Corporate Investment Over the Business Cycle

Thomas Dangl Youchang Wu

July 2013

Abstract

While firm-level capital growth rates exhibit positive spikes, and rise as fast as they fall, the average capital growth rate across firms exhibits negative spikes, and declines faster than it recovers. We develop a dynamic model of investment that reconciles these empirical patterns. The model features costly reversibility, cyclical macroeconomic shocks, and uncertainty about the state of the economy. A firm's optimal capacity is more sensitive to bad signals in good times than to good signals in bad times. The endogenous distribution of firms relative to their optimal capacities leads to sharp decline and slow recovery at the aggregate level.

JEL codes: G31, G30

Key Words: Investment, Real option, Learning, Business Cycle

Thomas Dangl is at the Institute of Management Science, Vienna University of Technology. Address: Theresianumgasse 27, A-1040 Vienna; Tel: +43-1-58801-33063; Fax: +43-1-58801-33096; Email: thomas.dangl@tuwien.ac.at. Youchang Wu is at the Wisconsin School of Business, University of Wisconsin-Madison. Address: 975 University Avenue, Madison, Wisconsin 53706; Tel: 608-8902595; Fax: 608-2654195; Email: ywu@bus.wisc.edu. We thank David Brown, Bjorn Eraker, Michael Gofman, Richard Green, Christopher Hennessy, Oliver Levine, Xiaoji Lin, Antonio Mello, Erwan Quintin, Mark Ready, Suresh Sundaresan, Sheridan Titman, Grigory Vilkov, Neng Wang, Toni Whited, Randy Wright, and Wei Zhang for helpful discussions. We also appreciate comments from participants at the European Winter Finance Summit 2013, Financial Intermediation Research Society meetings in 2012, Annual Conference of the Swiss Society for Financial Market Research in 2012, China International Conference in Finance in 2011, and seminars at University of Vienna, Vienna University of Economics and Business, Cheung Kong Graduate School of Business, University of Hong Kong, University of Wisconsin-Madison, University of Frankfurt, and Karlsruhe University. A grant from University of Wisconsin-Madison Graduate School is gratefully acknowledged.

Capital investment is one of the most important decisions in corporate finance. It not only determines the long-term prospects of a corporation and the welfare of its investors, but also has profound macroeconomic implications. Although the literature on this subject is vast in both economics and finance, what drives corporate investment remains an elusive question.[1] For example, why are firms in the U.S. reluctant to expand capital as the economy recovers from the recent financial crisis, despite record low interest rates, strong stock market performance, robust corporate earnings, and various stimulation packages?[2]

Using the quarterly Compustat-CRSP merged database from 1975 through 2011, we investigate the dynamics of corporate investment at the firm and aggregate levels. We first plot the average net capital growth rate and the average gross investment rate around the trough quarters dated by the National Bureau of Economic Research (NBER). The graphs show that the decline before the trough is much faster than the subsequent recovery. It takes sixteen quarters for the investment activity to get back to the level six quarters prior to the trough. We then test statistically whether the level and the slope (i.e., first-order difference) of the average capital growth rate are symmetric, and find both of them negatively skewed, especially when the growth rate is measured over a time interval of three or four quarters. The negative skewness of the slope confirms the visual impression that downturns are steeper than upturns (slope asymmetry), while the negative skewness of the level suggests the existence of negative spikes, and indicates that troughs are more extreme than peaks (level asymmetry).

By contrast, capital growth rates of individual firms are strongly positively skewed, suggesting the existence of positive spikes, a well-known feature of corporate investment activity. Furthermore, their first-order differences exhibit virtually zero skewness, indicating symmetric upward and downward slopes. That is, the growth rates run up to the spikes as fast as they come down.

The strikingly different patterns of capital growth rates at the aggregate and firm levels raise several interesting research questions. Why does the average capital growth rate across firms

[1] See Dixit and Pindyck (1994) and Stein (2003) for two excellent surveys of the literature.

[2] See, for example, the article "Investment falls off a cliff: U.S. companies cut spending plans amid fiscal and economic uncertainty" on Wall Street Journal (December 2, 2012).

exhibit negative spikes, while firm-level capital growth rates exhibit positive spikes? Why does the average capital growth rate decline much faster than it recovers, while firm-level capital growth rates rise and fall at an equal speed? What are the economic mechanisms behind these patterns? Can we reconcile them in a unified framework of optimal investment? These are the questions we explore in this paper.

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: (1) the profitability of an individual firm is strongly influenced by macroeconomic conditions; (2) firms face substantial uncertainty about the true state of the economy; (3) disinvestment is more costly than investment. We consider a cross-section of risk-neutral firms with an infinite horizon. Each firm faces its own business conditions, which are summarized by a random demand factor. The expected growth rate of a 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. The true state is not directly observable. Firms update their beliefs continuously by observing their own operating profits and a public signal. Capital can be expanded incrementally and instantaneously 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.

We calibrate the parameters of our model, and derive numerically the optimal investment/disinvestment policy of a typical firm. The policy is characterized by an upper bound and a lower bound on the firm's capital stock normalized by its current demand factor. The lower bound represents the firm's optimal normalized capacity, while the upper bound represents its maximum tolerated normalized capacity. Both bounds 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 with the posterior belief of being in an expansion. Furthermore, this relation is nonlinear. When the posterior probability of expansion is low, its increase leads to only a modest increase of optimal capacity. When this posterior probability is high, its decrease leads to a sharp decline of optimal capacity. By contrast, the disinvestment boundary is concave in the posterior probability of expansion.

The convex relation between the optimal capacity and the posterior probability of expansion is a result of the interaction between the expected growth and the option value of waiting arising from the time-varying uncertainty about the true state of the economy. The intuition is as follows.[3] 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 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 each other. While the signal increases the expected demand growth rate, it also brings more uncertainty, thus increasing the option value of waiting. Consequently, the firm adds capacity only modestly. The intuition for the concavity of the disinvestment boundary is similar.

Interestingly, when signals about the state of the economy are less noisy, the nonlinearities of both boundaries become more pronounced. This is because the benefit of waiting is higher when the signal is more precise. If the next signal can resolve a lot of uncertainty, firms do not rush to invest or disinvest. They behave conservatively with respect to both actions. That is, they invest (disinvest) as if the economy is in a recession (expansion) unless they are very confident that the opposite is true. As a result, both boundaries bend outward from the inaction region.

After deriving the optimal investment/disinvestment policy, we simulate a large number of firms following this strategy. These firms experience the same macroeconomic shocks and observe the same public signal. They also face heterogeneous shocks to their own business

[3] 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.

conditions. We examine the patterns of capital growth rate at both the firm and aggregate levels. Our simulated data display patterns that match the empirical patterns remarkably well. The average capital growth rate declines fast and recovers slowly. Both its level and its slope are negatively skewed. By contrast, the level of firm-level capital growth rates is strongly positively skewed, but their slope is symmetric.

The positive skewness of firm-level capital growth rates is a natural outcome of zero adjustment costs of investment and costly reversibility. The level asymmetry at the aggregate level comes from the fact that economic expansions usually last longer than recessions, which implies that the bulk of observations lie to the right of the sample mean. This asymmetry exists even when investment is perfectly reversible. The slope asymmetry, i.e., the sharp-decline-slow-recovery feature of the average capital growth rate, is due to the endogenous distribution of firms relative to their optimal capacities. During a recession, not only is the optimal capacity relatively insensitive to a modest improvement of the belief, the fraction of firms that are far away from their investment boundaries is also high because of low demand growth. The excess capacities cumulated over time prevent these firms from investing even if their beliefs are changed significantly by the arrival of positive signals. On the contrary, during an expansion, a large fraction of firms are pushed to the investment boundary by high demand growth. When a negative signal arrives, these firms stop investment immediately as their optimal capacities decrease, causing a sharp decline of the capital growth rate at the aggregate level.

Our comparative static analysis shows that a small cost of disinvestment can have big impacts. While the slope of the average capital growth rate has zero skewness under perfect reversibility, a small discount of 2% in the capital resale price is enough to generate significant slope asymmetry. A discount of 20% leads to capital growth rate patterns that are very similar to those obtained under perfect irreversibility.

Our paper contributes to the literature on capital investment using the real options approach. Our model echoes an important insight of this literature: uncertainty increases the option value of waiting and delays investment. This point has been made in several seminal papers, including Bernanke (1983), McDonald and Siegel (1986), Pindyck (1988). 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 at both turning points of the economic environment, the joint effects of signals on the first and the second moments of conditional beliefs, and the endogenous distribution of firms relative to their optimal capacities, lead to asymmetric reactions to signals at different phases of business cycles.

Our model builds on the structure of Guo, Miao, and Morellec (2005), who investigate optimal irreversible investment under random regime shifts. However, our paper is significantly different. 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 at both the firm and aggregate levels, and calibrate our model to replicate the major features of the data. Furthermore, we extend 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 as the high growth state starts, which is inconsistent with the slow recovery. Second, while investment is completely irreversible in Guo, Miao, and Morellec (2005), we allow for different degrees of reversibility, encompassing complete reversibility and irreversibility as two special cases. This allows us to investigate the economic impacts of reversibility.

Our paper contributes to the understanding of corporate investment under incomplete information. Alti (2003) investigates optimal investment policy when the productivity of capital is unknown and has to be inferred from realized cash flows. 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, rendering their models 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 downturns are more abrupt and violent than upturns.[4] The empirical evidence is somehow mixed. While Neftci (1985), Sichel (1993) find evidence supporting the asymmetry hypothesis, Falk (1986) and DeLong and Summers (1986) conclude that there is very little evidence of asymmetry. These studies examine only 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. To the best of our knowledge, we are the first to contrast directly the asymmetries of capital growth rates at the firm and aggregate levels, and provide a dynamic model to reconcile them. Treating macroeconomic conditions as an exogenous latent process, we do not aim to give an explanation for the sources of economic fluctuations. However, our model describes how firms at the micro level optimally respond to changing economic environments, and how such responses translate cyclical shocks into asymmetric time series of capital growth rates.

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, describe the optimal learning process, and derive the partial differential equation and boundary conditions for the optimal investment policy. In section 3 we calibrate the parameters for the base case of the model. Section 4 characterizes the optimal investment

[4] One of the most famous statements was 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."

policy. Section 5 compares the simulated and empirical patterns of capital growth rates at the firm and aggregate levels. Section 6 presents comparative static results with respect to depreciation and reversibility. Section 7 concludes the paper.

