They are not close to the disinvestment boundary either, because depreciation tends to reduce their net capital.
Consequently, when a positive signal arrives, the number of firms responding is small.

To illustrate this point more clearly, we present histograms of simulated normalized capital ($k/x$) at two turning points of a business cycle in Figure 6. Panel A shows the distribution at the end of an expansion, and Panel B shows the distribution at the end of a recession. One can see clearly that in Panel A firms are more concentrated along the investment boundary, while in Panel B they are much more dispersed, indicating a larger fraction of firms with excess capacity.

The convex relation between the optimal capacity and the belief of being in an expansion also tends to slow down the recovery as well. During an expansion, in which $\pi$ is close to one, a decrease in $\pi$ triggered by a negative signal leads to a sharp decline in a firm’s optimal capacity, as it not only lowers the expected demand growth rate, but also increases the option value of waiting. By contrast, when a positive signal arrives during a recession, its positive impact on the expected growth rate is partially offset by the higher option value of waiting, leaving optimal capacity largely unaffected. However, this effect is relatively small for two reasons. First, the shape of the boundary only affects the firms that are close to the boundary. Second, as we discuss in Section 4, the convexity is more pronounced when the signals are precise, but when signals are sufficiently precise, uncertainty does not have a strong effect because it is short-lived.

The simulation results presented in Figure 1 suggest that our model generates to a large extent the asymmetric average capital growth rate observed in the data. Nevertheless, the recession in our simulated data is deeper than its empirical counterpart, and the recovery appears to be faster. These discrepancies may arise because we assume perfectly synchronized regime shifts across firms. In reality firms do not enter a certain growth state simultaneously, thus the average growth rate is smoother. Another potential reason is that our model abstracts from any frictions in the financial market. In the real world, when the economy just recovers from a recession, it is usually difficult for firms to raise capital. Such financial constraints further reduce firms’ capital growth rates at the initial stages of

\[\text{Note that there are some firms located to the left of the peak of the histogram in both panels. The optimal capacities of these firms are lower than those of others because they have a more pessimistic view of the economy (i.e., a lower } \pi). \text{ The empirical distribution of normalized capital in our sample also exhibits a higher dispersion at the end of recessions, although its lower bound is not as clear-cut as that of the simulated data, potentially due to the heterogeneity of firms in the sample.}\]
This figure shows the histograms of simulated normalized capital \((k/x)\) at two turning points of a business cycle: at the end of an expansion and at the end of a recession. The parameter values used for the simulation are given in Table 1.
5.2 Skewness of Capital Growth Rates

To further gauge the empirical plausibility of our model, we examine the asymmetry of capital growth rates, both at the individual firm level and in the aggregate.

Panels (a) and (b) of Figure 7 show the skewnesses of firm-level capital growth rates and their first-order differences, respectively. Panels (c) and (d) show the skewnesses of average capital growth rates and their first-order differences, respectively. The length of the time interval over which the growth rate is measured varies from 1 to 10 quarters. The black curves are the empirical estimates of the skewnesses and their 95% confidence intervals reproduced from Figure 2. The red curves are the skewnesses of the simulated data and their confidence intervals.

As one can see from the figure, our model replicates the contrasting features of capital growth rates at the firm and aggregate levels. At the firm level, capital growth rates exhibit a strong level asymmetry with a positive skewness (Panel (a)), and virtually no slope asymmetry (Panel (b)). For both empirical and simulated firm-level capital growth rates, the positive skewness is highest at the one-quarter interval, and decreases steadily as the time interval lengthens, but remains positive even at the 10-quarter interval. This positive skewness is a natural outcome of costly reversibility, which makes firms reluctant to disinvest. Capital is built up for specific uses, and it may be very difficult to use it for other purposes. Thus firms may be able to expand their capital stock dramatically, but unable to reduce it as desired. The almost negligible slope asymmetry of firm-level capital growth rates in both empirical and simulated data suggests that for an individual firm, cutting back investment (instead of capital) is as fast as increasing it. A notable discrepancy between the simulated and the empirically estimated skewnesses in Panel (a) is that the simulated skewnesses appear to be too high. This is because we allow firms to add capital instantaneously with zero adjustment costs. In the real world, firms have to adjust costs for both investment and disinvestment, even though the costs for disinvestment are

