Equilibrium Value and Profitability Premiums

Hengjie Ai, Jun E. Li and Jincheng Tong *

October 17, 2021

Abstract Standard production-based asset pricing models cannot simultaneously explain the value and the profitability premia, because the value and the profitability factors are highly negatively correlated. Empirically, we show that value and profitability sorted portfolios differ in the persistence of productivity. We develop a general equilibrium model where firm-level productivity has a two factor structure with different persistence and demonstrate that heterogeneity in the persistence of productivity shocks can account for the coexistence of the profitability and the value premium.

JEL Code: E2, E3, G12

*Hengjie Ai (hengjie@umn.edu) is affiliated with the Carlson School of Management, University of Minnesota, Jun E. Li (junli@saif.sjtu.edu.cn) is affiliated with Shanghai Advanced Institute of Finance, Shanghai Jiao Tong University, and Jincheng Tong (jincheng.tong@rotman.utoronto.ca) is at the University of Toronto.

1 Introduction

Historically, firms with higher profitability ratios, defined as the ratio of gross profit to total assets, tend to earn higher average returns, which is often referred to as the profitability premium (Novy-Marx (2013)). At the same time, stocks with high book-to-market ratios, higher ratios of the book value of assets relative to their market value, or value firms, earn a higher average return than those with low book-to-market ratios (growth firms).

The coexistence of the value and profitability premia presents a challenge to the literature on production-based asset pricing models, for example, Zhang (2005). In these models, productivity is typically specified as an AR(1) process and is the only factor that determines the cross-section heterogeneity in firms' book-to-market ratios and profitability ratios. High productivity firms are high profitability ratio firms but are simultaneously also growth firms: they have a higher market value of assets relative to their book value and, therefore lower book-to-market ratios. As a result, traditional production-based asset pricing models cannot separate the value factor and the profitability factor. By construction, a profitability premium must imply a value discount. That is, these models can either generate a value or a profitability premium, but not both. In this paper, we develop a general equilibrium model where firm-level productivity has a two-factor structure that differs in persistence. We show that our model can not only distinguish the profitability factor from the value factor but also account for the co-existence of the value and profitability premium.

To distinguish profitability from value empirically, we first provide suggestive evidence for the two-factor structure in firm productivity. We show that consistent with the standard production-based asset pricing literature, both high profitability firms and growth firms have higher investment rates, higher cash flow growth rates, and faster sales growth rates at the time of portfolio formation when compared to low profitability and value firms, respectively. However, several differences between the profitability and value sorted portfolios suggest that book-to-market sorting differentiates firms in terms of the permanent component of productivity shocks, while portfolios sorted on profitability differ in terms of the transitory component of their productivity.

First, the difference in investment rate, cash flow growth rate, and sales growth rate is significantly higher for book-to-market sorted portfolios than profitability-sorted portfolios. For example, the average investment rate is about 7% for firms in the value portfolio, and more than twice as large: about 15% for the growth portfolio. In contrast, the average investment rate of low and high profitability sorted portfolios are 8% and 12%, respectively.

Second, the difference in investment rates, cash flow growth rates, and sales growth rates is much more persistent for book-to-market sorted portfolios than that for profitability sorted portfolios. Empirically, we show that the spread in investment rates, cash flow growth rates, and sales growth rates for book-to-market sorted portfolios remain significant five years after portfolio formation, while the difference in the above measures vanishes one year after portfolio formation for gross profitability sorted portfolios.

In addition, we formally estimate a two-factor model of firm cash flow and decompose cash flow at the firm level into a permanent and a transitory component. We demonstrate that book-to-market sorted portfolios differ more significantly in terms of the permanent component of cash flow, while gross profitability sorted portfolios have a higher spread in the transitory component of cash flow and a lower spread in the permanent component by comparison.

Motivated by the above empirical evidence, we develop a general equilibrium model where firm productivity has a permanent and transitory component. The transitory component has a large standard deviation in one-period innovations but a very small persistence. In contrast, the permanent component has a small conditional standard deviation but follows a random walk. This specification allows our model to distinguish the profitability factor from the value factor. A positive shock to the transitory component of productivity raises the current-period profitability of the firm but, due to its lack of persistence, has a quantitatively small impact on the book-to-market ratio. A positive shock to the permanent component of productivity, on the other hand, has a small impact on the current-period profitability as the size of the shock is typically small but significantly affects the market value of the firm, as the shock is expected to last into the future. As a result, from a quantitative perspective, consistent with the data, book-to-market sorting mainly differentiates firms along the permanent component of productivity shocks, while gross profitability sorting selects firms based mainly on the transitory component of their productivity.

We incorporate two sources of aggregate shocks into our model, total factor productivity shocks and shocks to the marginal cost of capital utilization. Our model generates a value premium as in standard adjustment cost based models. In our model, firms with low permanent components of the productivity shocks are value firms. The fixed operating cost and capital adjustment cost provide a form of operating leverage, making these firms riskier relative to growth firms, where adjustment cost is a smaller fraction of firm cash flow.

The key for our model to account for the profitability premium is variable capital utilization: firms can choose a higher rate of capital utilization at the expense of faster depreciation. In the presence of capital utilization costs, high profitability firms and growth firms respond very differently. High profitability firms, knowing that shocks are transitory, respond by increasing the rate of capital utilization and not investment. In contrast, growth firms are those who experience positive shocks in the permanent component of productivity.

Anticipating that the impact of these shocks will persist into the future, they respond by increasing investment and lowering capital utilization. In our model, the cost of capital utilization is a form of operating leverage and makes high profitability firms and value firms riskier.

In summary, in our model, in the cross-section, value and profitability sorting are determined by the permanent and transitory component of idiosyncratic productivity shocks. At the aggregate level, due to firms' optimal choices of investment and capital utilization, value firms are endogenously more exposed to aggregate productivity shocks, while high profitability firms are more sensitive to shocks to the marginal cost of capital utilization.

Our model is set in general equilibrium, while standard production-based asset pricing models of the cross-section of expected returns typically assume an exogenous pricing kernel. In general equilibrium setups, the cross-section distribution of firm types becomes a relevant state variable that determines the properties of the pricing kernel. Because endogenous investment decisions depend not only on productivity shocks but also on the property of the pricing kernel, partial equilibrium models cannot account for this equilibrium feedback mechanism. We develop an efficient method to simultaneously solve the evolution of firms' distribution and the equilibrium asset prices.

We calibrate our model to match aggregate moments and examine its ability to explain the cross-section. Our model matches well standard macroeconomic quantity dynamics. It also delivers realistic financial moments, including a high equity premium and a low and smooth risk-free interest rate. We report three key quantitative results in the cross-section of firms. First, the model generates a realistic distinction in firm characteristics for value/growth and high/low profitability sorted portfolios. The heterogeneity in the permanent component of firm productivity leads to a significant spread in book-to-market ratios. Shocks to the transitory component of firm productivity translate into differences in profitability. As a result, as in the data, book-to-market sorted portfolios exhibit large dispersion in investment rates, and profitability sorted portfolios do not. In addition, our model also captures well the heterogeneity in the persistence of cash-flow growth and investment growth for both value/growth and profitability sorted firms.

Second, variable capital utilization provides a resolution for the profitability premium. Utilizing capital more intensively raises a firm's current profit, leading to faster capital depreciation and incurring additional maintenance costs. Under the optimal policy, firms with higher transitory productivity shock tend to choose higher utilization rates than firms that experienced adverse transitory shocks. Faster capital depreciation due to higher utilization exposes profitable firms more to the aggregate utilization cost shock. As a result, high profitability firms are riskier and require a higher expected return in equilibrium. We show that this utilization channel gives rise to a significant profitability premium in our calibration.

Third, our equilibrium model also delivers a realistic value premium, and two economic mechanisms contribute to it. Aggregate investment is procyclical, and the growth firms' investment exhibits more procyclicality than value firms. Intuitively, growth firms are the high marginal product of capital firms because of the high level of permanent productivity. Efficient capital allocation requires that growth firms account for a larger fraction of aggregate investment in good times, and therefore, their investment is more procyclical. Dividends roughly equal to profits minus investment; because profits are smooth and growth firms' profits are small relative to their investment, procyclical investments translate into counter-cyclical dividends for growth firms, making growth firms good hedges for aggregate productivity shock. In addition, value firms tend to be firms with declining permanent productivity. They shrink in size via a combination of disinvestment and faster capital depreciation due to higher utilization rates. Higher utilization rates make value firms' cash flow more sensitive to aggregate utilization cost shock.

We develop an efficient approach to solve our model with heterogeneous firms. The entire distribution of firms is a relevant state variable, and firms need to forecast the distribution dynamics to set their policies optimally. Different from the standard Krusell and Smith (1998) method, which approximates firm distribution with key moments of it, we discretize the distribution into several grid points and keep track of the probability that firms fall into each grid. Furthermore, firms' transition dynamics between these grid points are consistent with firms' optimal policies. In the end, the equilibrium conditions and the law of motion for firm distribution are characterized by a set of non-linear equations that we solve efficiently with the perturbation method.