1 Asymmetries of Capital Growth Rates: Empirical Patterns

1.1 Capital Growth Rates Over the Business Cycle

To examine the empirical patterns of firms' investment behavior, we use the quarterly Compustat-CRSP merged database from 1975 through 2011.[5] 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. This is a normalized measure of net investment, i.e., gross investment minus depreciation. All nominal values are converted into year 2005 dollars using the quarterly GDP deflator. A firm enters the sample when its PPENT reaches \$1 million in year 2005 dollars, and stays in the sample even if its PPENT later drops below it. 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 exclude firms that are heavily involved in mergers and acquisitions in a give year. 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 in a give fiscal year, a firm acquires assets worth more than 20% of its total assets at the end of the prior fiscal year, we exclude that fiscal year from our sample. Our final sample consists of 532,306 firm-quarter observations, with an

[5]Before 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.

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

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

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

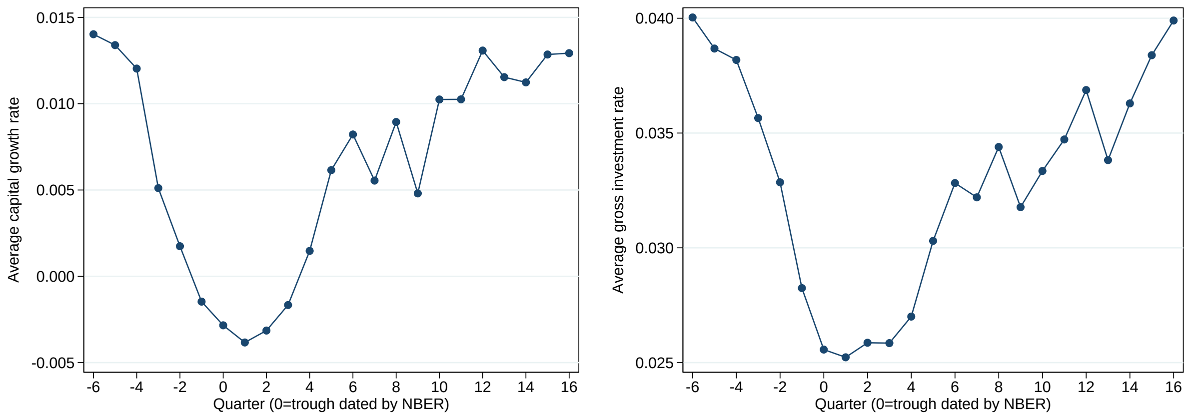
We first investigate the patterns of the average capital growth rate over the business cycle. We select an event window of 23 quarters around the quarters that are identified by the NBER as a business cycle trough. According to the NBER, there are six cycles between 1975 and 2011. A trough is a turning point that marks the end of a contraction and the start of an expansion.[6] Our event window goes from 6 quarters before the trough through 16 quarters after it.

Panel (a) of Figure 1 shows the average capital growth rate across firms, 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 the 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 quoted in footnote 4.

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

$$g'_{t,t-1} = ln(PPEGT_{t-1} + CAPX_t) - ln(PPEGT_{t-1}), \tag{2}$$

[6] The six trough quarters are 1975 quarter I, 1980 quarter III, 1982 quarter IV, 1991 quarter I, 2001 quarter IV, and 2009 quarter II.



(a) Average net capital growth rate

(b) Average gross investment rate

Figure 1: **Average net capital growth rate and gross investment rate over the business cycle.** Panel (a) shows the average capital growth rate across firms around the business cycle troughs dated by the NBER. Panel (b) shows the average gross investment rate across firms. 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.

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. Similar to the average net capital growth rate, the average gross investment rate also shows a sharp decline before the trough and a slow recovery subsequently. In the rest of the paper we focus on the net capital growth rate, on which a longer times series of data is available.[7]

1.2 Skewness of Capital Growth Rates

Figure 1 shows that the upturn and downturn of the average capital growth rate are asymmetric. To quantify the asymmetry, we use a standard measure 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 that the bulk of the observations lie to the right of the mean. A positive skewness value indicates the opposite. A symmetric distribution has zero skewness.

Following Sichel (1993), we distinguish between two types of asymmetry: level asymmetry

[7]The quarterly CAPX data is only available starting from 1984.

(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 the level asymmetry, and the skewness of its first-order difference to measure the slope asymmetry. We examine these asymmetries at both the firm and aggregate levels.

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

$$g_{t,t-n} \equiv ln(PPENT_t) - ln(PPENT_{t-n}), \tag{3}$$

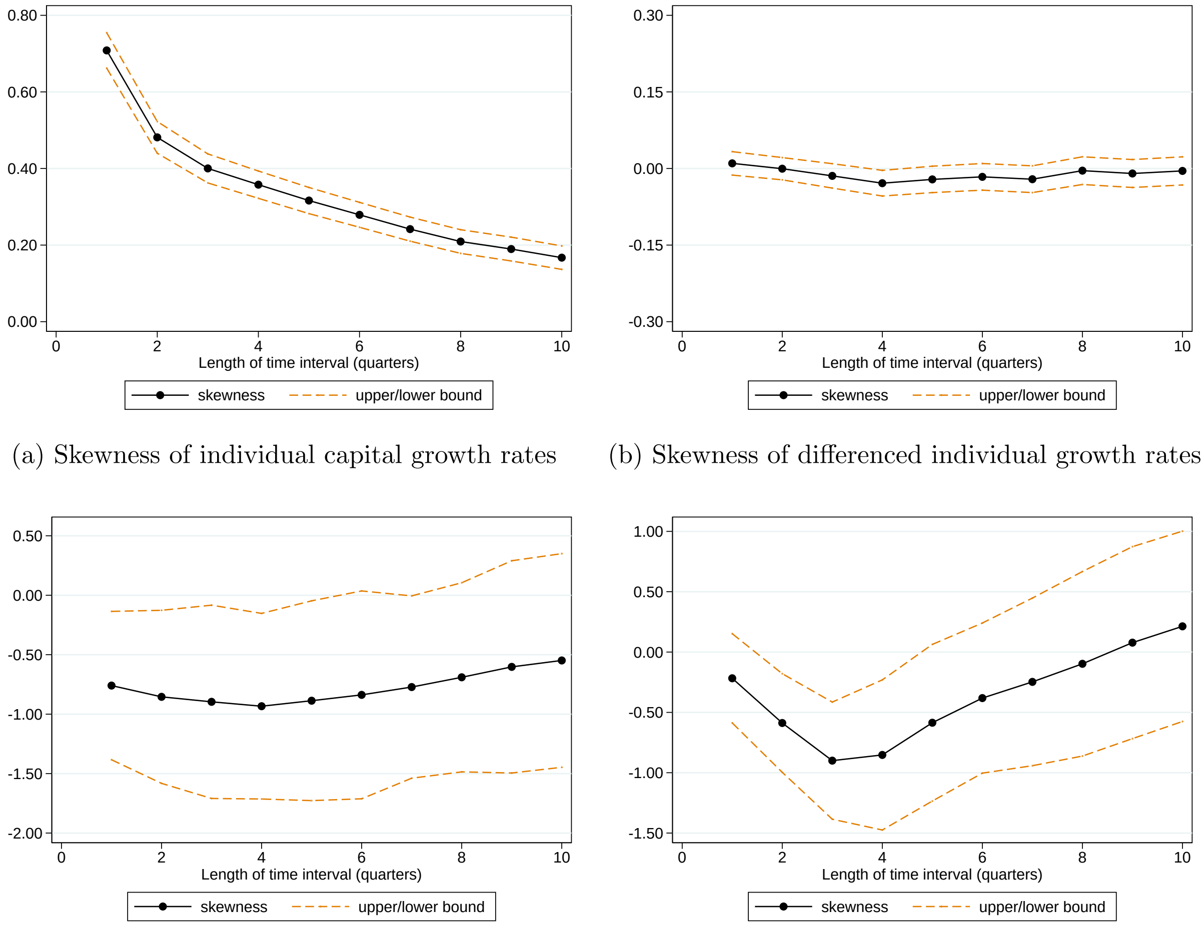
where n varies from 1 to 10. The first-order difference of the n-quarter capital growth rate is defined as

$$\Delta g_{t,t-n} \equiv g_{t,t-n} - g_{t-n,t-2n}. \tag{4}$$

For the asymmetries at the firm level, we first calculate the skewness values of the n-quarter capital growth rate and its first-order difference for each firm, and then average them across firms with at least 40 growth rate observations.[8] For the asymmetries 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 measure the level asymmetry, and the skewness of its first-order difference to measure the slope asymmetry.

Panels (a) and (b) show the average skewness value for the level and the slope of firm-level capital growth rates, respectively. Panels (c) and (d) show the skewness values for the level and slope of average capital growth rates, 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,

[8]For each growth rate series, we exclude the observations that are below the 1st or above the 99th percentile of the whole sample.



(a) Skewness of individual capital growth rates

(b) Skewness of differenced individual growth rates

(c) Skewness of average capital growth rates

(d) Skewness of differenced average growth rates

Figure 2: **Skewness of capital growth rates: empirical estimates.** Panels (a) and (b) show the average skewness values for the level and the slope of firm-level capital growth rates, respectively. Panels (c) and (d) show the skewness values for the level and slope of average capital growth rates, 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.

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 skewness values 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 computed using the standard method because of 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.[9] 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 skewness values 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 for the level 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-order differences of average capital growth rates are also negatively skewed, for time intervals up to eight quarters, indicating a steeper slope for the downturn, and lending support to the hypothesis of fast decline and slow recovery. The estimated values of skewness for the slope first decrease and then increase with the length of the time interval. At the quarterly interval ($n = 1$), it is -0.22, insignificantly different from zero. It becomes significantly negative for $n = 2, 3$ and 4. This suggests that while the decline of the average capital growth rate in a given quarter is not abnormally big, the consecutive declines over multiple quarters generate strong 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.

[9] High orders of the autoregressive model do not change the significance of our results.

By contrast, capital growth rates at the firm level exhibit a strong level asymmetry with positive skewness (Panel (a)) and almost no slope asymmetry (Panel (b)). For the level, the average skewness value across firms is 0.71 at the quarterly interval. It deceases 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 absolute values of skewness for the first-order differences are below 0.03 for all time intervals, indicating that at the firm level, changes in capital growth rates are symmetric.