\[ \text{confidence intervals for the simulated data, we repeat our simulations of 1000 firms 42 times. The point estimate of each skewness is the simple average across the 42 sets of simulations, and the standard error of the estimate is the standard deviation divided by } \sqrt{42}. \]
Panels (a) and (b) show the skewnesses of firm-level capital growth rates and their first-order differences, respectively. Panels (c) and (d) show the skewnesses of average capital growth rates and their first-order differences, respectively. The length of the time interval over which the growth rate is measured varies from 1 to 10 quarters. The black curves show the empirical estimates based on real capital growth rates of firms in the quarterly Compustat-CRSP merged database over 1975-2011. The red curves show results estimated from simulated data. The dashed curves indicate the 95% confidence interval. The parameter values used for the simulation are given in Table 1.
likely to be larger. Introducing some adjustment costs for capital expansion will make the positive skewness at the firm level less pronounced.

At the aggregate level, our model generates both the level asymmetry (Panel (c)) and slope asymmetry (Panel (d)) observed in the data. Both the average capital growth rates and their first-order differences are negatively skewed. Furthermore, both skewness coefficients are non-monotonic in the length of time interval, reaching their minimum values when growth rates are measured over an interval of 3 or 4 quarters, and slowly increases towards zero as the time interval further lengthens. Even though the magnitudes of the skewness coefficients are in general lower in the simulated data than in the real data, these results suggest that our model captures remarkably well the main features of the average capital growth rates across firms.

The level asymmetry of average capital growth rates has to do with the fact that expansions usually last longer than recessions do, which implies that the bulk of short-run growth rate observations are from the expansion period and therefore lie to the right of the mean. The relatively small number of observations from the recession period then form a long tail at the left side of the distribution, resulting in negative skewness. The slope asymmetry of the average capital growth rate reflects the sharp-decline-slow-recovery feature of the business cycle. As we discuss in section 5.1, this occurs mainly due to the endogenous distribution of firms relative to their optimal capacities over the business cycles. More firms are close to the investment boundary during an expansion than during a recession; therefore, more firms react when a negative signal arrives during an expansion than a positive signal arrives in a recession.

Figure 7 suggests that our model is able to reconcile two seemingly conflicting patterns of capital growth rates: negative skewness of the average capital growth rate, both in levels and in first-differences, and positive skewness at the firm level.

6 Effects of Depreciation and Reversibility

In this section we study the consequences of changes in depreciation and reversibility on investment and disinvestment boundaries, on the skewness of individual firms’ capital growth rates as well as on the skewness of first differences of the average capital growth. On
These figures compare the investment behavior of simulated firms under different degrees of investment reversibility. The resale price of capital, $b$, is 0.98 in the high reversibility case, 0.5 in the low reversibility case, and 0.8 in the base case. Other parameter values are given in Table 1.

On the one hand, this analysis serves as a robustness check about whether our base case results are qualitatively robust with respect to changes in the parametrization of the model. On the other hand, these comparative static results produce predictions about the dependency of investment behavior on certain firm characteristics.

### 6.1 Differences in the Degree of Reversibility

The degree of reversibility of investments is determined by the gap between the purchase price and the resale price of capital. Since we fix the purchase price of capital at 1 throughout our study, the specification of the resale price, $b$, characterizes reversibility, with values...
of $b$ close to 1 defining high reversibility and values of $b$ towards zero defining irreversibility.

Panels (a) and (b) of Figure 8 show the effects of investment reversibility on maximum tolerated excess capital and on the optimally invested capital, respectively. Solid curves correspond to the base case with $b = 0.80$. The high and low reversibility cases are represented by $b = 0.98$ and $b = 0.50$, respectively. Not surprisingly, with high reversibility firms are willing to invest and disinvest early, leading to a narrow inaction range compared to the case of low reversibility. Therefore, an increase in $b$ leads to a downward shift of the disinvestment boundary and an upward shift of the investment boundary. Furthermore, the right end of the disinvestment boundary moves more than the left and the left end of the investment boundary moves up more than the right end, suggesting lower sensitivity of the investment/disinvestment decision on the state of the economy. As $b$ approaches 1, investment becomes perfectly reversible, investment and disinvestment boundary will collapse into one and be perfectly flat (since optimal capital can costlessly be adjusted instantaneously to current demand).

With increased level of reversibility, firms’ disinvest capital actively to reduce excess capital in bad states. As a consequence, investment and disinvestment behavior become more and more symmetric. Thus, we see in Panel (c) of Figure 8 that average skewness of individual firms’ capital growth rate is reduced if reversibility increases with increasing resale price of capital.