Literature review

This paper that builds on a growing literature studying the profitability premium such as Dou et al. (2020), Kogan et al. (2020), Bouchaud et al. (2019), Ma and Yan (2018) and Deng (2020). Our paper is closely related with Dou et al. (2020), Kogan et al. (2020) and Ma and Yan (2018) because we provide a joint explanation for value and profitability premium at the same time. Different from these works, we focus on optimizing firms whose investment decisions are derived endogenously from their optimality conditions. This allows us to generate a key distinction between value and profitability strategies: profitable firms' investments are more transient than those made by growth firms.

Our paper is also related to the literature that emphasized the importance of permanent shocks at the firm level (see Gabaix (1999), Luttmer (2007) and Miao (2005)). Luttmer

(2007) shows that permanent productivity shock results in a power law distribution of firm size that is consistent with the U.S. firm size distribution. Gourio (2008) highlights the role of permanent shock to explain firm-level investment dynamics. We demonstrate the importance of permanent shock from a cross-sectional asset pricing perspective: allowing for permanent shocks captures the key difference between value (growth) firms and low (high) profitability firms.

This paper further belongs to the literature quantitatively explaining the cross-section of stock returns, in particular the production-based asset pricing literature.[1] Zhang (2005) provides an investment-based explanation for the value premium. Belo et al. (2017) investigates the relations between hiring decisions, cross-section of stock returns and the value premijm. Kogan and Papanikolaou (2014) explores whether investment-specific shock can explain stock market anomalies. Grigoris and Segal (2020) examines the relation between capital utilization and cross-section of stock returns. Hasler et al. (2020) shows that rational learning yields an downward-sloping term structure of market risk premium which helps understand the value premium. Gomes and Schmid (2021) examines the relations between financial leverage and the value premium.

Finally, this paper further belongs to the production-based asset pricing in general equilibrium model. From a methodological point of view, our general equilibrium model allows for a cross section of firms with heterogeneous productivity and is related to previous work including Gomes et al. (2003), Gârleanu et al. (2012), Ai and Kiku (2013), Favilukis and Lin (2015), Chen (2018), Bai et al. (2019) and Tong and Ying (2020). Compared with these papers, our model jointly explains value and profitability premium in general equilibrium. Our numerical approach is novel that the equilibrium in our model can be solved for efficiently using perturbation method.

2 Empirical Facts

In this section, we provide a set of empirical findings to further distinguish growth (value) and high (low) profitability firms. Empirical patterns in firms' fundamentals suggest that growth and profitable firms' investment and cash flow choices are very likely driven different types of firm-level shocks.

[1] Kogan and Papanikolaou (2012) provide an excellent survey for production-based asset pricing.

2.1 Data

We obtain firm-level balance sheet data from CRSP/Compustat Merged database and monthly stock returns from CRSP. We adopt the standard screening process for the CRSP/Compustat Merged Database. We exclude utilities and financial firms (SIC codes between 4900 and 4999 and between 6000 and 6999, respectively). Additionally, we only keep common stocks that are traded on NYSE, AMEX, and NASDAQ. The potential delisting bias of stock returns is corrected following Shumway (1997) and Shumway and Warther (1999). Our sample starts from January 1963 and ends in December 2020.

Following Novy-Marx (2013), we define gross profitability as revenue minus cost of goods sold divided by total asset, that is, (REVT-COGS)/AT in Compustat items. We define book-to-market ratio following Fama and French (2015). More detailed construction can be found in Table 1.

2.2 Growth perspectives of growth and profitable firms

Empirically, high gross profitability and growth firms share similar characteristics: they both have lower book-to-market ratios and higher investment rates. We present a set of empirical evidence that further distinguishes growth firms from high profitability firms. We show that they primarily differ in terms of their growth prospects. Growth firms are more exposed to permanent shocks, hence their capital and cash flow grow faster than value firms. On the other hand, high profitability firms are more exposed to transitory shocks, thus their high profitability is rather transient and they do not enjoy much higher growth than less profitable firms.

In order to compare growth and profitable firms, we sort firms into quintile portfolios based on their book-to-market ratios (BM) and gross profitability (GP/A), respectively. Panel A of Table 1 reports characteristics of the BM sorted portfolios. The results are consistent with the literature that growth firms have high gross profitability, higher investment rates (I/K), and lower equity returns than value firms. More importantly, growth firms enjoy much higher growth rates than value firms. We show two types of growth rates: one is based on capital, such as investment and capital growth rates; the other captures the growth rate of profits, such as the sales and cash flow growth rates.

Panel B of Table 1 reports the characteristics of GP/A sorted portfolios. It shows that firms with high gross profitability have lower book-to-market ratios and higher equity returns. These patterns are consistent with the literature, such as Novy-Marx (2013). However, if we focus on the growth perspective of high gross profitability firms, such as investment rates, capital, sales and, cash flow growth rates, in that case, their growth rates are very close to

Table 1: Firm characteristics

This table compares the median firm characteristics of portfolios sorted by book-to-market (BM) and gross profitability (GP/A). Panel A and B present the firm characteristics of portfolios sorted by BM and GP/A, respectively. The firm characteristics include investment rates (I/K), capital (PPEGT) growth rate (ΔK), sales growth rate (ΔSales), and cash flow (EBITDA) growth rate (ΔCF). We also report the annualized excess returns of value weighted portfolios $E[R^e]$. The sample period ranges from 1963 to 2020.

Panel A: Book-to-market sorted portfolios

	Value	2	3	4	Growth	Growth-Value
BM	1.53	0.89	0.61	0.39	0.18	-1.35
GP/A	0.17	0.19	0.26	0.32	0.36	0.19
I/K	0.07	0.09	0.10	0.12	0.15	0.08
ΔK	0.01	0.03	0.05	0.07	0.10	0.09
ΔSales	-0.01	0.03	0.05	0.08	0.12	0.13
ΔCF	-0.02	0.03	0.07	0.10	0.14	0.17
$E[R^e](\%)$	10.11	7.71	7.49	7.02	6.86	-3.26

Panel B: Gross profitability sorted portfolios

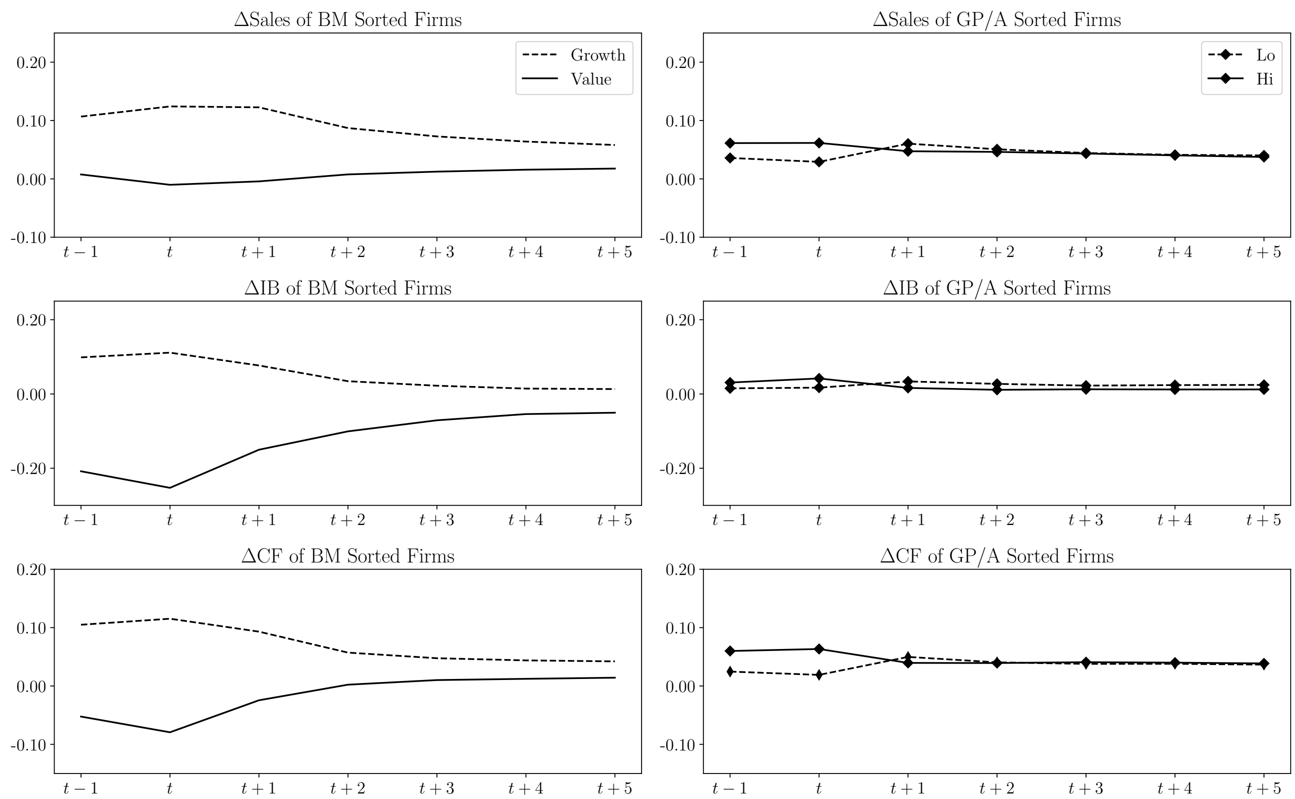
	Low	2	3	4	High	High-Low
BM	0.75	0.79	0.67	0.55	0.43	-0.32
GP/A	0.03	0.12	0.26	0.41	0.66	0.63
I/K	0.08	0.09	0.10	0.11	0.12	0.04
ΔK	0.03	0.04	0.05	0.05	0.06	0.03
ΔSales	0.03	0.04	0.05	0.06	0.06	0.03
ΔCF	0.04	0.05	0.06	0.08	0.09	0.05
$E[R^e](\%)$	4.70	6.33	5.44	7.28	8.02	3.32

the low profitability firms.