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

2 A Dynamic Model of Investment

2.1 Setup

We consider a cross-section of risk-neutral firms with an infinite time horizon. The firms are *ex ante* identical, but face both common and heterogeneous shocks. A typical firm operates in an environment similar to that of Guo, Miao, and Morellec (2005), except that it has incomplete information and that investment is not entirely irreversible. Time is continuous. Investment is incremental. Each firm's cash flow is driven by a distinct stochastic factor. Depending on the state of the macroeconomy, the expected growth rate of this factor shifts between a high level and a low level 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:

$$f(x_t, k_t) = \frac{1}{1-\alpha} x_t^{\alpha} k_t^{1-\alpha}, \tag{5}$$

where $(k_t)_{t\geq 0}$ represents the process of the firm's net capital stock, $(x_t)_{t\geq 0}$ represents the process of a demand factor.[10] 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:

$$d\,\ln(x_t) = \mu_t dt + \sigma_x dW_{xt},\ \ x_0 > 0, \tag{6}$$

where W_{xt} is a standard Wiener process, $ln(x)$ is the natural logarithm of x.[11] The volatility σ_x is a known constant. The expected (continuously-compounded) growth rate of the demand factor, μ_t, is determined by the macroeconomic condition. It is low in a recession ($\mu_t = \mu_l$) and high in an expansion ($\mu_t = \mu_h > \mu_l$). 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, μ_t switches randomly between μ_h and μ_l. More specifically, we assume that μ_t is driven by a continuous-time Markov jump process with the transition probabilities:

$$P(\Delta t) = \mathbb{I} + \begin{pmatrix} -\lambda_{h,l} & +\lambda_{h,l} \\ +\lambda_{l,h} & -\lambda_{l,h} \end{pmatrix} (\Delta t + o(\Delta t)), \tag{7}$$

[10] The factor x_t can be generally 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.

[11] We write the stochastic differential equation in terms of $d\ln(x)$ instead of dx so that the expected continuously compounded growth rate is not affected by the volatility. Note also that σ_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 σ_x differs across states.

where $\mathbb{I}$ is the identity matrix and $\lambda_{h,l}$ and $\lambda_{l,h}$ are the constant intensities of transition from μ_h to μ_l and vice versa.[12] 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, which is lower than one. The wedge between the purchase and resale prices of capital, $1 - b$, captures the feature of costly reversibility. Installed capital depreciates at a constant rate of ξ.

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 μ_t. In other words, the true state of the economy that determines μ_t is a hidden process. This implies that W_{xt} is not observable as well. When the firm observes a certain increase or decrease in the cash flow, it does not know which portion of it comes from the drift μ_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:

$$d\ln(s_t) = \mu_t dt + \sigma_s dW_{st}, \;\; s_0 > 0; \tag{8}$$

where W_{st} is a standard Wiener process, σ_s is a publicly known constant representing the volatility of the signal process. σ_s measures the noisiness of the public signal. It characterizes (inversely) the accuracy of the publicly available data about the economy. We assume the instantaneous correlation between the two Wiener processes, W_{st} and W_{xt}, to be a known

[12] 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.

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 the expected growth rates of x_t and s_t.

To summarize the information structure of our model, let $\mathcal{F}_t$ be the canonical nondecreasing 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 μ_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

By observing the demand factor x_t and the signal s_t over the time interval $[0, t)$, the firm can continuously update its belief about the state of the economy. To formalize the rational learning rule, we denote by π_t the probability that $\mu_t = \mu_h$ conditional on $\mathcal{F}_t$ and a prior π_0. Consequently, the rational conditional mean of the current continuously compounded growth rate is $\pi_t \mu_h + (1 - \pi_t)\mu_l$. 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.[13]

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

$$d\pi_t = [-\pi_t \lambda_{h,l} + (1 - \pi_t)\lambda_{l,h}]dt + (\mu_h - \mu_l)\pi_t(1 - \pi_t)\mathbf{1}'(\Phi')^{-1}\mathbf{dW_t^{\mathcal{F}}}, \tag{9}$$

where $\mathbf{W_t^{\mathcal{F}}}$ is a two-dimensional independent Wiener process with respect to $\mathcal{F}_t$ defined as

$$\mathbf{dW_t^{\mathcal{F}}} \equiv \begin{pmatrix} dW_{xt}^{\mathcal{F}} \\ dW_{st}^{\mathcal{F}} \end{pmatrix} \equiv \Phi^{-1} \begin{pmatrix} d\ln(x_t) - E(\mu_t|\mathcal{F}_t)dt \\ d\ln(s_t) - E(\mu_t|\mathcal{F}_t)dt \end{pmatrix}, \tag{10}$$

[13]See David (1997) for an early application of this filter in finance.

Φ is a 2×2 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 $\mathbf{1}$ *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 bewteen $dW_{xt}^{\mathcal{F}}$ and $dW_{st}^{\mathcal{F}}$ can be established by noting

$$dW_{xt}^{\mathcal{F}} dW_{st}^{\mathcal{F}} = (1,0)\,\Phi^{-1}(\Phi\Phi' dt)(\Phi^{-1})'(0,1)' = 0.$$

□

The diffusion process π_t is bounded between 0 ($\mu_t = \mu_l$, 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:

$$\bar{\pi} = \frac{\lambda_{l,h}}{\lambda_{h,l} + \lambda_{l,h}}, \tag{11}$$

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. $E(\mu_t|\mathcal{F}_t)dt$ represents the best forecast of $d\ln(x_t)$ and $d\ln(s_t)$ conditional on the information set $\mathcal{F}_t$, and $d\mathbf{W}_t^{\mathcal{F}}$ 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, $(\mu_h - \mu_l)$, is greater; or (ii) the uncertainty about the state of the economy, captured by the conditional variance of the belief, $\pi_t(1-\pi_t)$, 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., π_t 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 π_t is adjusted upward (downward). The information quality of this compounded signal is measured by the inverse of its variance, $\frac{1}{\sigma^2} = \mathbf{1}'(\Phi\Phi')^{-1}\mathbf{1}$. 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 $\mathbf{w}$ 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 $\mathbf{W}_t^{\mathcal{F}}$ is a Wiener process with respect to the filtration $\mathcal{F}_t$ 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 $\mathcal{F}_t$-Markov process.

Using $dW_{xt}^{\mathcal{F}}$ and $dW_{st}^{\mathcal{F}}$ defined by equation (10), we can rewrite the joint dynamics of x_t and s_t in terms of unexpected changes with respect to $\mathcal{F}_t$:

$$\begin{pmatrix} d\ln(x_t) \\ d\ln(s_t) \end{pmatrix} = [\pi_t\mu_h + (1-\pi_t)\mu_l] \begin{pmatrix} 1 \\ 1 \end{pmatrix} dt + \Phi \begin{pmatrix} dW_{xt}^{\mathcal{F}} \\ dW_{st}^{\mathcal{F}} \end{pmatrix}, \tag{12}$$

where π_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, an inaction region, and a disinvestment region. The firm remains inactive in the interior area of the inaction 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 inaction region, and then specify the boundary conditions that characterize the optimal investment/disinvestment policy.

From the previous section we know that the Bayesian belief π_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 π_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 π_t. 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 inaction region, the normalized firm value,* v*, has to satisfy the partial differential equation (Hamilton-Jacobi-Bellman equation):*

$$\begin{aligned}
[r - (\pi\mu_h + (1-\pi)\mu_l + \frac{1}{2}\sigma_x^2)]v \quad = \quad & \frac{h^{1-\alpha}}{1-\alpha} - h[\pi\mu_h + (1-\pi)\mu_l + \frac{1}{2}\sigma_x^2 + \xi]\frac{\partial v}{\partial h} + \frac{1}{2}\sigma_x^2 h^2 \frac{\partial^2 v}{\partial h^2} \\
& + [-\pi\lambda_{hl} + (1-\pi)\lambda_{lh} + \pi(1-\pi)(\mu_h - \mu_l)]\frac{\partial v}{\partial \pi} \\
& + \frac{[\pi(1-\pi)(\mu_h - \mu_l)]^2}{2\sigma^2}\frac{\partial^2 v}{\partial \pi^2} \\
& - h\pi(1-\pi)(\mu_h - \mu_l)\frac{\partial^2 v}{\partial h \partial \pi},
\end{aligned} \tag{13}$$

with $\sigma^2 \equiv \frac{1}{\mathbf{1}'(\Phi\Phi')^{-1}\mathbf{1}}$*, as defined in equation* (A.2).

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.[14]

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 π, because the marginal value of invested capital depends on the growth prospects of the firm.

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^*, which is also a function of the belief π, the firm finds it optimal to incrementally divest excess capital at a discount of $1 - b$. At this threshold, every negative shock in demand will be accommodated by proper disinvestment such that the region above h_d^* will never be entered.

[14] This homogeneity property simplifies the solution of the valuation equation because it reduces the dimensionality of the problem. It arises 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. 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 π_t.

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 corresponding boundary condition is

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

These conditions can be written in derivative form as

$$\lim_{k \to x\, h_i^*} \frac{\partial V}{\partial k} = 1, \quad \lim_{k \to x\, h_d^*} \frac{\partial V}{\partial k} = b.$$

The value-matching conditions above follow 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

$$\lim_{h \to h_i^*} \frac{\partial v(h, \pi)}{\partial h} = 1, \quad \lim_{h \to h_d^*} \frac{\partial v(h, \pi)}{\partial h} = b. \tag{14}$$

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 both 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 following conditions for $v(h, \pi)$ at both the investment and disinvest-

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.

Notation	Economic meaning	Value
α	operating profit parameter	0.74
μ_h	expected continuously compounded growth rate in expansion	0.0938
μ_l	expected continuously compounded growth rate in recession	-0.0629
$\lambda_{h,l}$	transition intensity from expansion to recession	0.2504
$\lambda_{l,h}$	transition intensity from recession to expansion	0.6345
ξ	depreciation rate	0.1211
σ_x	instantaneous volatility of demand factor x_t	0.3036
σ_s	instantaneous volatility of signal s_t	0.15
b	resale price of one unit of capital	0.80
r	risk-free rate	0.05
ρ	instantaneous correlation between dW_{st} and dW_{xt}	0.05

ment boundaries:

$$\lim_{h \to h^*} \frac{\partial^2 v(h,\pi)}{\partial h^2} = 0, \quad \lim_{h \to h^*} \frac{\partial^2 v(h,\pi)}{\partial h \partial \pi} = 0. \tag{15}$$

3 Calibration

Our model does not have an analytical solution, so we solve it numerically. We now describe our procedure to calibrate the model parameters, which are summarized in Table 1.