On the aggregate level, two mutually enforcing mechanisms are at work when reversibility of investment increases. First, the range of inactivity between investment boundary and disinvestment boundary becomes smaller, since firms are not willing to hold excess capital but rather disinvest it. Hence, firm’s heterogeneity with respect to excess capital is reduced and if the regime changes from recession to expansion, more firms move quickly towards the investment boundary. Second, the investment boundary becomes less steep and less convex (illustrated in panels (a) and (b) of Figure 8), as learning about the current state of the economy is less important if disinvestment can be done with only low losses. A flatter investment boundary, however, means that investment activity is less sensitive to the belief about the growth state. So we predict that among firms with high degree of reversibility, the slope asymmetry is less pronounced. See panel (d) of Figure 8 for an illustration.

Figure 8 also show another interesting feature. In all the four panels, while the gap
These figures compare the investment behavior of simulated firms under different depreciation rates. The depreciation rate is 6% per annum in the low depreciation case, 20% per annum in the high depreciation case, and 12.11% per annum in the base case. Other parameter values are given in Table 1.

between the $b = 0.98$ and $b = 0.80$ cases are rather big, the difference between the $b = 0.80$ and $b = 0.50$ cases are fairly small. This suggests a small cost of reversibility is enough to capture the majority of the effects of irreversibility.

### 6.2 Differences in Depreciation Rate

The impact of high depreciation on investment decisions is very similar to the effect of a high discount rate. The maximum tolerated excess capital and the optimally invested capital decrease, together with the sensitivity of investment/disinvestment decisions on information about the state of the economy (see panels (a) and (b) of Figure 9). This is so
because the effective planning horizon of decision makers decrease. A higher depreciation rate increases firms’ ability to adapt to negative demand shocks since it reduces excess capital at a faster speed. Consequently, our model predicts that higher depreciation reduces the positive skewness of individual investment, since it makes the firms’ investment and disinvestment more symmetric. See panel (c) of Figure 9 for an illustration.

The reduction of excess capital due to depreciation prevents firms from moving far-away from the investment boundary. So an increase in expected demand growth quickly translates into new investment by a broad range of firms. Hence, when firms have a higher depreciation rate, our model predicts that the average capital growth rate shows less slope asymmetry, compared to the case with a low depreciation rate. See panel (d) of Figure 9 for an illustration of this effect.

7 Conclusion

We develop a model of costly reversible incremental investment over the business cycle when there is uncertainty about the true state of the economy. We consider a cross-section of firms facing stochastic demand shocks. The expected growth rate of the demand factors depends on the state of the economy, which shifts between expansion and recession at random times. Because disinvestment is costly, a firm’s investment decision depends not only on the current demand factor but also on the belief about the current state of the economy. We show that a firm’s optimal investment threshold, defined in terms of a lower bound on the firm’s capital normalized by a demand factor, is a convex function of its posterior probability of being in an expansion. We then examine the dynamics of capital growth rate by simulating a large panel of firms following the same investment policy but facing heterogeneous demand shocks.

Our model replicates important features of the capital growth rate observed in the quarterly Compustat-CRSP merged database. Despite the strong positive skewness of capital growth rates at the firm level, the average capital growth rate across firms, as well as its first-order difference, are negatively skewed. The positive skewness of an individual firm’s capital growth rate is a natural outcome of investment costly reversibility, which limits the speed of capital shrinkage. The negative spike of the average capital growth rate
is due to the fact that the expansion normally lasts longer than the recession, which implies that the bulk of observations lie to the right of the mean. The sharp-decline-slow-recovery feature of the average capital growth rate is mainly due to the endogenous distribution of firms relative to the boundary. Since there are more firms at the investment boundary at the end of an expansion then at the end of recession, more firms react the negative signals coming in good times than to positive signals in bad times. As a result, the average capital growth rate across firms increases gradually during an expansion but drops sharply at the beginning of a recession.

One feature of our model is that it abstracts from any frictions in the financial markets. In the real world, when the economy just emerges from a recession, it is usually difficult for firms to raise capital. Such financial constraints tend to further delay firms’ investment at the initial stage of an expansion. Incorporating financial frictions over business cycles into the firm’s investment decision is a fruitful venue for future research. Another feature of our model is that it abstracts from the feedback effects of firm investment on product and capital prices. Our partial equilibrium approach is similar to that of Bertola and Caballero (1994), Bloom, Bond, and Reenen (2007) and Bloom (2009), in that we first derive the optimal policy of an individual firm, and then examine the aggregate behavior of a cross-section of firms following the policy. Endogenizing the feedback effects in a fully-specified general equilibrium model can potentially generate rich dynamics of various economic variables, but is beyond the scope this paper, and thus left for future research.

