

The Cyclical Composition of Startups

Eran B. Hoffmann*
Stanford University

[Please click here for the latest version](#)

November 20, 2017

Abstract

This paper proposes a new theory of business cycles based on the idea that financial uncertainty shocks change the nature of innovation. When investors become more risk tolerant, they fund riskier startups with greater growth potential. As these ambitious startups grow, the initial shock propagates and generates a boom in output and employment. I develop a heterogeneous firm industry model of the US business sector with countercyclical risk premia and innovation by startups and existing firms. The quantitative implementation of the model jointly matches time series properties of stock returns and macroeconomic aggregates, as well as micro evidence on firm cohort growth over the cycle.

*Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305 (email: eranhoff@stanford.edu). I am deeply indebted to Monika Piazzesi, Martin Schneider and Christopher Tonetti for invaluable guidance and support. I also thank Adrien Auclert, Jose Barrero, Nick Bloom, Emanuele Colonnelli, Patrick Kehoe, Pete Klenow, Pablo Kurlat, Moritz Lenel, Davide Malacrino, Sean Myers, Alessandra Peter, Cian Ruane, Itay Saporta-Ecksten, Alonso Villacorta, David Yang and seminar participants at Stanford University for helpful comments. Financial support from Stanford Institute for Economic Policy Research and Weiland Research Fellowship is gratefully acknowledged.

1 Introduction

The goal of macro-finance is understanding the connection between financial markets and the real economy. Empirical research has established that stock markets, which reflect the value of firms, are more volatile than the cash flows they generate. Furthermore, the effective discount rates on risky assets over the risk-free rate, also known as risk premia, appear to be low when economic conditions are good and high when economic conditions are bad. How do countercyclical risk premia contribute to macroeconomic fluctuations? Recent literature studies the impact of uncertainty shocks that generate time variation in risk premia.¹ A common finding is that in the presence of adjustment costs and other frictions, uncertainty delays investment and reduces hiring. This delay, however, does not inherently induce any persistent response and is quantitatively small.

This paper proposes a new theory of business cycles in which countercyclical variations in risk premia change the incentives to innovate in new and existing firms. When risk premia are low, investors fund riskier startups that are more innovative. If financial conditions remain favorable, these new firms keep innovating, expand, and generate an economic boom in output and employment that persists beyond business cycle frequency. Otherwise, the new firms forego growth opportunities and remain relatively small for the rest of their life cycle. When the price of risk is high, fewer firms enter overall. Those that do enter have a lower growth potential but are less exposed to future aggregate shocks. I find that the effect of countercyclical risk premia on innovation amplifies business cycle fluctuations in output and hours by 60 percent compared to fluctuations in productivity alone.

I develop a heterogeneous firm model of the US business sector that captures key features of firm creation, innovation, and exit. Firms maximize shareholder's value given state prices and differ in scale and in expected duration. Entrepreneurs create new firms and choose their expected duration, and incumbent firms endogenously innovate to increase their productive capacity. Aggregate shocks move the productivity and the effective discount rates on future profits in opposite directions, reflecting countercyclical risk premia. High expected duration firms choose to innovate more than low expected duration firms, making them more exposed to future risk premia, and therefore riskier. I calibrate the model to match the unconditional moments of stock returns and the detailed firm size by age distribution of all US private firms. The model also matches the stylized fact that firm cohorts that enter in booms have a greater share of large firms one year after entry than cohorts that enter in busts.

The quantitative implementation of the model jointly matches time-series properties of stock returns, output, hours, and firm entry, even though it is driven by just one random shock process. Stock returns in the model exhibit long booms between recessions and crashes at the start of every recession, corresponding to the timing and magnitude of realized stock returns in the data.

¹ See for example Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2016), Basu and Bundick (2017).