We first estimate the business cycle parameters, μ_h, μ_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:

$$x_t = \left[\frac{(1-\alpha) f(x_t, k_t)}{k_t^{1-\alpha}} \right]^{1/\alpha}. \tag{16}$$

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, all values are converted into year 2005 dollars using the annual GDP deflator. We exclude firm/year

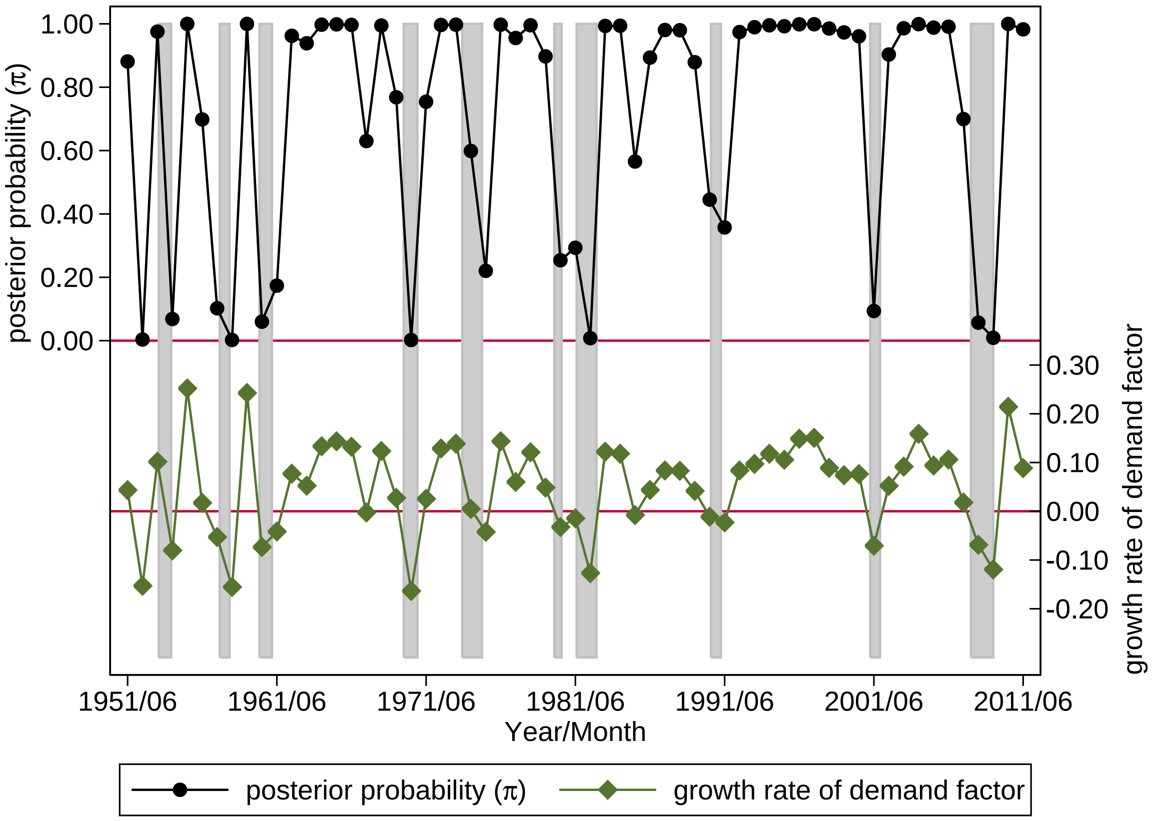


Figure 3: **Growth rate of the demand factor and posterior probability of expansion.** The bottom part of the figure (axis on the right) shows the annual growth rate (continuously-compounded) of the demand factor x, $\Delta ln(x)$, averaged across firms in the Compustat-CRSP merged database. The upper part of the figure shows the time series of the posterior belief π estimated from the time series of $\Delta ln(x)$ using the EM algorithm. The shaded areas indicate the recession periods dated by the NBER.

observations where the operating asset value is less than \$5 million (in year 2005 dollars). We set the operating profit parameter α equal to 0.74, along the line of Guo, Miao, and Morellec (2005).[15]

After estimating the x_t series for each firm, we calculate its continuously compounded annual growth rates, and average them across firms (after excluding observations below the 1st or above the 99th quantile of the sample). This gives us a time series of 60 annual observations.

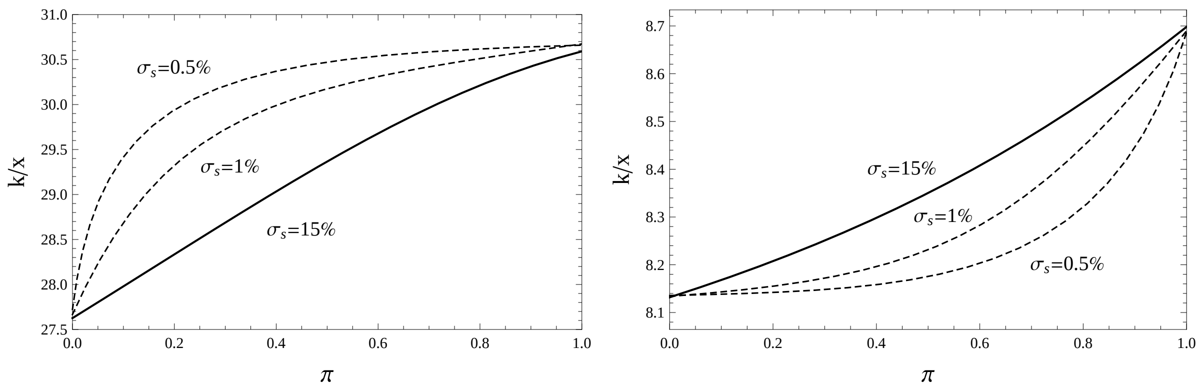
[15] As explained in footnote 7 of Guo, Miao, and Morellec (2005), the operating profit function (5) approximates the following specification: (1) Constant returns to scale Cobb-Douglas production function with labor and capital: $q = \lambda L^{\phi} K^{1-\phi}$; (2) Isoelastic demand function given 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 θ and ϕ through the following relation: $1 - \alpha = (1-\phi)\theta/(1-\theta\phi)$. Labors 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 $1 - \alpha$. Using alternative values of α changes the estimates of μ_h and μ_l accordingly, but has little effect on our main results.

We apply the Expectation-Maximization (EM) algorithm developed by Dempster, Laird, and Rubin (1977) to estimate the parameters of the hidden Markov chain: $\lambda_{h,l}$, $\lambda_{l,h}$, μ_h, μ_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 on average 10.5 quarters and recessions 4.1 quarters. We plot the time series of estimated average annual growth rate of the demand factor, and the posterior beliefs π_t in Figure 3. The shaded areas indicate the historical expansion and recession periods as identified by the NBER. One can see that the π_t series matches the NBER-dated business cycles very well. Recession periods are always associated with π_t less than 0.5, while expansion periods are associated with π_t higher than 0.5 except for the year 1952. Our results also indicate that expected growth rates in expansions and recessions are quite different: 0.0938 vs. -0.0629. This highlights the importance of the macroeconomic state for firms' profitability and investment decisions.

The other parameter values for the base case are determined as follows. The depreciation rate ξ is 0.1211. This is estimated using the same 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, σ_x, is 0.3036. We regress the annual demand growth rate on an NBER recession dummy firm by firm, using firms with at least 30 annual observations, and then take a simple average of the root mean squared error of each regression to obtain this estimate.

The instantaneous volatility, σ_s, of the public signal is 0.15, which is equal to the annual volatility of the U.S. stock market returns from 1951 to 2011. The risk-free rate is 0.05. This value is relatively high compared to the historical data. Firms in our model are risk neutral, and use the risk-free rate to discount future operating profits. In practice the discount rate is higher than the risk-free rate because of risk aversion. Our risk-free rate can be interpreted as a certainty equivalent of the discount rate under a more general setup with risk aversion. The



(a) Optimal disinvestment boundary (b) Optimal investment boundary

Figure 4: **Optimal investment and disinvestment boundaries.** This figure depicts the optimal investment boundary (Panel (b)) and disinvestment boundary (Panel (a)) as a function of the belief π for different levels of information quality. The solid curves represent the base case ($\sigma_s = 15\%$). The dashed curves represent cases with higher information quality (σ_s=1% and σ_s=0.5%, respectively). All other parameter values are fixed at the base case level specified in Table 1.

resale price of capital, b, is 0.80. As we show in Section 6, while lower values of b are associated with a more pronounced slope asymmetry of the average capital growth rate, the impact of a further decrease of b beyond 0.8 is fairly small. The instantaneous correlation, ρ, between the public signal and the firm's demand factor is 0.05. Our results are robust to other values of ρ.

4 Optimal Investment Policy

4.1 Optimal Investment Boundary

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 π 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.

Figure 4 shows the investment and disinvestment boundaries as a function of the current belief π. These boundaries represent, respectively, the optimal and maximum tolerated capacities normalized by the current demand factor. The area between them is the inaction region, in which the firm remains inactive, i.e., it does not adjust its capital. The solid curves represent the boundaries for the base case parameterization summarized in Table 1. The dashed curves represent cases with higher information quality, i.e., lower σ_s of the public signal s_t.

Panel (b) plots the investment boundary, i.e., the lower bound for the firm's normalized capital stock. The firm increases its capital whenever k/x hits the boundary from above (as capital depreciates or demand increases) or from the left (as π 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 π, indicating that the firm's investment decision is relatively insensitive to changes in the belief when π is low. This results from the interaction of the expected growth rate and uncertainty about the growth rate. Starting from π 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 π 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.

Comparing the investment boundaries under different levels of information quality, we find that more precise signals make the investment 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, and a higher option value of waiting. Higher information quality thus amplifies the convexity of the boundary. For σ_s close to zero, the option value of waiting is so high that the firm always invests conservatively as if it were in a recession unless it is almost sure of being in an expansion. As a result, the investment boundary is almost flat for π below one and increase sharply as π approaches one.

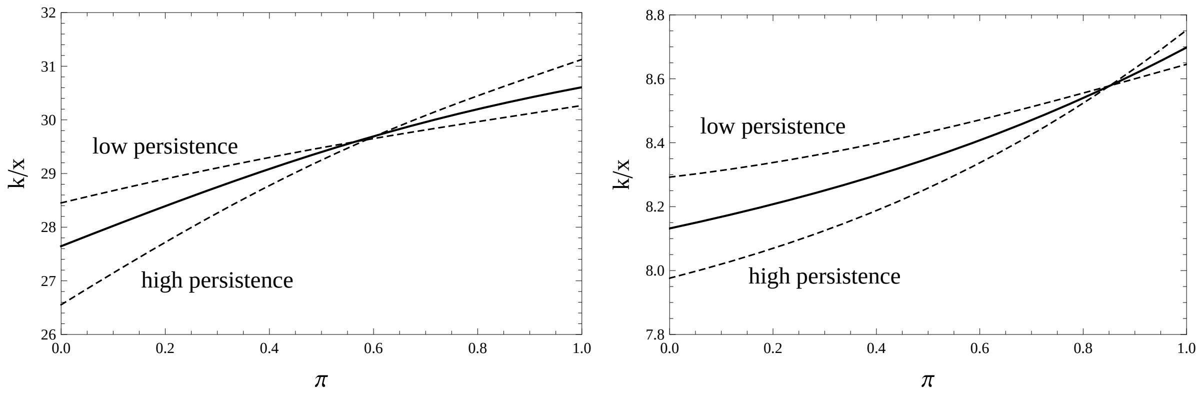
Panel (a) of of Figure 4 shows the disinvestment boundaries for different levels of information. These boundaries characterize the maximum tolerated normalized capacity, k/x. When a firm experiences negative demand shocks such that invested capital is high relative to demand, demand persistence creates an incentive for the firm to sell capital even at a discount. Thus, if a firm's normalized capacity hits this boundary (from below or from the right), disinvestment takes place, which pushes k/x back towards the inaction region.