Persistence of growth perspectives In order to further distinguish growth and profitable firms, we show their time series median of firm characteristics after their corresponding portfolio formation year, as in Figure 1. The left and right panel shows the results of BM and portfolio sorted portfolios, respectively.

Figure 1: Firm characteristics around portfolio formation

This table reports the time series median of firm characteristics around portfolio formation year t. The left and right panel report the firm characteristics around portfolio formation year of BM and GP/A sorted portfolios, respectively. The definition of book-to-market and profitability follows Fama and French (2015) and Novy-Marx (2013), respectively. The sample starts in 1963 and ends in 2020, at annual frequency.



First, as shown in the left panel of Figure 1, the growth perspectives of growth firms are significantly and persistently outperforming value firms around the portfolio formation year. More importantly, the gap between growth and value firms is large. This evidence suggests that growth firms persistently have much higher growth perspectives than value firms around the portfolio formation year.

Second, the right panel of Figure 1 suggests that the firms with the highest gross profitability do not enjoy much higher growth rates than ones with the lowest gross profitability. More importantly, the high growth rates of the most profitable firms are rather transitory. For the growth rates of sales, income before extraordinary items, and cash flow, the firms with the highest gross profitability only outperform those with the lowest gross profitability at the portfolio formation year. After the portfolio formation year, the growth rates of the least profitable firms are even persistently higher than the most profitable firms.

The empirical evidence suggests that high growth rates of growth firms relative to value firms are very persistent. In contrast, the higher growth rates of more profitable firms are rather transitory. Additionally, the gap between high and low profitability firms is much smaller compared to the gap between growth and value firms.

Overall, growth firms do persistently have significantly higher growth rates than value firms, but high gross profitability firms' growth rates do not outperform low gross profitability firms significantly. These empirical facts motivate us to connect the different growth rates their exposure to permanent and transitory shocks.

2.3 Unobserved components decomposition of cash flow

This section decomposes firm-level cash flow into two components: permanent and transitory. Our decomposition results show that growth firms' cash flows are more exposed to the permanent component, whereas the high gross profitability firms' cash flows are more exposed to the transitory component.

The unobserved component approach, introduced by Harvey (1985), decomposes a time series into a trend and a transitory component. We assume the unobserved component representation takes the following form

$$y_t = x_t + z_t, \tag{1}$$

$$x_t = \mu + x_{t-1} + \varepsilon_t, \tag{2}$$

$$z_t = \rho_z z_{t-1} + \eta_t \tag{3}$$

where y_t is the log of cash flow (EBITDA), x_t is the unobserved trend component, assumed to be a random walk with average growth rate μ, and z_t is the unobserved stationary component. Equation (3) assumes the stationary component takes an AR(1) process, which makes our empirical analysis quantitatively comparable to our general equilibrium model described in Section 3. Detailed procedure of this decomposition can be found in Appendix A.

Table 2 presents the decomposition results. We normalize the trend component by physical capital (PPEGT) because it is growing over time. As shown in Panel A of Table 2, the spread of the trend component is 0.39 for book-to-market sorted portfolios, and it is 20% larger than the gross profitability sorted portfolios, as in Panel B. On the other hand, the spread for transitory components of the book-to-market sorted portfolios is less than 50% of the gross profitability sorted portfolios.

Table 2: Trend and transitory component

This table shows the trend and transitory components of cash flow. We decompose firm-level cash flow, log(EBITDA), into a trend and a transitory component, using the unobserved components method. Trend/K denotes the trend component normalized by physical capital (PPEGT), and Trans represents the transitory component of cash flow.

Panel A: Book-to-market sorted portfolios

	Value	2	3	4	Growth	Growth-Value
BM	1.53	0.89	0.61	0.39	0.18	-1.35
Trend/K	0.16	0.21	0.28	0.38	0.55	0.39
Trans	0.13	0.58	0.85	1.01	0.46	0.34

Panel B: Gross profitability sorted portfolios

	Low	2	3	4	High	High-Low
GP/A	0.03	0.12	0.26	0.41	0.66	0.63
Trend/K	0.13	0.15	0.24	0.36	0.45	0.32
Trans	0.08	0.65	0.60	0.90	0.87	0.80

3 Model setup

We begin by constructing a general equilibrium model with a representative household and heterogeneous firms in this section. Motivated by the empirical evidences that value/growth and high/low profitability firms differ in the persistence of cash-flow, We assume that firms in the model economy are subject to both transitory and permanent shocks to firm-level productivity. These characteristics enable us to investigate the dynamics of the value and profitability premiums within the context of a unified general equilibrium model.

3.1 Households

Time is infinite and discrete. We assume that the representative household has a recursive preference with risk aversion γ and intertemporal elasticity of substitution (IES) ψ as in Epstein and Zin (1989) :

$$U_t = \left\{ (1-\beta) C_t^{1-\frac{1}{\psi}} + \beta \left(E_t \left[U_{t+1}^{1-\gamma} \right] \right)^{\frac{1-\frac{1}{\psi}}{1-\gamma}} \right\}^{\frac{1}{1-\frac{1}{\psi}}}, \tag{4}$$

where U_t is time-t utility and C_t is time-t consumption. The household receives both labor and capital incomes. They trade equities and risk-free bond. In every period t, the household purchases B_t units of risk-free bond and $\omega_{j,t}$ shares of equity for firm j. Investment in risk-free bond delivers a return identical to risk-free interest rate $R_{f,t}$. The payoff of equity consists of capital income and dividend payment $V_{j,t+1} + D_{j,t+1}$ in the next period. The labor market is frictionless and perfectly competitive. Labor income from firm i is therefore the product of competitive wage rate W_t and hours spent at that firm $L_{j,t}$. Household supplies labor inelastically. The household budget constraint at time t can be written as

$$C_t + \int \omega_{j,t} V_{j,t} dj + B_t = W_t \int L_{j,t} dj + R_{f,t-1} B_{t-1} + \int \omega_{j,t-1} \left(V_{j,t} + D_{j,t} \right) dj \tag{5}$$

The pricing kernel is given by the household's intertemporal marginal rate of substitution:

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} \left(\frac{U_{t+1}^{1-\gamma}}{E_t \left[U_{t+1}^{1-\gamma} \right]} \right)^{\frac{1/\psi - \gamma}{1-\gamma}} \tag{6}$$

3.2 Technology

There is a continuum of incumbent firms indexed by $j \in [0,1]$. Incumbent firms produce output $Y_{j,t}$ at time t with capital $K_{j,t}$ and labor $L_{j,t}$ as two production factors. The production function follows:

$$Y_{j,t} = A_t \left[X_{j,t}^{1-\nu} Z_{j,t}^{1-v} (u_{j,t} K_{j,t})^{\nu} \right]^{\alpha} L_{j,t}^{1-\alpha} \tag{7}$$

Here, α denotes the capital share, and ν is the span of control parameter, as in Atkeson and Kehoe (2005). $u_{j,t}$ is the utilization rate of physical capital. The growth rate of aggregate productivity contains a deterministic component μ_A, a persistent component ζ_t, and a Gaussian component ε_t^A:

$$\frac{A_{t+1}}{A_t} = e^{\mu_A+\zeta_t+\varepsilon_{t+1}^A}$$
$$\zeta_t = \rho_\zeta \zeta_{t-1} + \varepsilon_t^\zeta \tag{8}$$

in which both ε_t^A and ε_t^ζ follow normal distributions with zero mean and standard deviations are denoted as σ_A, σ_ζ respectively. The persistent component ζ_t introduces long run risk in the spirit of Bansal and Yaron (2004) into our production economy.

Firm-level productivity contains a permanent component $X_{j,t}$ and a transitory component $Z_{j,t}$. The growth rate of permanent component is determined by an i.i.d normal shock $\varepsilon_{x,t+1}^j$ with mean μ_x and variance σ_x^2:

$$\frac{X_{j,t+1}}{X_{j,t}} = e^{\epsilon_{x,t+1}^j} \tag{9}$$

The transitory component $Z_{j,t}$ follows a discrete time Markov Chain with transition probability matrix P_Z.