Appendixes

A.1 An Alternative Formulation of the Optimal Updating Rule

Equation (9) can be rewritten as:

\[
d\pi_t = [-\pi_t \lambda_{h,l} + (1 - \pi_t) \lambda_{l,h}]dt + \frac{(\mu_h - \mu_l)\pi_t(1 - \pi_t)}{\sigma^2} \left( \begin{array}{c} d\pi_x \cr d\pi_s \end{array} \right) - E_t(\mu_t | \mathcal{F}_t)dt \right)
\]

(A.1)

where

\[
\sigma^2 = \frac{1}{1'(\Phi \Phi')^{-1}1},
\]

(A.2)
\[ w \equiv \frac{1'(\Phi\Phi')^{-1}}{1'(\Phi\Phi')^{-1}1} = \left( \frac{\sigma_s^2 - \rho \sigma_x \sigma_s}{\sigma_s^2 + \sigma_s^2 - 2 \rho \sigma_x \sigma_s}, \frac{\sigma_x^2 - \rho \sigma_x \sigma_s}{\sigma_x^2 + \sigma_x^2 - 2 \rho \sigma_x \sigma_s} \right). \tag{A.3} \]

Note that $\Phi\Phi'$ is simply the instantaneous variance-covariance matrix of $\frac{dx_t}{x_t}$ and $\frac{ds_t}{s_t}$. Readers familiar with the classic mean-variance portfolio analysis will immediately recognize that $\sigma^2$ is the minimum instantaneous variance that can be obtained using all possible linear combinations of $\frac{dx_t}{x_t}$ and $\frac{ds_t}{s_t}$, while $w$ is a vector that specifies the weights of each individual signal in the minimum variance combination. This formulation thus reveals an important feature of the Bayesian learning process. When there are multiple jointly normally distributed signals, the agent can form a minimum variance “portfolio” of all the available signals, and base learning on this compound signal. When the realized value of this compound signal is higher than expected, the posterior belief $\pi_t$ is adjusted upward. Conversely, when the realized value is lower than expected, $\pi_t$ is adjusted downward.

The standard deviation of this optimally constructed compound signal, $\sigma$, measures the noisiness of the overall information of all the signals. The learning equation (9) thus implies that the response to forecasting errors is more pronounced when signals are more precise. Furthermore, by the nature of the minimum variance portfolio, the optimal weighting vector $w$ assigns more weight to the signal with lower variance, indicating that agents pays more attention to the signal that has less noise. In particular, when $\sigma_s = \rho \sigma_x < \sigma_x$, the optimal weight of $\frac{dx_t}{x_t}$ in the compound signal is zero. Learning is entirely based on the signal $\frac{ds_t}{s_t}$. More surprisingly, when $\sigma_s < \rho \sigma_x$, the optimal weight of $\frac{dx_t}{x_t}$ is even negative. This implies that when the realized value of the more precise external signal $s_t$ is just as expected, while the firm’s own demand factor $x_t$ is higher than expected, agents will adjust their belief $\pi_t$ downward. The intuition is as follows. Since the shocks to the two signals are highly correlated, a higher-than-expected realized value of $\frac{dx_t}{x_t}$ suggests that it is very likely that $\frac{ds_t}{s_t}$ has also received a positive shock; i.e., it is above its true mean. This therefore suggests that the current belief of the mean, which is right at the realized value of $\frac{dx_t}{x_t}$, is too high and should be revised downward.

**A.2 Proof of Proposition 2**

Consider firm value $V$ as a claim on the firm’s operating profit as a function of its current demand factor $x$, the installed capital $k$, and the belief about the current state of the
economy $\pi$, i.e., $V = V(x,k,\pi)$. For the risk-neutral decision maker, the value function must satisfy the following Hamilton-Jacobi-Bellman equation:

$$rV(x,k,\pi)dt = f(x,k)dt + E(dV(x,k,\pi)_{F_t})$$

The expectation of $dV$ is to be determined by Itô’s lemma using the $F_t$-dynamics of the state variables $x$, $k$, and $\pi$. Note that since $dW^{F}_{xt}$ and $dW^{\mathbb{F}}_{st}$ are uncorrelated, we have