The disinvestment boundaries are concave, and the concavity becomes more pronounced as the information quality increases. The intuition for this concavity is similar to that for the convexity of investment boundary. 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. By contrast, 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. When information quality of the signal is high, the option value of waiting is high, and the disinvestment boundary becomes flat at high values of π. The firm only starts to sell capital aggressively (i.e., at relatively low values of k/x) when π is close to zero, i.e., when it is almost sure that the current state is the low growth state.

4.2 Effects of Regime Persistence

Figure 5 shows the effects of regime persistence on the investment and disinvestment boundaries. When both states of the business cycles become more persistent (as the transition intensities $\lambda_{h,l}$ and $\lambda_{l,h}$ decrease proportionally), the optimal investment and disinvestment boundaries become steeper. This is quite intuitive. If both the expansion and the recession are more likely to persist, the gap between the marginal values of capital in these two regimes is wider, and the boundaries are more sensitive to the belief.

More subtly, The figure also shows that the negative impact of higher persistence on the left end ($\pi \to 0$) of the investment boundary is stronger than its positive impact on the right end ($\pi \to 1$). Costly reversibility is key for this asymmetric effect. Since increasing capacity



(a) Disinvestment boundary

(b) Investment boundary

Figure 5: **Optimal investment and disinvestment boundaries: effects of regime persistence.** This figure shows the optimal investment (Panel(b)) and disinvestment (Panel (a)) boundaries under different degrees of regime persistence. The solid curves in both panels represent the base case parameterization. The dashed curves represent two alternative degrees of persistence. The transition intensities ($\lambda_{h,l}$ and $\lambda_{l,h}$) in the low (high) persistence case is double (half) of their base case values. All other parameters are fixed at the base case level specified in Table 1.

after a shift from a recession to an expansion requires no adjustment costs, higher persistence of the expansion has almost no effect on the investment boundary in the recession. However, disinvestment after a reverse shift is costly, therefore higher persistence of the recession has a strong negative impact on the investment boundary in both the recession and the expansion. Anticipating that the excess capacity problem is more severe after a shift to a more persistent recession, the firm rationally refrains from investing too aggressively in the expansion, even though the expansion is expected to last longer.

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 π converges to zero or one. Furthermore, at the aggregate level, the speed of capital adjustment depends crucially on the distribution of firms' normalized capacities relative to the 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.[16] The common macroeconomic regime follows 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 6 shows the equal-weighted average capital growth rate over a full business cycle (starting with a recession). The solid line is empirical average capital growth rate reproduced from Panel (a) of Figure 1. The black dashed line is the simulated average growth rate under the base case parameterization specified in Table 1 (with $\sigma_s = 15\%$). 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.

Our simulated capital growth rate under the baseline parameterization features a sharp decline and a slow recovery, although the recovery is not as slow as it is empirically. The simulation under almost complete information shows a more dramatic decline at the beginning of a recession. At the same time, it also exhibits a sharp and immediate rebound at the

[16] 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.

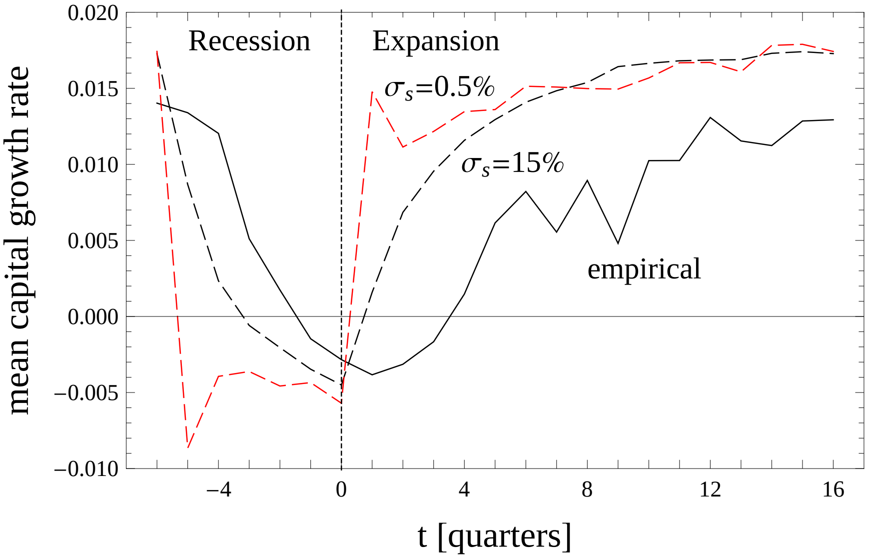
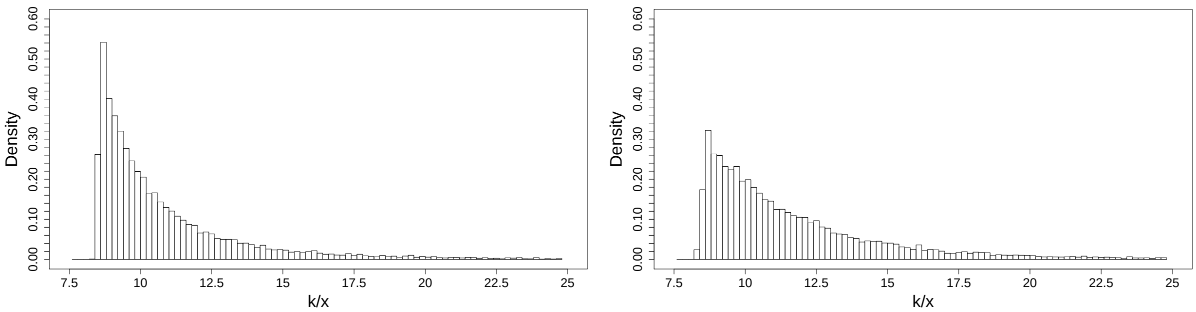


Figure 6: **Mean capital growth rate over the business cycle.** This graph shows the simulated and the empirically estimated equal-weighted average capital growth rates over a full business cycle. The solid line represents the empirical quarterly real capital growth rate around the business cycle troughs dated by the NBER, reproduced from Panel (a) of Figure 1. The black and red dashed lines represent the simulated average quarterly capital growth rates with $\sigma_s = 15\%$ and $\sigma_s = 0.5\%$, respectively. Other parameter values are given in Table 1. 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.

beginning of an expansion. 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.[17] These dramatic changes do not match well the empirical data.

The reason for the asymmetry of decline and recovery in our baseline model is the endogenous distribution of firms relative to their optimal capacities. Since the optimal capacity is high and demand grows fast in an expansion, a large fraction of 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 immediately as their optimal capacities decrease, generating a sharp decline in the average capital growth rate. By contrast, at the end of a recession, not only is the optimal capacity relatively insensitive to a

[17]Note that the capital growth rate in the almost-perfect information 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.



(a) At the end of an expansion

(b) At the end of a recession

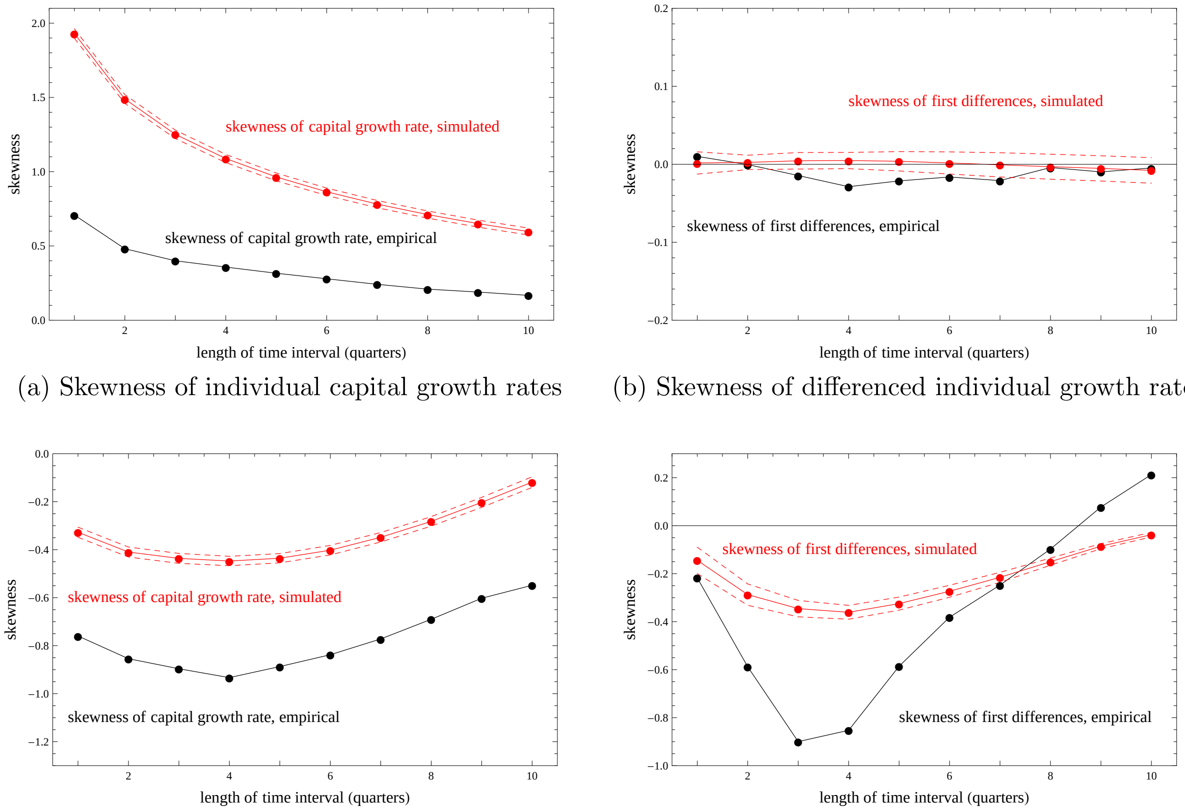
Figure 7: **Histograms of normalized capital: Expansion vs. recession.** This figure shows the histograms of simulated normalized capital (k/x) at two turning points of a business cycle: at the end of an expansion and at the end of a recession. The parameter values used for the simulation are given in Table 1.

modest improvement of the belief, the fraction of firms that are far away from their investment boundaries is also high because of the low demand growth. The excess capacities cumulated over time prevent these firms from investing even if their beliefs are changed significantly by the arrival of a positive signal — their capacities are too high even compared to the boundary for an expansion period. Consequently, the reaction to the positive signal is small, and the recovery is slow.[18]

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 7. 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 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 recovery appears to be faster in our model than in the data. This discrepancy may arise because we assume perfectly synchronized regime shifts across firms. In reality firms do not

[18]The fraction of firms at the disinvestment boundary is generally small due to the costs of disinvestment. Those firms do not have much impact on the speed of the decline or the recovery.