The model has two key ingredients. First, there are three firm types that can be chosen at entry. *Traditional firms*, such as a neighborhood restaurant, provide a standard service to an existing market and do not intend to grow much beyond their initial size. *Innovative firms*, such as a new Google or Walmart, may also start small, but intend to introduce new goods or services and expand to more markets. These firms have a lower exit rate and hence a higher expected duration. Since they have more time to recover the cost of innovation, they prioritize innovation over generating short-term profits. Innovation in the model stands in for activities such as development of new products, branding, and training of personnel, which are contributing to the growth of the firm. As innovative firms grow old, they become *mature firms*, which is the third firm type in the model. Mature firms maintain a moderate level of innovation but grow at a slower pace than innovative firms. The distinction between traditional, innovative, and mature firms captures the empirical fact that most firms stay small throughout their life cycle, while some firms grow to be very large. The transformation of innovative firms to mature firms captures the low growth rate of older firms in the data. Survey evidence suggesting that entrepreneurs differ in their ex-ante expectations of firm growth justifies the determination of firm type at entry.²

Second, the market price of risk (MPR) is an exogenous time-varying state variable that determines the discount rate of future risky payoffs but does not change the risk-free rate. Empirical research in finance has established that the expected returns on risky assets above the risk-free rate are countercyclical.³ In the model, firms are risky because their profits are exposed to shocks to the aggregate total factor productivity (TFP). Therefore, the model captures countercyclical risk premia with aggregate shocks that move TFP and the MPR in opposite directions. When TFP is rising and the MPR is falling, the value of firms increases more than the cash flows they generate. The technology of firm entry and incumbent firm innovation is time invariant in the model. Thus, when the value of firms increases, startup entry and incumbent firm innovation increase in response.⁴ Other models of firm entry and firm dynamics typically emphasize other frictions and mechanisms and hold the price of risk constant.

The key mechanism of the model operates through changes in the relative valuation of traditional and innovative firms over the business cycle. Intuitively, two opposing forces affect the relative value of firms. Traditional firms have more front-loaded profits, and so their values

² [Hurst and Pugsley \(2011\)](#) show that entrepreneurs differ in their desire to grow big and in their expectations for innovation activity, such as the introduction of new products and application for patents, trademarks, and copyrights. [Schoar \(2010\)](#) argues for a distinction between “subsistence” and “transformational” entrepreneurs, who have different goals in business creation.

³ See for example [Fama and French \(1989\)](#), [Lettau and Ludvigson \(2001\)](#) and [Cochrane and Piazzesi \(2005\)](#).

⁴ [Kaplan and Schoar \(2005\)](#) provide evidence that venture capital returns are similar to the returns on publicly traded stocks, and that the inflow of funds into venture capital is high when stock markets are high. [Brown, Fazzari, and Petersen \(2009\)](#) provide evidence that R&D expenditure is sensitive to the availability of external equity.

increase more when profits increase in response to a positive shock to TFP. This would be the dominant force in an environment with risk premia set to zero. But innovative firms own more growth options that are riskier than claims to cash flows from the current productive capacity of firms. Growth options are riskier because they act as implicit leveraged claims to the future value of the firm, which is more volatile than the firm’s cash flow.⁵ A negative shock to the MPR reduces the effective discount rate on the risky growth-option component of innovative firm values, and increases their total value relative to traditional firms. With fluctuations in TFP and MPR corresponding to the co-movement of asset prices and cash flows in the data, the cyclical effect of risk premia dominates that of profits. Therefore the share of innovative firms within firm cohorts is procyclical.

I calibrate the model to match unconditional moments of macroeconomic and financial variables, then evaluate its success at matching untargeted moments and time series properties of firm entry and financial returns. The calibration of the model also matches the firm size by age distribution of all US firms based on public-use Census data, which provide discipline for the entry, exit and growth rates of the three firm types. Despite the parsimonious nature of the firm type space, the model is able to match the detailed firm distribution data. It also matches other essential features of firm dynamics, including the observed declining exit rates and growth rates with firm age, and a conditional version of the empirical regularity known as Gibrat’s law.⁶

Stock returns in the model match the unconditional expected stock returns and unconditional volatility of returns in CRSP data. The model can also deliver a downward sloping term-structure of risk premia on levered equity, consistent with the empirical findings of [van Binsbergen, Brandt, and Koijen \(2012\)](#). I construct a simple dividend payout rule that keeps leverage stationary at the firm level, similar to the rules suggested by [Belo, Collin-Dufresne, and Goldstein \(2015\)](#). When the ratio of debt to firm value is high, firms reduce their dividend payments. This makes dividends more exposed to fluctuations in the value of the firms, and thus more risky in the short-run. Interestingly, the slope of the term structure becomes steeper in high leverage states of the world, such as in recessions, consistent with the recent findings of [Bansal, Miller, and Yaron \(2017\)](#).