$$\begin{align*}
(d\pi)^2 &= d\pi(d\pi)' \\
&= [\pi(1 - \pi)(\mu_h - \mu_l)]^2 \left[1'(\Phi')^{-1}dW^{F}_t\right] \left[1'(\Phi')^{-1}dW^{\mathbb{F}}_t\right]' \\
&= [\pi(1 - \pi)(\mu_h - \mu_l)]^2 \frac{1}{\sigma^2}dt,
\end{align*}$$

$$\begin{align*}
dx \ d\pi &= dx \ (d\pi)' \\
&= x\pi(1 - \pi)(\mu_h - \mu_l) \left[(1,0) \Phi dW^{F}_t\right] \left[1'(\Phi')^{-1}dW^{\mathbb{F}}_t\right]' \\
&= x\pi(1 - \pi)(\mu_h - \mu_l)dt.
\end{align*}$$

Therefore, in the inactivity region, where $dk = -\xi kdt$, we have:

$$E[dV(x,k,\pi)|F_t] = \left[-\frac{\partial V}{\partial k}\xi k + x(\pi \mu_h + (1 - \pi)\mu_l) + \frac{1}{2}\frac{\partial^2 V}{\partial x^2}x^2 \sigma^2 + \frac{\partial V}{\partial \pi}(-\pi \lambda_{h,l} + (1 - \pi)\lambda_{l,h}) + \frac{1}{2}\frac{\partial^2 V}{\partial \pi^2}[\pi(1 - \pi)(\mu_h - \mu_l)]^2 \frac{1}{\sigma^2} + \frac{\partial^2 V}{\partial x \partial \pi}x\pi(1 - \pi)(\mu_h - \mu_l)\right] dt.$$

Substituting this expression into the Hamilton-Jacobi-Bellman equation and dropping $dt$ from both sides of the equation yields

$$rV = \frac{1}{1 - \alpha}x^\alpha k^{(1 - \alpha)} - \frac{\partial V}{\partial k}\xi k + x(\pi \mu_h + (1 - \pi)\mu_l) \frac{\partial V}{\partial x} + \frac{1}{2}x^2 \sigma^2 \frac{\partial^2 V}{\partial x^2} + \frac{(-\pi \lambda_{h,l} + (1 - \pi)\lambda_{l,h})}{2}\frac{\partial V}{\partial \pi} + \frac{[\pi(1 - \pi)(\mu_h - \mu_l)]^2}{2\sigma^2} \frac{\partial^2 V}{\partial \pi^2} + x\pi(1 - \pi)(\mu_h - \mu_l) \frac{\partial^2 V}{\partial x \partial \pi} \quad (A.4)$$

The last part of the proof to show that writing $V(x,k,\pi)$ as $V = xv(h,\pi)$ with $h = \frac{k}{x}$ gives equation (13). This is done by substituting the partial derivatives below into equation
\[ \frac{\partial V}{\partial k} = \frac{\partial v(h, \pi)}{\partial h}, \]
\[ \frac{\partial V}{\partial x} = v(h, \pi) - h \frac{\partial v(h, \pi)}{\partial h}, \]
\[ \frac{\partial^2 V}{\partial x^2} = \frac{1}{x} \frac{\partial^2 v(h, \pi)}{\partial h^2}, \]
\[ \frac{\partial V}{\partial \pi} = x \frac{\partial v(h, \pi)}{\partial \pi}, \]
\[ \frac{\partial^2 V}{\partial \pi^2} = \frac{\partial^2 v(h, \pi)}{\partial \pi^2}, \]
\[ \frac{\partial^2 V}{\partial x \partial \pi} = \frac{\partial v}{\partial \pi} - h \frac{\partial^2 v(h, \pi)}{\partial x \partial \pi}. \]

### A.3 Numerical Optimization of Investment and Disinvestment Boundaries

We solve the Hamilton-Jacobi-Bellman equation (13) together with the free boundary conditions for incremental investment and incremental disinvestment numerically by numerically solving the underlying stochastic dynamic program. We discretize (A.2)

\[ V(x, f, \pi) \Delta t = f(x, k) \Delta t + e^{-r \Delta t} E(V(x + \Delta x, k + \Delta k, \pi + \Delta \pi)) \]

and use its homogeneity property in \( k \) to write the program in terms of Tobin’s average \( Q \)

\[ V(x, t, \pi) = k V(x/k, 1, \pi) = k Q(g, \pi) \]

with \( g = x/k \). Since in the inaction region of the problem – where the firm neither invests nor disinvests – capital, \( k \), is constant, thus, \( g \) is just a scaled version of \( x \) there.