(a) Skewness of individual capital growth rates

(b) Skewness of differenced individual growth rates

(c) Skewness of average capital growth rates

(d) Skewness of differenced average growth rates

Figure 8: **Skewness of capital growth rates: empirical estimates vs. simulated results.** Panels (a) and (b) show the average skewness values for the level and the slope of firm-level capital growth rates, respectively. Panels (c) and (d) show the skewness values for the level and slope of average capital growth rates, 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 skewness reproduced from Figure 2. The red curves are the skewness values estimated from the simulated data and their 95% confidence intervals. The parameter values used for the simulation are given in Table 1.

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 an expansion.

5.2 Skewness of Capital Growth Rates

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

Panels (a) and (b) show the average skewness values for the level and the slope of firm-level capital growth rates, respectively. Panels (c) and (d) show the skewness values for the level and slope of average capital growth rates, 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 skewness reproduced from Figure 2. The red curves are the skewness values estimated from the simulated data and their 95% confidence intervals.[19]

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 positive skewness (Panel (a)), and virtually no slop asymmetry (Panel (b)). For both empirical and simulated firm-level capital growth rates, the positive skewness is highest at the one-quarter interval, and deceases 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 skewness values in Panel (a) is that the simulated skewness values 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 pay adjustment costs for both investment and disinvestment, even though the costs for disinvestment are likely to be higher. 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 level and the slope of the average capital growth rates are negatively skewed. Furthermore, the skewness values for both the level and

[19] We report the confidence intervals to account for simulation errors. To calculate the confidence intervals, 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 $\sqrt{42}$.

the slope are non-monotonic in the length of time interval, reaching their minimums when growth rates are measured over an interval of three- and four-quarters, and slowly increasing towards zero as the time interval further lengthens. The particularly strong skewness at the three- and four-quarter intervals has to do with the fact that recessions on average last about six quarters. Even though the magnitudes of the skewness values 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 rate across firms.

The level asymmetry of the average capital growth rates has to do with the fact that expansions usually last longer than recessions, which implies that the bulk of short-run growth rate observations are from the expansion periods and therefore lie to the right of the mean. The relatively small number of observations from the recession periods 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 corporate investment. As we discuss in section 5.1, this occurs due to the endogenous distribution of firms relative to their optimal capacities over the business cycle. 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 8 suggests that our model is able to reconcile two seemingly conflicting patterns of firms' capital growth rates: for individual firms, the level of capital growth rates is positively skewed, but the slope is symmetric; in the aggregate, both the level and the slope are negatively skewed.

6 Effects of Reversibility and Depreciation

In this section we study the consequences of changes in reversibility and depreciation rate on investment and disinvestment boundaries, and on the characteristics of capital growth rates. On the one hand, this analysis serves as a robustness check for our base case results. On the other hand, the comparative static results obtained through this analysis generate predictions about the dependence of investment behavior on certain firm characteristics.

6.1 Differences in the Degree of Reversibility

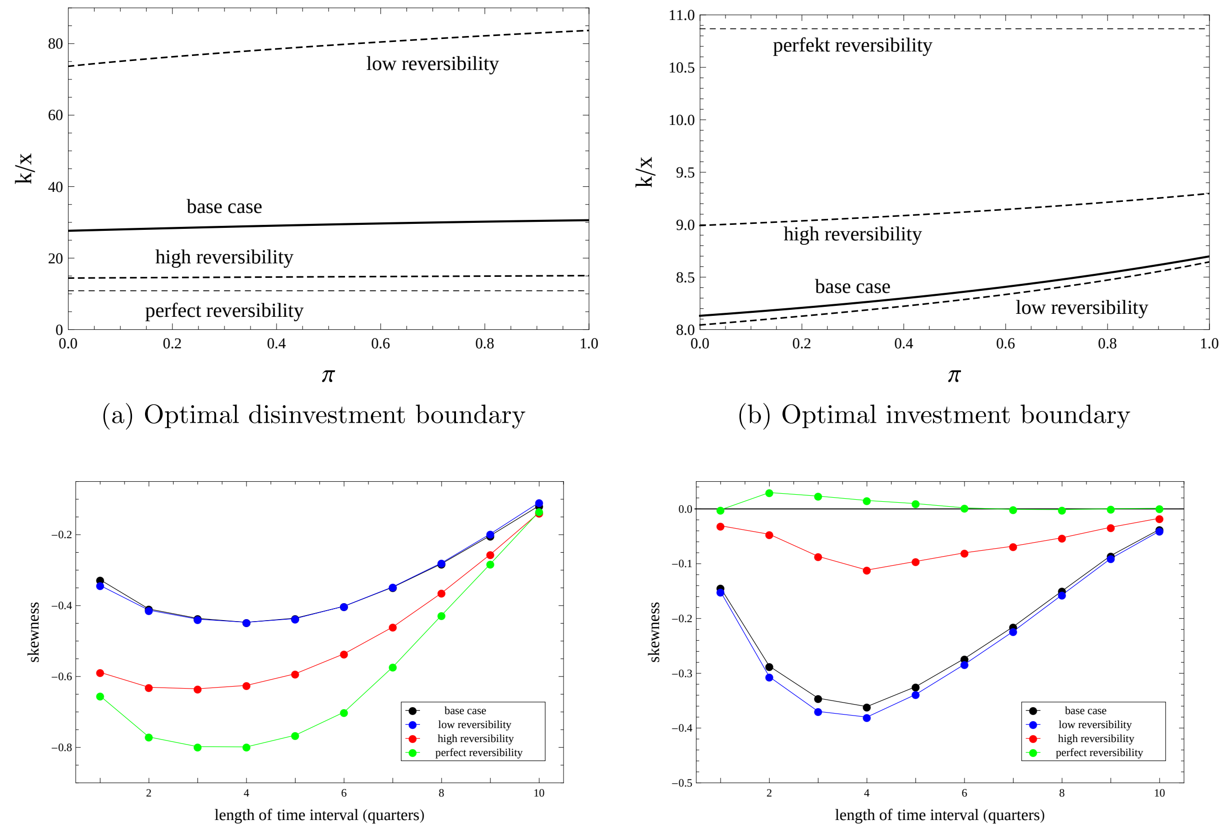
The degree of investment reversibility is determined by the gap between the purchase price and the resale price of capital. Since we fix the purchase price of capital at 1 throughout our study, the specification of the resale price, b, characterizes reversibility, with $b = 1$ defining perfect reversibility and $b = 0$ defining perfect irreversibility.

Perfect Reversibility Case: If investment is perfectly reversible, firms adjust capital stock instantaneously to any level appropriate for the realized demand factor, because both upward and downward adjustments are costless. The belief about future growth rates becomes irrelevant. The optimal capacity is determined by the classical optimality condition of Jorgenson (1963), i.e., the equality of the marginal revenue product of capital ($\frac{\partial f}{\partial K}$) and the user cost of capital, which in our case is simply $r+\xi$. Using equation (5), this optimality condition leads to:

$$\frac{k}{x} = (r+\xi)^{-\frac{1}{\alpha}}. \tag{17}$$

Based on our base case values of r, ξ and α in Table 1, the optimal normalized capacity (k/x) in this case is then 10.8679. Since firms maintain their capacities constantly at this optimal ratio, the capital growth rate is identical to the growth rate of the demand factor. This serves as a good benchmark for our study of the effects of costly reversibility.

Panels (a) and (b) of Figure 9 show the effects of investment reversibility on the disinvestment and investment boundaries, respectively. Not surprisingly, with a high degree of reversibility, firms invest and disinvest actively. Therefore, an increase in b leads to a downward shift of the disinvestment boundary and an upward shift of the investment boundary, leading to a narrow inaction region compared to the case of low reversibility. Furthermore, both boundaries become flatter, suggesting a lower sensitivity of the investment/disinvestment decision on the state of the economy. For the perfect reversibility case ($b = 1$), investment and disinvestment boundaries collapse into one, and the combined boundary is perfectly flat, as it



(a) Optimal disinvestment boundary

(b) Optimal investment boundary

(c) Skewness of average capital growth rates

(d) Skewness of differenced average growth rates

Figure 9: **Comparative statics with respect to the degree of reversibility.** This figure compares the investment behavior of simulated firms under different degrees of investment reversibility. The resale price of capital (b) is 1 in the perfect reversibility case, 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.

is independent of the belief.

With an increased degree of reversibility, firms' disinvest capital actively to reduce excess capital in bad states. As a consequence, investment and disinvestment become more symmetric. Thus the average skewness of individual firms' capital growth rates decreases as the resale price of capital increases (not plotted in the figure).

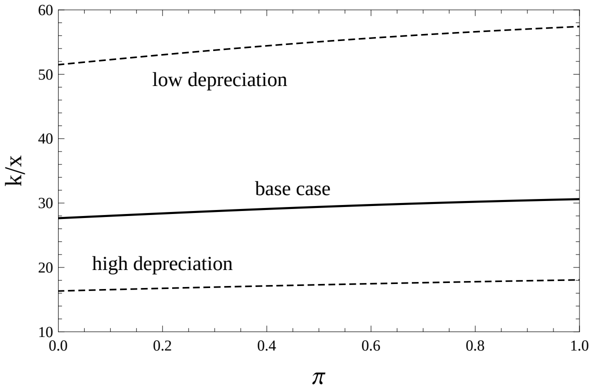
At the aggregate level, two mutually enforcing mechanisms weaken the slope asymmetry as investment reversibility increases. First, since the inaction region between investment and disinvestment boundaries becomes smaller, the heterogeneity among firms with respect to excess capital is lower. When the regime changes from a recession to an expansion, more firms move quickly towards the investment boundary. Second, the investment boundary becomes less steep and less convex, as learning about future growth is less important if disinvestment can be done at low costs. A flatter investment boundary means that firms do not move between substantially different thresholds from time to time. This weakens the effect of both negative and positive shocks. So we predict that a higher degree of reversibility lowers the slope asymmetry of the capital growth rate at the aggregate level. Panel (d) of Figure 9 confirms this intuition. In particular, it shows that the slope asymmetry of the average capital growth rate completely disappears when investment is perfectly reversible. This suggests that costs of reversibility are a necessary condition for the slope asymmetry at the aggregate level in our model.