I assess the quantitative success of the model in two ways. First, I compare simulated model moments with untargeted moments of firm cohort cyclical characteristics. A testable

⁵ [Berk, Green, and Naik \(1999\)](#), [Gomes, Kogan, and Zhang \(2003\)](#), [Garleanu, Panageas, and Yu \(2012\)](#), and [Kogan and Papanikolaou \(2014\)](#) point out that the value of growth options has a different risk profile than the value of assets in place. In their models, firms are endowed with future opportunities to invest in new projects, that generate implicit leveraged claims on the future value of projects. In my model, firms have opportunities to innovate, i.e. add new productive capacity *and* add new innovation opportunities. These generate implicit leveraged claims on the future value of firms. This paper focuses on aggregate fluctuations and captures features of firm entry, firm life cycle, and the firm size distribution, while the other papers focus on explaining documented facts on the cross-section and predictability of stock returns.

⁶ Conditional on firm type, the growth rate of firms is independent of their size.

prediction of the model is that the share of large firms in cohorts that enter in booms is higher than average several years after entry. And indeed, the data support this prediction. I regress the log number of one year old firms on real GDP growth at the time of their entry. Cohorts that enter when GDP growth is 3 percentage point higher have 7.2 percent more firms, and 15.3 percent more firms with more than 100 employees when one year old. The same regression on model simulations matches the coefficients, despite being untargeted in the model calibration.

Second, I evaluate the success of the model at generating aggregate fluctuations. The model generates aggregate fluctuations in output, hours, entry and stock returns that are observed in the data. Using the time series of the US real output over the period 1979:Q1 to 2016:Q4, I recover the model implied TFP and MPR, and construct the model implied series for hours, entry and stock returns. The model implied time series capture the timing and the magnitude of the aggregate fluctuations. In particular, model-implied realized stock returns exhibit long booms between recessions and sharp busts in every recession in the sample period, closely resembling the realized returns in the data. This success is surprising because the model parameters are only chosen to match unconditional moments, and the time series of shocks is recovered only from realized output. This result highlights the importance of countercyclical risk premia in explaining aggregate fluctuations in both asset prices and aggregate quantities.

What is the contribution of fluctuations in risk premia and innovation to fluctuations in the real economy? I decompose the implied output and hours time series into components related to TFP and the MPR by shutting off the fluctuations in one state variable at a time. I find that fluctuations in MPR, and hence in the risk premia, increase the volatility of output and hours by approximately 60 percent, by increasing innovation in booms and decreasing innovation in busts.

Fluctuations in risk premia and innovation also slow down recoveries from financial recession such as the Great Recession. Previous work on the Great Recession emphasized the role of credit and collateral channels in reducing investment and slowing down the recovery.⁷ Here, countercyclical risk premia reduce innovation in financial recessions and slow down recoveries, even when markets are complete and firms can freely issue equity and risk-free obligations. I measure the role of the fall in innovation during the Great Recession by replacing firm entry and innovation by incumbents in the period 2007:Q4 to 2010:Q1 with unconditional average values, and holding fixed all other shocks and innovation decisions before and after the recession. The counterfactual exercise reveals a loss of 3.6 percent of GDP in 2016 due to reduced innovation during the Great Recession, which accounts for half of the deviation from linear trend at the time.

Finally, the model provides a new narrative for the differences between the outcomes of the 2001 recession and the 2007-2009 Great Recession. In the mid to late 1990s risk premia were low and many innovative firms entered and grew quickly, creating an economic boom. Financial conditions started deteriorating in 2000, and stock prices fell by almost one half in

⁷ See for example Gilchrist and Zakrajšek (2012), Garcia-Macia (2017) and Villacorta (2017).

the following two years.⁸ But according to the model, there were many large innovative firms active at the time, that offset the decline in productivity and kept output and employment relatively high throughout the short 2001 recession. In contrast, the model suggests fewer innovative firms entered between 2002 and 2007, and those that had entered did not grow as fast. During the Great Recession, stock markets also fell by one half,⁹ reducing the incentives to innovate. This time however, innovative firms made up a smaller fraction of the existing US business sector, which rapidly shrank in response to the high risk premia. Fewer new firms entered, and even fewer of them were innovative firms. The lack of innovation in startups and incumbent firms during the Great Recession lead to a persistent decline in output, that a weak recovery of productivity and risk premia after the Great Recession helped propagate.