Firm j's profit at time t, $\pi(A_t, Z_{j,t}, X_{j,t}, K_{j,t})$, is given as

$$\begin{aligned}\pi\left(A_t, Z_{j,t}, X_{j,t}, K_{j,t}\right) &= \max_{L_{j,t}} Y_{j,t} - W_t L_{j,t} \\ &= \max_{L_{j,t}} A_t \left[X_{j,t}^{1-\nu} Z_{j,t}^{1-\nu} (u_{j,t}K_{j,t})^\nu\right]^\alpha L_{j,t}^{1-\alpha} - W_t L_{j,t}\end{aligned} \tag{10}$$

It is convenient to write the profit function explicitly by maximizing out labor in equation 10 and using the labor market clearing condition $\int L_{j,t}dj = 1$ and obtain an expression for optimal labor input at firm j:

$$L_{j,t} = \frac{X_{j,t}^{1-\nu} Z_{j,t}^{1-\nu} (u_{j,t}K_{j,t})^\nu}{\int X_{i,t}^{1-v} Z_{i,t}^{1-\nu} (u_{i,t}K_{i,t})^\nu di} \tag{11}$$

so that firm j's operating profit function becomes

$$\pi(A_t, Z_{j,t}, X_{j,t}, K_{j,t}) = \alpha A_t \left[X_{j,t}^{1-\nu} Z_{j,t}^{1-\nu} (u_{j,t}K_{j,t})^\nu\right]^\alpha \left[\int X_{i,t}^{1-v} Z_{i,t}^{1-\nu} (u_{i,t}K_{i,t})^\nu di\right]^{\alpha-1} \tag{12}$$

3.3 Capital utilization

Let $I_{j,t}$ denote firm j's investment at time t. Capital accumulates as follows:

$$K_{j,t+1} = \big(1 - \delta(u_{j,t}, \theta_{t+1})\big) K_{j,t} + I_{j,t} \tag{13}$$

in which $\delta(u_{j,t}, \theta_{t+1})$ is the capital depreciation rate that depends on capital utilization rate $u_{j,t}$ as well as a random shock θ_{t+1}. In particular, we assume the following functional form for depreciation:

$$\delta\left(u_{j,t}, \theta_{t+1}\right) = \delta_k + \theta_{t+1} \frac{u_{j,t}^{1+\lambda} - 1}{1+\lambda} \tag{14}$$

where λ is the curvature parameter for the depreciation rate function and δ_k determines the unconditional depreciation rate at the deterministic steady state. And the random process θ_{t+1} follows an mean-reverting process with mean $\bar{\theta}$ and persistence ρ_θ. Innovations to such a process, $\varepsilon_{\theta,t}$ is a normal random variable with mean zero and standard deviation σ_θ:

$$\log\left(\theta_{t+1}\right) - \overline{\theta} = \rho_\delta \left(\log\left(\theta_t\right) - \log\left(\bar{\theta}\right)\right) + \varepsilon_{\theta,t+1} \tag{15}$$

$\delta(u, \theta)$ function is increasing and convex with respect to u. This captures the idea that more intensive use of capital accelerate the depreciation of it. Firms choose capital utilization rate optimally with all information available up to time t. But the actual amount of depreciated capital is unknown until θ_{t+1} is realized at time $t+1$.

The effect of a depreciation shock on the capital depreciation of a firm highly depends on its utilization rate. For firms that choose utilization rates greater than the steady state value 1, a positive surprise in *theta* accelerates capital depreciation. The same shock, on the other hand, has the opposite effect on firms with utilization rates below one: it preserves capital capacity for future production, rather than rapidly depreciating it. This heterogeneous impact of depreciation shock generates a sizable profitability premium. Later in the paper, we demonstrate that profitable firms have high transitory productivity Z and increase capital utilization above 1 in order to seize transient profitable opportunities, resulting in positive and significant exposures to aggregate depreciation risk.

3.4 Firms' problem

Let Γ_t denote the cross-sectional distribution of capital $K_{i,t}$, permanent productivity $X_{i,t}$ and transitory productivity $Z_{i,t}$. Upon observing the exogenous aggregate state A_t, the entire distribution of firms Γ_t, firm specific productivity, $X_{j,t}, Z_{j,t}$ and the endogenous firm capital $K_{j,t}$, firm j makes optimal investment decision $I_{j,t}$, and optimal capital utilization rate $u_{j,t}$, to maximize the present value of future dividend:

$$V(K_{j,t}, Z_{j,t}, X_{j,t}|A_t, \Gamma_t) = \max_{I_{j,t}, u_{j,t}} D_{j,t} + E_t\left[M_{t+1} V(K_{j,t+1}, X_{j,t+1}, Z_{j,t+1}|A_{t+1}, \Gamma_{t+1})\right] \tag{16}$$

where dividend $D_{j,t}$ satisfies

$$D_{j,t} \leq \pi(A_t, Z_{j,t}, X_{j,t}, K_{j,t}) - I_{j,t} - H(I_{j,t}, K_{j,t}) - F(K_{j,t}) \tag{17}$$

and capital accumulation equation 13.

Investment entails standard convex adjustment cost that has a quadratic form, $H(I_{j,t}, K_{j,t}) = \frac{\eta}{2}(\frac{I_{j,t}}{K_{j,t}} - \delta)^2 K_{j,t}$. $F(K_{j,t})$ denotes the nonnegative fixed cost of production. Fixed costs are proportional to firm size $K_{j,t}$ to ensure that they do not become trivially low as the firm grows. Previous literature (Zhang (2005)) have established that fixed cost of production is a crucial ingredient that enables standard Q model to match the value premium.

3.5 Entry and exit

Now we describe the entry of new firms and the death of incumbents. At the end of each period t, after the investment decisions have been made, firms receive a random death shock with probability κ_D and exit the economy. Upon receiving the death shock, a firm pays out all dividend and its capital stock evaporates.

At the same time, κ_D fraction of new firms come into existence. New firms are endowed with initial productivity $Z_{0,t}$ and $X_{0,t}$. The representative agent optimally decides the amount of initial capital $K_{0,t}$ to new entrants and new-born firms start to produce one period after they enter the economy. New entrants solve the following optimization problem:

$$V_0(Z_{0,t}, X_{0,t}|A_t, \Gamma_t) = \max_{K_{0,t}} -K_{0,t} - H_0(K_{0,t}) + (1-\kappa_D)E_t\left[M_{t+1}V(K_{0,t}, Z_{j,t+1}, X_{j,t+1}|A_{t+1}, \Gamma_{t+1})\right] \tag{18}$$

New entrants decide the optimal level of initial investment $K_{0,t}$. Setting up a new firm entails additional cost $H_0(K_{0,t})$ which we interpret as entry cost. With probability of $(1 - \kappa_D)$, new entrants survive. Conditioning on survival, new entrants carry $K_{0,t}$ amount of capital to the next period and become an incumbent firm to produce output. Initial transitory productivity $Z_{0,t}$ is fixed in our calibration. We choose the initial permenant productivity $X_{0,t}$ so that the total amount of X shock firms is normalized to 1 at all time.

Given the entry and exit dynamics described above, the law of motion of firm distribution Γ can be written as:

$$\begin{aligned}\Gamma(K_{a',t}, X_{i',t}, Z_{u',t}) =& (1-\kappa_D)\int \mathbb{I}_{\{I(K_{a,t-1}, X_{i,t-1}, Z_{u,t-1}) = K_{a',t} - (1-\delta(u_{t-1}))K_{a,t-1}\}} \mathrm{P}(e^{\varepsilon_t} = \frac{X_{i',t}}{X_{i,t-1}}) \mathrm{P}(Z_{u',t}|Z_{u,t-1}) \\ & d\Gamma(K_{a,t-1}, X_{v,t-1}, Z_{u,t-1}) + \kappa_D \mathbb{I}_{\{K_{a',t} = K_{0,t-1}\}} \mathrm{P}(e^{\varepsilon_t} = \frac{X_{i',t}}{X_{0,t-1}}) \mathrm{P}(Z_{u',t}|Z_{0,t})\end{aligned} \tag{19}$$

which says that at time t the total measure of firm with firm-level state variable $(K_{a',t}, X_{i',t}, Z_{u',t})$ comes from two sources at time $t-1$. First, incumbent firms who survive the death shock. For these surviving firms, the optimal investment and utilization decision result in the amount of capital that is identical to $K_{a',t}$. At the same time, the realization of permanent and transitory firm-level shocks bring their current level of productivities to $(X_{i',t}, Z_{u',t})$. Second, new entrants at time $t-1$ whose optimal initial size coincides with $K_{a',t}$. Realizations for shocks also bring them to the state space $(X_{i',t}, Z_{u',t})$.

3.6 Competitive equilibrium

A competitive equilibrium consists of an optimal investment policy, $I(K_{j,t}, Z_{j,t}, X_{j,t}, A_t, \Gamma_t)$, an optimal utilization rate $u(K_{j,t}, Z_{j,t}, X_{j,t}, A_t, \Gamma_t)$, value function $V(K_{j,t}, Z_{j,t}, X_{j,t}, A_t, \Gamma_t)$, optimal initial capital stock for new entrants $K_0(A_t, \Gamma_t)$, a law of motion for the firm distribution Γ_t, and the a pricing kernel function $M_{t+1}(\Gamma_{t+1}, A_{t+1}|\Gamma_t, A_t)$ so that the following conditions hold:

- Given the law of firm distribution Γ_t, $I(K_{j,t}, Z_{j,t}, X_{j,t}, A_t, \Gamma_t)$ and its induced pricing kernel dynamics $M_{t+1}(\Gamma_{t+1}, A_{t+1}|\Gamma_t, A_t)$, $u(K_{j,t}, Z_{j,t}, X_{j,t}, A_t, \Gamma_t)$ and $V(K_{j,t}, Z_{j,t}, X_{j,t}, A_t, \Gamma_t)$ solve the dynamic optimization problem 16; $K_0(A_t, \Gamma_t)$ solves new entrants' optimization problem 18.

- The law of motion for the firm distribution, 19, is consistent with the optimal decisions for both incumbent and new entrant firms. Also it pins down the pricing kernel dynamics $M_{t+1}(\Gamma_{t+1}, A_{t+1}|\Gamma_t, A_t)$.