Interestingly, the level asymmetry of the average capital growth rate becomes more pronounced as investment becomes more reversible (Panel (c)). This confirms our intuition that the negative spikes at the aggregate level are not due to costly reversibility. Instead they arise because expansions generally last longer than recessions, as the relatively smaller number of observations drawn from the recession periods tend to form a long tail of on the left side of the growth rate distribution. Higher reversibility induces firms to reduce capital more aggressively in the recession. This results in troughs even deeper below the mean, and explains why the level asymmetry is more pronounced when reversibility is high.

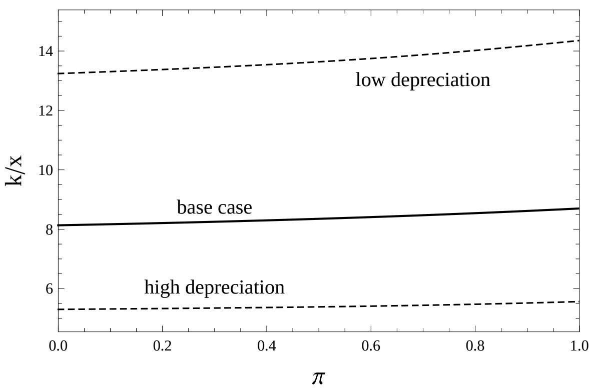
The patterns of capital growth rates in Panels (c) and (d) also show another interesting feature. While a small cost of 2% ($b = 0.98$) is enough to generate substantial deviations from the perfect reversibility case ($b = 0$), the differences between the $b = 0.80$ case and the $b = 0.50$ case are very small. This suggests a small cost of reversibility is enough to capture the major effects of irreversibility.

6.2 Differences in Depreciation Rate

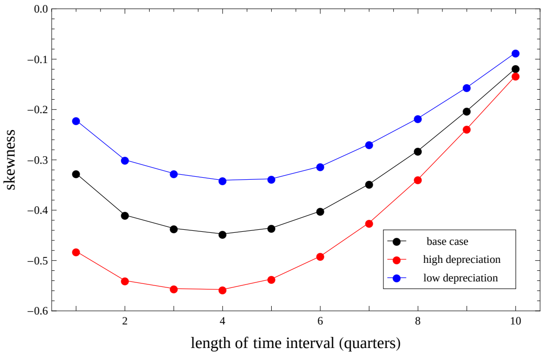
A high depreciation rate has effects similar to those of a high discount rate. It reduces both the maximum tolerated excess capital and the optimal capacity (see panels (a) and (b) of Figure 10). This is because a higher depreciate rate lowers the the marginal value of invested capital. A higher depreciate rate also makes the investment/disinvestment boundaries flatter, indicting



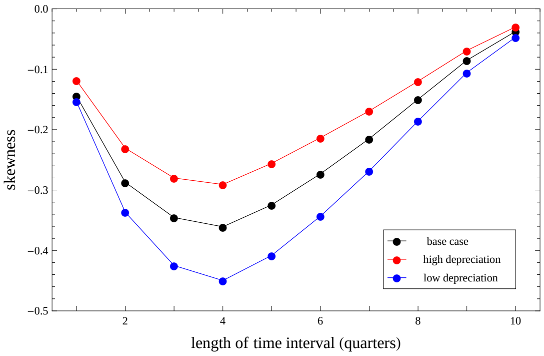
(a) Optimal disinvestment boundary



(b) Optimal investment boundary



(c) Skewness of average capital growth rates



(d) Skewness of differenced average growth rates

Figure 10: **Comparative statics with respect to the depreciation rate.** This figure compares the investment behavior of simulated firms under different depreciation rates. The depreciation rate (ξ) 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.

a lower sensitivity to the belief. This is because a firm's effective planning horizon decreases as the depreciation rate increases, which reduces the importance of expected future growth rate.

Like high reversibility, a high depreciation rate increases firms' ability to adapt to negative demand shocks, since it reduces excess capital at a faster speed. This makes firms' investment and disinvestment more symmetric. Consequently, a higher depreciation rate reduces the positive skewness of capital growth rates at the firm level (not plotted).

At the aggregate level, the effects of a high depreciation rate are also similar to those of high reversibility. It allows firms to reduce excess capital at a faster speed without paying the disinvestment cost. This makes the negative skewness of the average capital growth rate more pronounced (Panel (c)). On the other hand, the reduction of excess capital through high depreciation prevents firms from moving faraway from the investment boundary. So an increase in expected demand growth quickly translates into new investment by a broad range of firms. Hence a higher depreciation rate is associated with a less pronounced slope asymmetry at the aggregate level. Panel (d) of Figure 10 illustrates this effect.

7 Conclusion

Our empirical analysis of firms in the quarterly Compustat-CRSP merged database shows strikingly different patterns of corporate investment at the firm and aggregate levels. The average capital growth rate across firms exhibits negative spikes, despite the positive spikes at the firm level. Furthermore, while individual firms run up to their investment spikes and come down at an equal speed, the average capital growth rate declines much faster than it recovers.

We develop a dynamic model of investment that replicates these empirical patterns. The model features costly reversibility, cyclical macroeconomic shocks, and uncertainty about the true state of the economy. We consider a cross-section of firms facing heterogeneous demand shocks. The expected growth rate of firms' 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 the demand factor, is a convex function of its posterior belief 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.

Our model replicates the features of capital growth rates at both the firm and aggregate levels. For individual firms, the level of capital growth rates is positively skewed, but the slope is symmetric. In the aggregate, both the level and the slope are negatively skewed. The positive skewness of firm-level capital growth rates is a natural outcome of 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 slope asymmetry of the average capital growth rate is due to the endogenous distribution of firms relative to the boundary. More firms are close to the investment boundary at the end of an expansion then at the end of a recession, therefore more firms react to negative signals arriving in good times than to positive signals arriving in bad times. As a result, the average capital growth rate across firms increases gradually during a recovery but drops sharply at the beginning of a recession.

One feature of our model is that it abstracts from 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 further delay firms' investment at the initial stage of an expansion. Incorporating financial frictions over the business cycle 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) 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.

Appendices

A.1 An Alternative Formulation of the Optimal Updating Rule

Equation (9) can be rewritten as:

$$d\pi_t = [-\pi_t\lambda_{h,l} + (1-\pi_t)\lambda_{l,h}]dt + \frac{(\mu_h - \mu_l)\pi_t(1-\pi_t)}{\sigma^2}\mathbf{w}\begin{pmatrix} d\ln(x_t) - E_t(\mu_t|\mathcal{F}_t)dt \\ d\ln(s_t) - E_t(\mu_t|\mathcal{F}_t)dt \end{pmatrix} \quad \text{(A.1)}$$

where

$$\sigma^2 \equiv \frac{1}{\mathbf{1}'(\Phi\Phi')^{-1}\mathbf{1}}, \tag{A.2}$$

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

Note that $\Phi\Phi'$ is simply the instantaneous variance-covariance matrix of $d\ln(x_t)$ and $d\ln(s_t)$. Readers familiar with the classic mean-variance portfolio analysis will immediately recognize that σ^2 is the minimum instantaneous variance that can be obtained using all possible linear combinations of $d\ln(x_t)$ and $d\ln(s_t)$, while $\mathbf{w}$ 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 π_t is adjusted upward. Conversely, when the realized value is lower than expected, π_t is adjusted downward.

The standard deviation of this optimally constructed compound signal, σ, 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 $\mathbf{w}$ 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 $d\ln(x_t)$ in the compound signal is zero. Learning is entirely based on the signal $d\ln(s_t)$. More surprisingly, when $\sigma_s < \rho\sigma_x$, the optimal weight of $d\ln(x_t)$ 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 π_t downward. The intuition is as follows. Since the shocks to the two signals are highly correlated, a higher-than-expected realized value of $d\ln(x_t)$ suggests that it is very likely that $d\ln(s_t)$ 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 $d\ln(s_t)$, 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 π, i.e., $V = V(x, k, \pi)$. For the risk-neutral decision maker, the value function must satisfy the following Hamilton-Jacobi-Bellman equation:

$$rV(x,k,\pi)dt = f(x,k)\ dt + E(dV(x,k,\pi)|\mathcal{F}_t) \tag{A.4}$$

The expectation of dV is to be determined by Itô's lemma using the $\mathcal{F}_t$-dynamics of the state variables x, k, and π. Note that since $dW_{xt}^{\mathcal{F}}$ and $dW_{st}^{\mathcal{F}}$ are uncorrelated, we have

$$\begin{aligned}
dx &= x(\pi\mu_h + (1-\pi)\mu_l + \frac{1}{2}\sigma_x^2)dt + x\sigma_x dW_{xt}^{\mathcal{F}}, \\
(d\pi)^2 &= d\pi(d\pi)' \\
&= [\pi(1-\pi)(\mu_h-\mu_l)]^2 \left[\mathbf{1}'(\Phi')^{-1}\mathbf{dW}_t^{\mathcal{F}}\right]\left[\mathbf{1}'(\Phi')^{-1}\mathbf{dW}_t^{\mathcal{F}}\right]' \\
&= [\pi(1-\pi)(\mu_h-\mu_l)]^2 \frac{1}{\sigma^2}dt, \\
dx\,d\pi &= dx\,(d\pi)' \\
&= x\pi(1-\pi)(\mu_h-\mu_l)\left[(1,0)\,\Phi\mathbf{dW}_t^{\mathcal{F}}\right]\left[\mathbf{1}'(\Phi')^{-1}\mathbf{dW}_t^{\mathcal{F}}\right]' \\
&= x\pi(1-\pi)(\mu_h-\mu_l)dt.
\end{aligned}$$

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

$$\begin{aligned}
E[dV(x,k,\pi)|\mathcal{F}_t] = \Big[&-\frac{\partial V}{\partial k}\xi k + \frac{\partial V}{\partial x}x[\pi\mu_h + (1-\pi)\mu_l + \frac{1}{2}\sigma_x^2] \\
&+\frac{1}{2}\frac{\partial^2 V}{\partial x^2}x^2\sigma_x^2 + \frac{\partial V}{\partial \pi}[-\pi\lambda_{h,l} + (1-\pi)\lambda_{l,h}] \\
&+\frac{1}{2}\frac{\partial^2 V}{\partial \pi^2}[\pi(1-\pi)(\mu_h-\mu_l)]^2\frac{1}{\sigma^2} + \frac{\partial^2 V}{\partial x \partial \pi}x\pi(1-\pi)(\mu_h-\mu_l)\Big]\,dt.
\end{aligned}$$