In the investment region, we set Tobin’s marginal \( Q \) equal to 1 (the purchase price of capital) and in the disinvestment region we set Tobin’s marginal \( Q \) equal to \( b \) (the resale price of capital). This is so because a firm that enters the investment region will immediately invest the appropriate magnitude of new capital to find its location at the investment boundary thereby incurring the respective investment cost. The argument in
the disinvestment region is analogous. We derive

\[ \frac{\partial}{\partial k}[V(x, h, \pi)] = \frac{\partial}{\partial k} \left[ k Q \left( x, h, \pi \right) \right] = Q(g, \pi) - g \frac{\partial Q(g, \pi)}{\partial g}. \]

Hence, for given investment and disinvestment boundaries \( g^*_i(\pi) \) and \( g^*_d(\pi) \), respectively, the value function inside the action regions is set according to

\[ Q(g, \pi) = \begin{cases} 
\frac{g}{g^*_i(\pi)} Q(g^*_i(\pi), \pi) - \frac{g - g^*_i(\pi)}{g^*_i(\pi)} & g > g^*_i(\pi), \\
\frac{g}{g^*_d(\pi)} Q(g^*_d(\pi), \pi) - \frac{g - g^*_d(\pi)}{g^*_d(\pi)} & g < g^*_d(\pi).
\end{cases} \]

Inside the inaction region, we model the joint dynamics of \( g \) and \( \pi \) as a two-dimensional binomial tree. The difficulty in doing so is that \( \pi \) follows a mean reversion process with non-constant volatility. Therefore, we employ the approach of Nelson and Ramaswamy (1990). Consider the process

\[ Z(\pi) = \int_{1/2}^{\pi} \frac{\sigma}{(\mu_h - \mu_l) p(1 - p)} dp = \frac{\sigma}{\mu_h - \mu_l} \ln \left( \frac{\pi}{1 - \pi} \right), \] (A.5)

with \( \sigma \) defined as the positive square root of \( \sigma^2 \) in Equation (A.2).

Then the process \( z_t = Z(\pi_t) \) has constant volatility of 1 and follows the dynamics

\[ \begin{align*}
\frac{dz}{dt} &= \left[ \mu_\pi \frac{dZ(\pi)}{d\pi} + \frac{1}{2} \sigma^2 \frac{d^2 Z(\pi)}{d\pi^2} \right] dt + \sigma (\Phi')^{-1} dW^F_t \\
\mu_\pi &= \frac{\sigma}{\mu_h - \mu_l} \\
z_0 &= Z(\pi_0)
\end{align*} \] (A.6)

where \( \sigma_\pi \) is the volatility and \( \mu_\pi \) is the risk neutral drift of the belief (see Proposition 1) given by

\[ \begin{align*}
\sigma_\pi &= \frac{(\mu_h - \mu_l) \pi (1 - \pi)}{\sigma} \\
\mu_\pi &= -\pi \lambda_{hl} + (1 - \pi) \lambda_{lh}
\end{align*} \] (A.7)

and \( W^F_{zt} \) is a standard Brownian motion. Consequently, \( z_t \) and \( g_t \) can be modeled on a two-dimensional binomial tree following Boyle, Evnine, and Gibbs (1989), regarding the correlation of the two processes that is implicitly given. Then, the belief process \( \pi \) results
form applying the inverse, \( \pi_t = Z^{-1}(z_t) \), that is given by

\[
Z^{-1}(z) = \frac{1}{1 + \exp\left\{- \left( \frac{\mu_h - \mu_l}{\sigma} \right) z \right\}}.
\]  

(A.8)

The optimization of the free boundaries \( g_i^* \) and \( g_d^* \) is done via value function iteration with the goal to have \( Q(g, \pi) \) smooth at the boundaries, which implies that Tobin’s marginal \( Q \) converges to 1 when moving from the inside of the inaction region towards the investment boundary, and that it converges to \( b \) when moving from the inside of the inaction region towards the disinvestment boundary.

The boundaries discussed in the text, \( h_i^* \) and \( h_d^* \), are calculated simply by

\[
\begin{align*}
    h_i^* &= \frac{1}{g_i^*}, \\
    h_d^* &= \frac{1}{g_d^*}.
\end{align*}
\]
References