- goods market clearing: aggregate output equals aggregate consumption plus aggregate investment that includes investment costs for both incumbent and new entrant firms:

$$
\begin{aligned}
C_t + \int \Big(I(K_{a,t}, Z_{j,t}, X_{i,t}) + H(I(K_{a,t}, Z_{j,t}, X_{i,t}), K_{a,t}) \Big) d\Gamma(K_{a,t}, Z_{j,t}, X_{i,t}) + K_{0,t} + H_0(K_{0,t}) \\
\leq \int A_t \Big[u_t^{\nu} K_{a,t}^{\nu} Z_{j,t}^{1-\nu} X_{i,t}^{1-\nu} \Big]^{\alpha} L^{1-\alpha}(K_{a,t}, Z_{j,t}, X_{i,t}) d\Gamma(K_{a,t}, Z_{j,t}, X_{i,t})
\end{aligned}
\tag{20}
$$

- labor market clearing gives rise to the optimal hour allocation equation 11.

3.7 Homogeneity property and the solution for the competitive equilibrium

A significant obstacle to numerically solving and analyzing our general model with heterogeneous firms is the fact that the entire firm distribution, Γ_t, is an aggregate and endogenous state variable. Through the market clearing condition 20, Γ_t affects current consumption C_t. More importantly, firms must forecast the law of motion of this distribution in order to evaluate cash flows intertemporally, as described in the firms' objective function. We use the homogeneity structure of our problem, which is a result of production technologies and shock structures, to reduce the dimension of the distribution by one. Then, we develop a numerical method for solving the competitive equilibrium that can be efficiently implemented using the perturbation approach.

The model features a balanced growth path with average growth rate μ_A. At firm-level, firm variables are driven by a random walk process X. We first reformulate the model in terms of stationary variables and then solve for the competitive equilibrium. We define the following stationary aggregate variables: $c_t = \frac{C_t}{\bar{A}_t}$. At firm-level, we scale firm variables with the product of permanent component of idiosyncratic productivity $X_{j,t}$ and the aggregate trend A_t: $k_{j,t} = \frac{K_{j,t}}{X_{j,t}\bar{A}_t}, y_{j,t} = \frac{Y_{j,t}}{X_{j,t}\bar{A}_t}, \tilde{I}_{j,t} = \frac{I_{j,t}}{X_{j,t}\bar{A}_t}, k_{0,t} = \frac{K_{0,t}}{X_{0,t}\bar{A}_t}$. We find it convenient to define what we call summary measure to replace the distribution Γ_t.

$$m_t(k, Z) = E[X \mid k, Z]\phi_t(k, Z) \tag{21}$$

where $\phi(k, Z)$ is the joint density for (k, Z). Summary measure m records the total amount of X shock for a firm with capital k and transitory productivity Z. When we solve for the equilibrium numerically, we discretize the state space of normalized capital k into nK different grid points. nZ refers to the total number of discretized Markov states for transitory productivity Z.

To see why it is more convenient to replace Γ_t with summary measure m_t, we re-write the resource constraint 20 using stationary variables and summary measure:

$$c_t + \sum_{v=1}^{nZ}\sum_{j=1}^{nK}\left[\tilde{I}\left(k_j, Z_v\right) + H(\tilde{I}\left(k_j, Z_v\right), k_j)\right] m_t\left(k_j, Z_v\right) + X_{0,t}\cdot\left[k_0 + H_0\left(k_{0,t}\right)\right] \\ \leq \theta_t\left[\sum_{v=1}^{nZ}\sum_{j=1}^{nK} Z_v^{1-\nu} k_j^{\nu} m_t\left(k_j, Z_v\right)\right]^{\alpha} \tag{22}$$

Because firm-level decisions are homogeneous of degree 1 with respect to the permanent component X , aggregating firm-level resources can be achieved with a lower dimension distribution m_t rather than keeping track of firm-level resources using the whole distribution Γ_t itself.

Now we are ready to compute how summary measure m evolves over time. For a particular capital productivity pair (k_u, Z_v), the total amount of X at this location at $t+1$ is given by:

$$m_{t+1}(k_u, Z_v) = \sum_{j=1}^{nZ}\sum_{i=1}^{nK}(1-\kappa_D)\, P\left(e^{-\varepsilon_x'} = \frac{k_u}{g(k_i, Z_j)}e^{\mu_A}\right) m_t(k_i, Z_j)\frac{g(k_i, Z_j)}{k_u}e^{\mu_A+\zeta_t+\varepsilon_{t+1}^A}\pi_Z(Z_v \mid Z_j)$$
$$+(1-\kappa_D)\, P\left(e^{-\varepsilon_x'} = \frac{k_u}{k_0}e^{\mu_A}\right) m_{0,t}(X_{0,t})\frac{k_{0,t}}{k_u}e^{\mu_A+\zeta_t+\varepsilon_{t+1}^A}\pi_Z(Z_v \mid Z_0) \tag{23}$$

where $g(k_i, Z_j) = (1-\delta(u_{k_i,Z_j}))k_i + \tilde{I}(k_i, Z_j)$ denotes the total amount of capital to carry to the next period, before shocks for the next period realize. The total measure of X can come from both incumbents and new entrants. Incumbent firms that survive to the next period whose optimal decisions and realizations of shocks lead them to (k_u, Z_v). New entrants with optimal initial size k_u upon entry with the right value of shocks also end up with (k_u, Z_v). $m_{0,t}(X_{0,t})$ denotes the measure for the new entrants. We set $m_{0,t}$ so that the total measure of X across all types of firms is normalized to be 1.

In the Appendix B, we demonstrate that the solution to the competitive equilibrium consists of a set of non-linear equation system that jointly determine optimal investment/utilization policies, the law of motion for summary measure m_t under the market clearing conditions. We show how to solve the system of equations efficiently using perturbation method.

4 Quantitative results

In this section, we calibrate our model and evaluate its ability to replicate key moments of both macroeconomic quantities and asset prices at the aggregate level. More importantly, we investigate its performance in terms of quantitatively accounting for key features of firm characteristics and producing both a value premium and profitability premium in the cross-section. For macroeconomic quantities, we focus on a long sample of U.S. annual data from 1930 to 2018. All macroeconomic variables are real and per capita. Consumption, output, and physical investment data are from the Bureau of Economic Analysis (BEA). For the purpose of cross-sectional analyses, we make use of several data sources at the micro level, including (1) firm-level balance sheet data from the CRSP/Compustat Merged Fundamentals Annual Files, (2) monthly stock returns from CRSP.

Table 3: Calibration

Parameter	Symbol	Value
Relative risk aversion	γ	22
IES	ψ	0.95
Capital share in production	α	0.33
Span of control parameter	ν	0.85
Mean productivity growth rate	$e^{a_{ss}}$	0.02
Time discount rate	β	0.98
Capital depreciation rate	δ	0.10
Capital adj. cost paramter	η	5.5
Exit probability	κ_D	0.09
Initial productivity	X_0	0.053
Persistence of ζ	ρ_ζ	0.97
Volatility of ζ shock	σ_ζ	0.75%
Mean of idio. permanent productivity shock	μ_X	-0.02
Volatility of idio. permanent productivity shock	σ_X	0.2
Persistent of idio. transitory productivity shock	ρ_Z	0.5
Volatility of idio. transitory productivity shock	σ_Z	0.7

4.1 Calibration

We calibrate our model at annual frequency and present the parameters in Table 3. The first group of parameters are standard in the literature. In particular, we set capital share parameter α to 0.33, as in the standard business-cycle literature (Kydland and Prescott (1982)), to match the average capital share in the U.S. economy. The span of control parameter ν is set to 0.85, which broadly falls into the range of this parameter used in the literature by, for example, Atkeson and Kehoe (2005), Gollin (2008), Hsieh and Klenow (2009) and Ai et al. (2020).

The parameters in the second group are determined by matching a set of first moments of quantities and prices. We set the long-term average output growth rate to match a value for the annual U.S. growth rate of 2%. The time discount factor β is set to match a low risk-free interest rate. The capital depreciation rate is fixed to be 10% per year following the RBC literature (Kydland and Prescott (1982)). The persistence parameter of the persistent component of TFP shocks ρ_ζ and the standard deviation of it σ_ζ are set to 0.97 and 0.75% as in Croce (2014). The adjustment cost parameter is set to deliver a reasonable time-series variation of aggregate investment. At firm-level, we distinguish two types of firm-specific shocks, transitory component Z and permanent shock X. We set the standard deviation of permanent shock ε_X to be 20% and the standard deviation of transitory shock Z to be 70% based on estimates reported in Gourio (2008). We choose the mean of permanent component to correct for the Jensen term so that firm grows at the same rate as the aggregate economy.

We calibrate the exogenous firm death rate κ_D to be 9% per year to match the average exit rate in the data. The new entrant firms initial productivity is set such that the total measure of X across all types of firms is normalized to be 1. In our model model, we do not have financial leverage, all firms are purely equity financed.

4.2 Aggregate moments

We now turn to the quantitative performance of the model at the aggregate level. We solve and simulate our model at the annual frequency and compute model-implied annual moments. We show that our model is broadly consistent with the key empirical features of macroeconomic quantities and asset prices. Table 4 reports the model-simulated moments of macroeconomic quantities and asset returns and compares them to their counterparts in the data.