Substituting this expression into the Hamilton-Jacobi-Bellman equation (A.4) and dropping dt from both sides of the equation yields

$$\begin{aligned} rV &= \frac{1}{1-\alpha}x^{\alpha}k^{(1-\alpha)} - \frac{\partial V}{\partial k}\xi k + x[\pi\mu_h + (1-\pi)\mu_l + \frac{1}{2}\sigma_x^2]\frac{\partial V}{\partial x} + \frac{1}{2}x^2\sigma_x^2\frac{\partial^2 V}{\partial x^2} \\ &+ [-\pi\lambda_{h,l} + (1-\pi)\lambda_{l,h}]\frac{\partial V}{\partial \pi} \\ &+ \frac{[\pi(1-\pi)(\mu_h-\mu_l)]^2}{2\sigma^2}\frac{\partial^2 V}{\partial \pi^2} + x\pi(1-\pi)(\mu_h-\mu_l)\frac{\partial^2 V}{\partial x \partial \pi} \end{aligned} \tag{A.5}$$

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 (A.5):

$$\begin{aligned} \frac{\partial V}{\partial k} &= \frac{\partial v(h,\pi)}{\partial h}, \\ \frac{\partial V}{\partial x} &= v(h,\pi) - h\frac{\partial v(h,\pi)}{\partial h}, \\ \frac{\partial^2 V}{\partial x^2} &= \frac{1}{x}h^2\frac{\partial^2 v(h,\pi)}{\partial h^2}, \\ \frac{\partial V}{\partial \pi} &= x\frac{\partial v(h,\pi)}{\partial \pi}, \\ \frac{\partial^2 V}{\partial \pi^2} &= x\frac{\partial^2 v(h,\pi)}{\partial \pi^2}, \\ \frac{\partial^2 V}{\partial x \partial \pi} &= \frac{\partial v}{\partial \pi} - h\frac{\partial^2 v(h,\pi)}{\partial x \partial \pi}. \end{aligned}$$

A.3 Numerical Optimization of Investment and Disinvestment Boundaries

We solve the Hamilton-Jacobi-Bellman equation (13) together with the free boundary conditions by solving numerically the underlying stochastic dynamic programming problem. We discretize (A.4)

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

and use its homogeneity property in k to write the program in terms of Tobin's average Q

$$V(x,t,\pi) = kV(x/k,1,\pi) = kQ(g,\pi)$$

with $g = x/k$. Since in the inaction region, where the firm neither invests nor disinvests, capital, k, is constant, g_t is just a scaled version of x_t 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 amount of new capital to locate itself at the investment boundary. The argument for the disinvestment region is analogous. We derive

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

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

$$\begin{aligned}
Q(g,\pi) &= \frac{g}{g_i^*(\pi)}Q(g_i^*(\pi),\pi) - \frac{g-g_i^*(\pi)}{g_i^*(\pi)} \qquad g > g_i^*(\pi),\\
Q(g,\pi) &= \frac{g}{g_d^*(\pi)}Q(g_d^*(\pi),\pi) - \frac{g-g_d^*(\pi)}{g_d^*(\pi)}b \qquad g < g_d^*(\pi).
\end{aligned}$$

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

$$Z(\pi) = \int_{\frac{1}{2}}^{\pi} \frac{\sigma}{(\mu_h-\mu_l)p(1-p)}dp = \frac{\sigma}{\mu_h-\mu_l}\ln\left(\frac{\pi}{1-\pi}\right), \tag{A.6}$$

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

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

$$\begin{aligned} dz &= \underbrace{\left[\mu_\pi \frac{dZ(\pi)}{d\pi} + \frac{1}{2}\sigma_\pi^2 \frac{d^2 Z(\pi)}{d\pi^2}\right]}_{\mu_z} dt + \underbrace{\sigma \mathbf{1}'(\Phi')^{-1}\mathbf{dW_t^{\mathcal{F}}}}_{dW_{zt}^{\mathcal{F}}} \\ z_0 &= Z(\pi_0) \end{aligned} \tag{A.7}$$

where σ_π is the volatility and μ_π is the drift of the belief (see Proposition 1) given by

$$\begin{aligned} \sigma_\pi &= \frac{(\mu_h - \mu_l)\pi(1-\pi)}{\sigma} \\ \mu_\pi &= -\pi\lambda_{hl} + (1-\pi)\lambda_{lh} \end{aligned} \tag{A.8}$$

and $W_{zt}^{\mathcal{F}}$ 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), taking into account the correlation of the two processes that is implicitly given. Then, the belief process π results from applying the inverse, $\pi_t = Z^{-1}(z_t)$, which is given by

$$Z^{-1}(z) = \frac{1}{1 + \exp\left\{-\left(\frac{\mu_h - \mu_l}{\sigma}\right) z\right\}}. \tag{A.9}$$

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

$$\begin{aligned} h_i^* &= \frac{1}{g_i^*}, \\ h_d^* &= \frac{1}{g_d^*}. \end{aligned}$$

References

Abel, Andrew B., and Janice C. Eberly, 1996, Optimal Investment with Costly Reversibility, *Review of Economic Studies* 63, 581–593.

Acemoglu, Daron, and Andrew Scott, 2003, Asymmetric Business Cycles: Theory and Time-Series Evidence, *Journal of Monetary Economics* 40, 501–533.

Alti, Aydogan, 2003, How Sensitive Is Investment to Cash Flow When Financing Is Frictionless?, *Journal of Finance* 58, 707–722.

Bernanke, Ben, 1983, Irreversibility, Uncertainty, and Cyclical Investment, *Quarterly Journal of Economics* 8, 85–106.

Bertola, Guiseppe, and Ricardo J. Caballero, 1994, Irreversibility and Aggregate Investment, *Review of Economic Studies* 61, 223–246.

Bloom, Nick, 2009, The Impact of Uncertainty Shocks, *Econometrica* 77, 623–685.

Bloom, Nick, Stephen Bond, and John Van Reenen, 2007, Uncertainty and Investment Dynamics, *Review of Economic Studies* 74, 391–415.

Boyle, Phelim P., Jeremy Evnine, and Steven Gibbs, 1989, Numerical Evaluation of Multivariate Contingent Claims, *Review of Financial Studies* 2, 241–250.

Chalkley, Martin, and In Ho Lee, 1998, Learning and Asymmetric Business Cycles, *Review of Economic Dynamics* 1, 623–645.

Chamley, Christophe, and Douglas Gale, 1994, Information Revelation and Strategic Delay in a Model of Investment, *Econometrica* 62, 1065–1085.

David, Alexander, 1997, Fluctuating Confidence in Stock Markets: Implications for Returns and Volatility, *Journal of Financial and Quantitative Analysis* 32, 427–462.

Decamps, Jean-Paul, Thomas Mariotti, and Stephane Villeneuve, 2005, Investment Timing Under Incomplete Information, *Mathematics of Operations Research* 20, 472–500.

DeLong, Bradford, and Lawrence Summers, 1986, Are Business Cycles Symmetric?, in *American Business Cycle: Continuity and Change*, 166-79, edited by Robert Gordon. (University of Chicago Press: Chicago).

Dempster, Arthur, Nan Laird, and Donald Rubin, 1977, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society* Series B 39, 1–38.

Dixit, Avinash, and Robert Pindyck, 1994, *Investment Under Uncertainty.* (Princeton University Press, Princeton, New Jersey).

Dumas, Bernard, 2001, Super Contact and Related Optimality Conditions, *Journal of Economic Dynamics and Control* 15, 675–685.

Falk, Barry, 1986, Further Evidence on the Asymmetric Behavior of Economic Time Series over the Business Cycle, *Journal of Political Economy* 94, 1096–1109.

Gale, Douglas, 1996, Delay and Cycles, *Review of Economic Studies* 63, 169–198.

Grenadier, Steven R., and Andrey Malenko, 2010, A Bayesian Approach to Real Options: The Case of Distinguishing Between Temporary and Permanent Shocks, *Journal of Finance* 65, 1949–1986.

Guo, Xin, Jianjun Miao, and Erwan Morellec, 2005, Irreversible Investment with Regime Shifts, *Journal of Economic Theory* 132, 37–59.

Hamilton, James, 1989, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* 57, 357–384.

Jorgenson, Dale W., 1963, Capital Theory and Investment Behavior, *American Economic Review Papers and Proceedings* 53, 247–259.

Keynes, John M., 1936, *The General Theory of Employment, Interest, and Money.* (Macmillan, London).

Klein, Manuel, 2009, Comment on Investment Timing Under Incomplete Information, *Mathematics of Operations Research* 35, 249–254.

Liptser, Robert S., and Albert N. Shiryaev, 2001, *Statistics of Random Processes I, General Theory.* (Springer Berlin).

McDonald, Robert, and Daniel Siegel, 1986, The Value of Waiting to Invest, *Quarterly Journal of Economics* 101, 707–728.

Morellec, Erwan, 2001, Asset Liquidity, Capital Structure and Secured Debt, *Journal of Financial Economics* 61, 173–206.

Neftci, Salih N., 1985, Are Economic Time Series Asymmetric over the Business Cycle?, *Journal of Political Economy* 92, 307–328.

Nelson, Daniel B., and Krishna Ramaswamy, 1990, Simple Binomial Processes as Diffusion Approximations in Financial Models, *Review of Financial Studies* 3, 393–430.

Nieuwerburgh, Stijn Van, and Laura Veldkamp, 2006, Learning Asymmetries in Real Business Cycles, *Journal of Monetary Economics* 53, 753–772.

Pindyck, Robert, 1988, Irreversible Investment, Capacity Choice, and the Value of the Firm, *American Economic Review* 78, 969–985.

Sichel, Daniel E., 1993, Business Cycle Asymmetry: A Deeper Look, *Economic Inquiry* 31, 224–236.

Stein, Jeremy, 2003, Agency, Information, and Corporate Investment, in *Handbook of the Economics of Finance*, Volume 1, 111-165. Edited by George Constantinides, Milton Harris, and Rene Stulz. (Elsevier Science, Amsterdam).

Veldkamp, Laura, 2005, Slow Boom, Sudden Crash, *Journal of Economic Theory* 124, 230–257.

Veronesi, Pietro, 1999, Stock Market Overreaction on Bad News in Good Times: A Rational Expectations Equilibrium Model, *Review of Financial Studies* 12, 975–1007.

Veronesi, Pietro, 2000, How does Information Quality Affect Stock Returns?, *Journal of Finance* 55, 807–837.