Table 4: Aggregate moments

This table presents moments from the model simulation and the data, at annual frequency. The top panel reports the basic statistics of macroeconomic quantities, and the bottom panel reports the moments for asset prices.

	Data	Benchmark
$\sigma(\Delta y)$	3.05	2.84
$\sigma(\Delta c)/\sigma(\Delta y)$	0.83	1.05
$\sigma(\Delta i)/\sigma(\Delta y)$	2.61	1.79
$E[R^M - R^f]$	5.71	6.65
$\sigma(R^M - R^f)$	20.89	5.84
$E[R^f]$	1.10	4.84
$\sigma(R^f)$	0.97	3.06

In terms of aggregate moments for macro quantities (top panel), our calibration features a low volatility of consumption growth (3%) and a relatively high volatility of investment (5.1%). The investment to output ratio is 22%, which is close to the value of 17% in the data. Our model inherits the success of real business-cycle models with respect to the quantity side of the economy. Turning the attention to the asset pricing moments (bottom panel), our model produces an average risk-free rate 4.8% and a high equity premium 6.65%, comparable to key empirical moments for aggregate asset markets. Overall, our model performs quite well in terms of standard macro and asset pricing moments at the aggregate level.

4.3 Optimal utilization choice

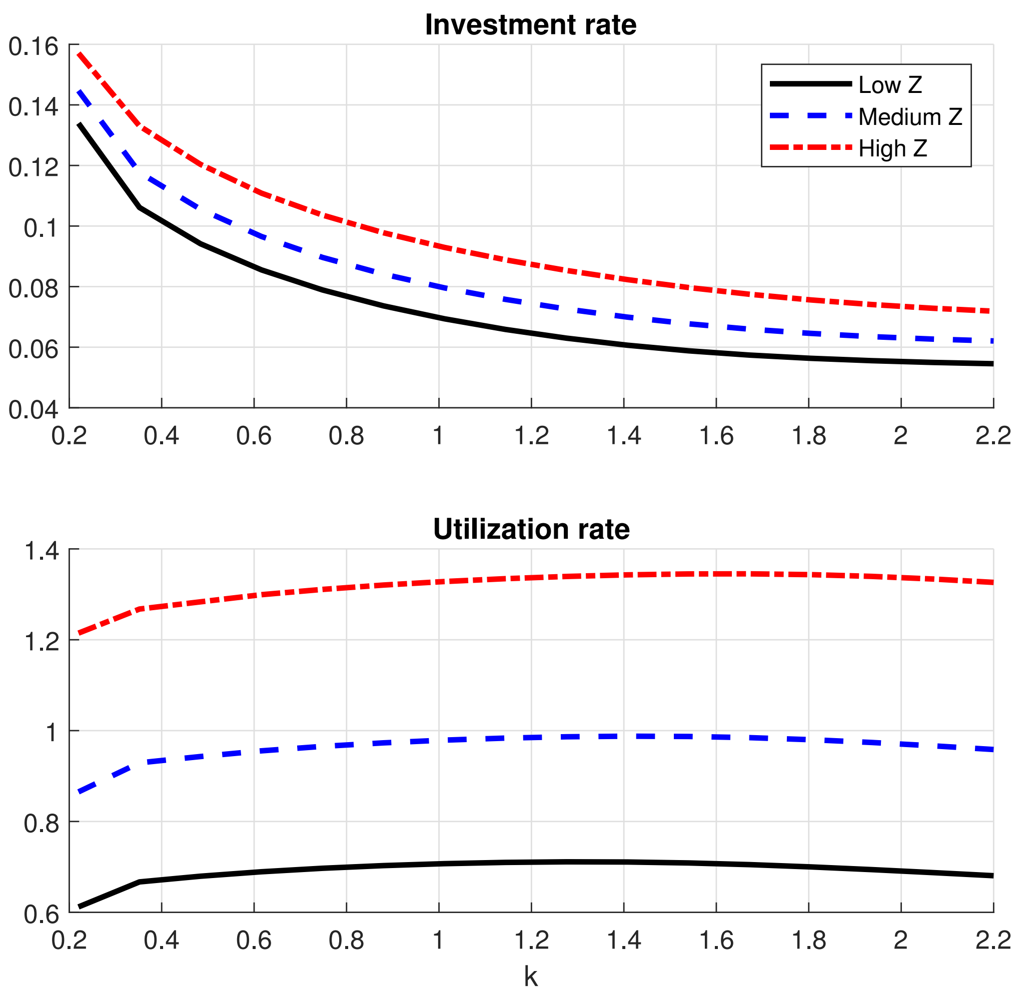
The optimal choice of capital utilization is determined by balancing the marginal benefit of increased capital utilization and the marginal cost of increased capital utilization. Increased capital utilization accelerates capital depreciation and reduces future capital capacity for production. On the plus side, increased utilization results in increased current output and dividend payments to shareholders. Therefore, firms that hit by different types of idiosyncratic shocks will choose utilization differently.

To understand the cross-sectional heterogeneity in utilization and investment, we plot optimal utilization choices and investment rates across all types of firms in Figure 2. The lower panel of Figure 2 shows that high Z and large firms tend to utilize capital more. Due to the decreasing return to scale technology, the marginal product of capital for large firms is typically low. They downsize their capital stock by a combination of disinvestment, and high capital depreciation resulting from high utilization rate. High Z firms take advantage of current transient productivity to raise their contemporaneous outputs, by use current capital intensively. Interestingly, we observe a significant difference in utilization rate for high vs low Z firm while their optimal investment policies do not differ that much, as shown in the upper panel of Figure 2. This distinction highlights the roles of transitory vs permanent shock in determining firms' optimal production policies. Since investment matters more for long term growth of a firm, it is intuitive for firms' to adjust their investment rates following permanent productivity shocks, rather than transitory shock. As a transient profitable opportunity arrives, firms respond by utilizing existing capital more, without excessive changes in investment policy because transient favorable productivity shock may not persist.

The utilization and depreciation relation 14 provides a novel economic channel to jointly explain value and profitability premium, as illustrated in the lower panel of figure 2. Holding size k fixed, high Z firms utilize capital above its long run steady state level one which imposes themselves more to the aggregate depreciation shock. As a result, investors require a higher risk premium when investing in the stocks of high Z firms versus low Z firms. Given that variation in Z is largely responsible for variation in the profitability characteristic, increased exposure to aggregate depreciation shocks, due to a higher level of utilization, results in a sizable profitability premium.

Figure 2: Policy functions: utilization vs investment

This figure plots how investment rate and utilization rate change across different grid point of capital k and transitory shock Z.



In addition to the established mechanism that operating leverage induced by fixed production costs generates value premiums, our model creates a new channel for the value premium, through capital utilization and depreciation risk. While we keep transitory productivity Z constant in the lower panel of figure 2, we observe some variation in utilization rates across firms of various sizes. Value firms (those with a high k size) have a low permanent productivity X and a high level of capital, which contradicts their low productivity. Value firms prefer to use capital more intensively in order to reduce their capital stock, which results in increased exposure to depreciation risk. Quantitatively, our model produces a sizable value premium thanks to both the operating leverage and the endogenous utilization channel.

4.4 Value and Profitability premium

We present the model simulation results for the value and profitability premium, as well as cross-sectional firm characteristics, in this section. Our model is capable of quantitatively accounting for both the value and profitability premiums. More importantly, because value and profitability sorted portfolios are driven by distinct types of firm-level shocks, our model achieves a clear separation between these two factors and their correlation is closer to the data than in standard models.

Value premium in the model Table 5 compares the moments of model simulated data and the empirical data for book-to-market sorted portfolios. Our model generates a similar pattern as in the data for firm characteristics and expected returns. Value portfolios earn a higher excess return than profitable firms, around 2.31 percent. Small growth firms also have higher gross profitability and investment rates that are consistent with the data. At the same time, growth firms receive sequences of positive permanent shock which raise their marginal product of capital. They choose to preserve capital for future production rather than utilizing it more intensively as value firms do. Overall, our model produces a sizable value premium as well as realistic firm characteristics for Book-to-Market sorted portfolios.

Profitability premium in the model We sort firms into quintiles based on their gross profitability, defined as operating profits to capital ratio, and compare the moments calculated from the model and the empirical data, as shown in Table 6. The model delivers a reasonable profitability premium at 1.13 percent. At the same time, our model replicates the flat pattern for investment rates across profitability sorted portfolios in the data. Because high profitability is mainly driven by high transitory shock Z, profitable firms prefer the intensive margin of investment (utilization) than the extensive margin (invest to expand capital stock).

Table 5: Book-to-market portfolios: Data and model

This table compares the moments in the empirical data (Panel A) and the model simulated data (Panel B) at the portfolio level. Firms are soted into quintile portfolios based on their book-to-market (BM) ratios. Panel A reports the median of firm characteristics, such as book-to-market ratio (BM), gross profitability (GP/A) and investment rate (I/K). We also report the excess return of value-weighted portfolios, $E[R^e]$, they are annualized by multiplying by 12, in percentage. Panel B reports the corresponding moments computed from model simulations, at annual frequency.

Panel A: Data

	Value	2	3	4	Growth	Growth-Value
BM	1.53	0.89	0.61	0.39	0.18	-1.35
GP/A	0.17	0.19	0.26	0.32	0.36	0.19
I/K	0.07	0.09	0.10	0.12	0.15	0.08
ΔK	0.01	0.03	0.05	0.07	0.10	0.09
Perm/K	0.16	0.21	0.28	0.38	0.55	0.39
Trans	0.13	0.58	0.85	1.01	0.46	0.34
$E[R^e](\%)$	10.11	7.71	7.49	7.02	6.86	-3.26

Panel B: Model

	Value	2	3	4	Growth	Growth-Value
BM	1.00	0.89	0.78	0.66	0.53	-0.46
GP/A	0.23	0.26	0.26	0.25	0.23	0.00
I/K	0.05	0.06	0.07	0.08	0.11	0.06
ΔK	-0.09	-0.08	-0.07	-0.04	0.00	0.09
X/K	0.65	0.99	1.34	2.07	2.84	2.19
Z	1.20	1.32	1.41	1.57	1.71	0.51
$E[R^e](\%)$	7.79	7.73	7.37	6.73	5.41	-2.31

As a result, the investment dimension along the profitability sorted portfolio is flat but the utilization dimension exhibit a larger spread.

Table 6: Gross profitability portfolios: Data and model

This table compares the moments in the empirical data (Panel A) and the model simulated data (Panel B) at the portfolio level. Firms are soted into quintile portfolios based on their gross profitability (GP/A) ratios. Panel A reports the median of firm characteristics, such as gross profitability (GP/A), book-to-market ratio (BM), and investment rate (I/K). We also report the excess return of value-weighted portfolios, $E[R^e]$, they are annualized by multiplying by 12, in percentage. Panel B reports the corresponding moments computed from model simulations, at annual frequency.

Panel A: Data

	Low	2	3	4	High	High-Low
BM	0.75	0.79	0.67	0.55	0.43	-0.32
GP/A	0.03	0.12	0.26	0.41	0.66	0.63
I/K	0.08	0.09	0.10	0.11	0.12	0.04
ΔK	0.03	0.04	0.05	0.05	0.06	0.03
Perm/K	0.13	0.15	0.24	0.36	0.45	0.32
Trans	0.08	0.65	0.60	0.90	0.87	0.80
$E[R^e](\%)$	4.70	6.33	5.44	7.28	8.02	3.32

Panel B: Model

	Low	2	3	4	High	High-Low
BM	0.88	0.77	0.86	0.75	0.80	-0.08
GP/A	0.20	0.23	0.25	0.27	0.31	0.11
I/K	0.06	0.07	0.06	0.07	0.06	0.01
ΔK	-0.05	-0.06	-0.07	-0.06	-0.08	-0.03
X/K	1.14	1.62	1.14	1.62	1.34	0.20
Z	0.48	0.70	1.06	1.77	3.47	2.99
$E[R^e](\%)$	6.61	7.12	7.20	7.33	7.74	1.13

The separation between value and profitability premium We demonstrate that value and growth firms' cash flow growth consists of distinct components that differ in persistence and therefore provide a separation between value and profitability empirically. We show that such a separation is also present in our quantitative model.

Profitable firms have a high transitory productivity Z, whereas unprofitable firms have a low transitory productivity Z, as illustrated in as shown in panel B of table 6. However, the Z dimension along the BM sorted portfolio is rather flat. The other important firm-level state variable is the normalized firm size $k(\frac{K}{X})$ that contains information about the permanent productivity X. Small growth firms that receive a sequence of positive X shock generally have lower k and we see a clear distinction in k along the BM sorted portfolios. On the other hand, profitability is unrelated to k or the permanent component X, as BM sorted portfolios exhibit little variation in k. Thus, variation in transitory productivity Z accounts for the majority of the variation in profitability, whereas heterogeneity along k (or X) accounts for the variation in BM ratios.

The investment pattern across profitability and BM sorted portfolios exemplifies distinct determinations of these portfolios. Due to the fact that growth and value firms differ primarily in terms of their permanent productivity X, we observe a significant difference in investment rates between value and growth firms, while the dispersion in investment rates between high and low profitability firms is much smaller. A permanent productivity shock X permanently shifts firms' marginal product of capital. As a result, firms' investment policies are more responsive to permanent shocks than to transitory shocks.

A puzzling fact in the data is that profitable firms tend have lower BM ratios which make them look a lot like growth firms, despite their high stock returns. We replicate this fact in panel A of table 6. Indeed, firms with higher GP/A tend to carry lower BM ratios. Earlier works (see Kogan and Papanikolaou (2013)) that can produce value and profitability premium simultaneously typically makes the counterfactual prediction that value and profitable firms both have high BM ratios. In other words, they fail to make a distinction between value and profitable firms, despite jointly accounting for the two premium. Our separation of value and profitability premium by distinct types of idiosyncratic shocks allows our model to match this fact. As shown in panel B of table 6, we observe the same pattern as in the data that high GP/A firms are associated with lower BM ratios and the pattern hold when we examine the BM sorted portfolios as in panel B of table 5.

5 Conclusion

In this paper, we develop a general equilibrium model with heterogeneous firms that jointly explain value and profitability premium in a unified framework. We emphasize the importance of two types of firm-level shocks to distinguish value and profitability strategies separately. This rich set of firm-level shocks allow us to match key firm characteristics for growth and profitable firms, which a single firm-level shock setup would fail to do so.

Variable utilization is a critical ingredient in our model that shifts firm cash-flow payouts intertemporally, which ultimately determine the high aggregate risk exposures for profitable firms. Optimal allocation of aggregate investment endogenously gives rise to highly procyclical investment processes for growth firms, which makes growth firms good hedges for business cycle risks. When we calibrate our model to the dynamics of macroeconomic quantities, we show that the utilization and investment channel are quantitatively important determinants for the cross-section of asset returns.

References

Ai, H. and D. Kiku (2013): "Growth to value: Option exercise and the cross section of equity returns," *Journal of Financial Economics*, 107, 325–349.

Ai, H., J. E. Li, K. Li, and C. Schlag (2020): "The collateralizability premium," *The Review of Financial Studies*, 33, 5821–5855.

Atkeson, A. and P. J. Kehoe (2005): "Modeling and Measuring Organization Capital," *Journal of Political Economy*, 113, 1026–53.

Bai, H., K. Hou, H. Kung, E. X. Li, and L. Zhang (2019): "The CAPM strikes back? An equilibrium model with disasters," *Journal of Financial Economics*, 131, 269 – 298.

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Belo, F., J. Li, X. Lin, and X. Zhao (2017): "Labor-force heterogeneity and asset prices: The importance of skilled labor," *Review of Financial Studies*, 30, 3669–3709.

Bouchaud, J.-P., P. Krueger, A. Landier, and D. Thesmar (2019): "Sticky expectations and the profitability anomaly," *The Journal of Finance*, 74, 639–674.

Chen, A. Y. (2018): "A general equilibrium model of the value premium with time-varying risk premia," *The Review of Asset Pricing Studies*, 8, 337–374.

Croce, M. (2014): "Long-run productivity risk: A new hope for production-based asset pricing?" *Journal of Monetary Economics*, 66, 13–31.

Deng, Y. (2020): "Product Market Competition and the Profitability Premium," *Working Paper.*

Dou, W., Y. Ji, and W. Wu (2020): "The Oligopoly Lucas Tree: Consumption Risk and Industry-Level Risk Exposure," *working paper.*

Epstein, L. G. and S. E. Zin (1989): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework," *Econometrica*, 57, 937–969.

Fama, E. F. and K. R. French (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

Favilukis, J. and X. Lin (2015): "Wage rigidity: A quantitative solution to several asset pricing puzzles," *Review of Financial Studies*, 29, 148–192.

Gabaix, X. (1999): "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, 114, 739–767.

Gârleanu, N., L. Kogan, and S. Panageas (2012): "Displacement risk and asset returns," *Journal of Financial Economics*, 105, 491–510.

Gollin, D. (2008): "Nobody's business but my own: Self-employment and small enterprise in economic development," *Journal of Monetary Economics*, 55, 219–233.

Gomes, J., L. Kogan, and L. Zhang (2003): "Equilibrium Cross Section of Returns," *Journal of Political Economy*, 111, 693–732.

Gomes, j. f. and l. Schmid (2021): "Equilibrium Asset Pricing with Leverage and Default," *The Journal of Finance*, 76, 977–1018.

Gourio, F. (2008): "Estimating Firm-Level Risk," *working paper.*

Granger, C. W. J. and P. Newbold (2014): *Forecasting economic time series*, Academic Press.

Grigoris, F. and G. Segal (2020): "The Utilization Premium," *working paper.*

Harvey, A. C. (1985): "Trends and cycles in macroeconomic time series," *Journal of Business & Economic Statistics*, 3, 216–227.

Hasler, M., M. Khapko, and R. Marfe (2020): "Rational Learning and the Term Structures of Value and Growth Risk Premia," *working paper.*

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124, 1403–1448.

Kogan, L., J. Li, and H. Zhang (2020): "Operating Hedge and Gross Profitability Premium," *working paper.*

Kogan, L. and D. Papanikolaou (2012): "Economic Activity of Firms and Asset Prices," *Annual Review of Financial Economics*, 4, 1–24.

——— (2013): "Firm Characteristics and Stock Returns: The Role of Investment-Specific Shocks," *The Review of Financial Studies*, 26, 2718–2759.

——— (2014): "Growth opportunities, technology shocks, and asset prices," *Journal of Finance*, 69, 675–718.

Krusell, P. and A. A. Smith, Jr (1998): "Income and wealth heterogeneity in the macroeconomy," *Journal of Political Economy*, 106, 867–896.

Kydland, F. E. and E. C. Prescott (1982): "Time to build and aggregate fluctuations," *Econometrica*, 1345–1370.

Luttmer, E. G. (2007): "Selection, growth, and the size distribution of firms," *Quarterly Journal of Economics*, 122, 1103–1144.

Ma, L. and H. Yan (2018): "The value and profitability premiums," *Working paper.*

Miao, J. (2005): "Optimal Capital Structure and Industry Dynamics," *The Journal of Finance*, 60, 2621–2659.

Novy-Marx, R. (2013): "The other side of value: The gross profitability premium," *Journal of Financial Economics*, 108, 1–28.

Shumway, T. (1997): "The delisting bias in CRSP data," *Journal of Finance*, 52, 327–340.

Shumway, T. and V. A. Warther (1999): "The delisting bias in CRSP's Nasdaq data and its implications for the size effect," *Journal of Finance*, 54, 2361–2379.

Tong, J. and C. Ying (2020): "A Dynamic Agency Based Asset Pricing Model with Production," *working paper.*

Zhang, L. (2005): "The value premium," *Journal of Finance*, 60, 67–103.

Appendix

A Unobserved components decomposition

The unobserved components representation typically takes the form

$$\begin{aligned} y_t &= x_t + z_t, \\ x_t &= \mu + x_{t-1} + \varepsilon_t, \\ \phi_p(L) z_t &= \theta_q(L) \eta_t, \end{aligned} \tag{A1}$$

where y_t is the observed series (in log terms), x_t is the unobserved trend component, which is assumed to be a random walk with average growth rate μ, and z_t is the unobserved stationary component. The stationary component described in Equation (A1) can be specified using a very general ARMA(p,q) specification, with L being the lag operator.

To ensure that our decomposition is consistent with the productivity specification in our general equilibrium model, we assume that the innovation to the trend and transitory components are independent, thus $cov(\varepsilon_t, \eta_{t+j}) = 0, \forall j$. Furthermore, the stationary component is assumed to be an AR(1) process, as in Equation (3).

Substituting Equation (2) and (3) into Equation (1), taking first differences, yield,

$$\Delta y_t - \rho_z \Delta y_{t-1} = (1 - \rho_z)\mu + \varepsilon_t - \rho_z \varepsilon_{t-1} + \eta_t - \eta_{t-1}. \tag{A2}$$

Apply Granger's lemma (Granger and Newbold (2014)), the right hand side of Equation (A2) can be represented by an MA(1) process

$$\Delta y_t - \rho_z \Delta y_{t-1} = (1 - \rho_z)\mu + u_t + \hat{\theta} u_{t-1}. \tag{A3}$$

We can recover the trend component x_t and stationary component z_t by comparing the predicted values left-hand side of Equation (A2) and (A3), $E_t[\Delta y_{t+1} - \rho_z \Delta y_t]$, we can obtain the following relationship

$$\hat{\theta} u_t = -\rho_z \varepsilon_t - \eta_t.$$

Together with Equation (A2), we can solve for the innovations to the unobserved trend and stationary components,

$$\varepsilon_t = \frac{1}{1-\rho_z}\big[\Delta y_t - \rho_z \Delta y_{t-1} - (1-\rho_z)\mu + \hat{\theta}(u_t - u_{t-1})\big],$$
$$\eta_t = -\rho_z \varepsilon_t - \hat{\theta} u_t,$$

where $\hat{\theta}$ and u_t can be obtained by estimating Equation (A3) using an ARMA(1,1) process. By adding up the past series of shocks, we can recover the decomposed trend and stationary components

$$x_t = x_0 + \mu t + \sum_{s=1}^{t} \varepsilon_t,$$
$$z_t = \rho_z z_{t-1} + \eta_t.$$

B Numerical method

The optimal allocations and prices are pinned down by the following set of nonlinear equations. Let $\{k_j\}_{i=1}^{nK}, \{Z_j\}_{j=1}^{nZ}$ denote two sets of pre-determined grid points. We solve the system of nonliear equations on these points.

- optimal investment for a incumbent firm with state variable (k_i, Z_j)

$$\begin{aligned} 1 + h_0\left(\frac{\tilde{I}\left(k_i, Z_j\right)}{k_i}\right) =& \beta\left(1-\kappa_D\right)\mathbb{E}\left(\Lambda\left(A' \mid A, m\right)\right. \\ & \sum_{u=1}^{nK}\sum_{v=1}^{nZ} P\left(\frac{g\left(k_i, Z_j\right)}{k_u}\right) P\left(Z_v \mid Z_j\right) \\ & \left\{\theta' \alpha \nu Z_v^{1-\nu} k_u^{\nu-1} \Theta^{\alpha-1} + (1-\delta)\left(1 + h_0\left(\frac{\tilde{I}\left(k_u, Z_j\right)}{k_u} - h_1\right)\right)\right. \\ & \left.\left. + \frac{h_0}{2}\left(\frac{\tilde{I}^2\left(k_u, Z_j\right)}{k_u^2} - \delta^2\right)\right\}\right) \end{aligned} \tag{B4}$$

- optimal investment by new entrants

$$
\begin{aligned}
1+h'(k_0) =& \beta(1-\kappa_D)\mathbb{E}(\Lambda(A' \mid A, m) \\
& \sum_{u=1}^{nK}\sum_{v=1}^{nZ} P\left(\frac{k_0}{k_u}\right) P(Z_v \mid Z_0) \\
& \left\{\left(\frac{k_0}{k_u}\right)^{1-\nu} \theta' \alpha\nu Z_v^{1-\nu} k_0^{\nu-1}\Theta^{\alpha-1} + (1-\delta)\left(1+h_0\left(\frac{\tilde{I}(k_u, Z_v)}{k_0}\frac{k_0}{k_u} - h_1\right)\right)\right. \\
& \left. + \frac{h_0}{2}\left(\frac{\tilde{I}^2(k_u, Z_v)}{k_0^2}\left(\frac{k_0}{k_u}\right)^2 - \delta^2\right)\right\}\Bigg)
\end{aligned} \tag{B5}
$$

- where Θ is given by

$$
\Theta = \sum_i \sum_j m(k_i, Z_j) Z_j^{1-\nu} k_i^{\nu} \tag{B6}
$$

- law of motion for summary measure $m'_{u,v}, \forall u \in \{1,..,nK\}, \forall v \in \{1,...,nZ\}$

$$
\begin{aligned}
m'_{u,v} = & (1-\kappa_D)\sum_{i=1}^{nK}\sum_{j=1}^{nZ} P\left(\frac{g(k_i, Z_j)}{k_u}\right)\frac{g(k_i, Z_j)}{k_u} e^{-\Delta a'} m_{i,j}\pi_Z(Z_v \mid Z_j) \\
& + (1-\kappa_D) P\left(\frac{k_0}{k_u}\right)\frac{k_0}{k_u} e^{-\Delta a'} m_0 \pi_Z(Z_v \mid Z_0)
\end{aligned} \tag{B7}
$$

where Z_0 denotes initial productivity level upon entry and

$$
g(k_i, Z_j) = (1-\delta)k_i + \tilde{I}(k_i, Z_j).
$$

- aggregate resource constraint

$$
c + \sum_{v=1}^{nZ}\sum_{j=1}^{nK} \tilde{I}(k_j, Z_v) m(k_j, Z_v) + \sum_{v=1}^{nZ}\sum_{j=1}^{nK} \frac{\eta}{2}\left(\frac{\tilde{I}(k_j, Z_v)}{k_j} - h_0\right)^2 k_j m(k_j, Z_v) + X_0 \cdot [k_0 + h(k_0)] \leq \theta\Theta^{\alpha} \tag{B8}
$$

We firstly solve the above equation systems on each grid point without aggregate shocks, to obtain the steady state equilibrium allocations. We can determine the optimal allocations such as investment and utilization rate for each type of firms, as well as aggregate variables including aggregate consumption, output, etc. Then we apply perturbation method which solves the economy with aggregate shocks. In the end, we compute the present value of firm cash flows given optimal allocation as firm equity value.

C Variable definitions

Table 1: Variable definitions

Variables	Definition	Sources
BE	Book value of equity, computed as the book value of stockholders' equity, plus balance sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use the redemption, liquidation, or par value (in that order) as the book value of preferred stock.	Compustat
ME	Market value of equity is copmputed as price per share times the number of shares outstanding. The share price is taken from CRSP, the number of shares outstandings from Compustat or CRSP, depending on availability.	Compustat
BM	Book to market value of equity ratio.	Compustat
GP/A	Compustat item REVT minus COGS divided by AT.	Compustat
ΔSales	Growth rate of Compustat item SALE.	Compustat
I/K	Compustat item CAPX divided by PPEGT.	Compustat
ΔPPEGT	Growth rate of Compustat item PPEGT.	Compustat
ΔAT	Growth rate of Compustat item AT.	Compustat
ΔCF	Growth rate of Compustat item EBITDA.	Compustat
Macro variables	NIPA tables.	Compustat