This paper is most closely related to the literature that studies business cycles in models with heterogeneous firms and investment in intangibles. Its main contribution is in explaining the fluctuations in output, hours and firm entry with a single shock process, that also captures the co-movement of stock prices with macroeconomic aggregates through countercyclical risk premia. Section 5 provides a detailed discussion of the contribution to the literature.

The rest of the paper is organized as follows. Section 2 describes the model. Section 4 outlines the calibration and the quantitative evaluation of the model. Section 3 discusses the key mechanism of the model. Section 5 discusses the contribution to the literature. Section 6 concludes.

2 Model

2.1 Overview and Setup

This section develops a heterogeneous firm industry model of the US business sector. In the model the boundary of the firm is technological: firms produce homogeneous goods using labor and non-transferable firm-specific organization capital.¹⁰ Startups innovate by creating new firms, and incumbent firms innovate by increasing their stock of organization capital. All firms take prices as given, including wages and state prices which are exogenous to the model.

The model departs from the literature in two key elements that capture the effects of financial shocks on the nature of innovation. First, I introduce three ex-ante different types of firms: *traditional*, *innovative* and *mature*. The technological difference between them is in the duration of their organization capital. The organization capital of traditional firms is of shorter duration, implying that they have less time to recover the costs of innovation. Traditional firms therefore endogenously choose to innovate less and grow slower than innovative and mature

⁸ The S&P 500 dropped from 1,527 in March 24, 2000 to 800 in October 4, 2002, completing a 47.59% drop.

⁹ The S&P 500 dropped from 1,562 in October 12, 2007 to 683 in March 6, 2009, completing a 56.24% drop.

¹⁰ My concept of organization capital draws on previous work of [Prescott and Visscher \(1980\)](#) and [Atkeson and Kehoe \(2005\)](#). It includes among other things brand value, formal and informal knowledge, training and task assignments that contribute to the firms productive capacity and are non-transferable to other firms.

firms. Second, I introduce state prices that capture countercyclical risk premia. When the aggregate productivity unexpectedly falls, investors become less willing to hold risky assets and demand higher expected rates of return. When risk premia are high, fewer startups are funded and less resources are devoted to innovation within firms.

Time is discrete and indexed by t . The economy is populated by a large number of heterogeneous firms that take prices as given and indexed by i . Asset markets are complete, and firms maximize shareholders value. Each firm allocates labor to production, innovation and management. Entrepreneurs create startups, choose the duration type of the new firms, and are free to enter and thus earn zero profits. Entry cost is type dependent and exhibits a congestion externality. Aggregate TFP, wages and state prices follow exogenous stochastic processes.

2.2 Incumbent Firms

Firms produce a homogeneous good Y_{it} using firm-specific organization capital K_{it} and labor L_{yit} in a Cobb-Douglas production function. They also allocate labor to innovation L_{git} , and management. Management requires λ units of labor per unit of organization capital. Productivity is determined by an aggregate TFP state variable A_t and wages W_t , that are taken as given. The profits of the firm are then the difference between output and the cost of labor,

$$\Pi_{it} = A_t K_{it}^\alpha L_{yit}^{1-\alpha} - W_t(L_{yit} + L_{git} + \lambda K_{it}), \quad (1)$$

where $\alpha \in (0, 1)$ is the organization capital share in production.

Besides their stock of organization capital, firms are also characterized by their firm type $\theta_{it} \equiv (\delta_{it}, \psi_{it})$, consisting of a depreciation rate δ_{it} and exogenous exit rate ψ_{it} . I choose a structure for the firm types that will later allow the model to match important empirical features of firm distribution with few parameters. There are three firm types indexed by $j \in \{tr, in, ma\}$: traditional (*tr*), innovative (*in*) and mature (*ma*). The depreciation rate of innovative firms δ_{in} is lower than that of traditional δ_{tr} and mature firms δ_{ma} , allowing innovative firms to accumulate organization capital quicker. The exit rate of traditional firms ψ_{tr} is higher than that of innovative ψ_{in} and mature ψ_{ma} firms. Innovative firms can unexpectedly transform into mature firms with a time invariant probability $P(\theta_{ma}|\theta_{in})$. This captures the decline in mean growth rate of firms as they age. For simplicity, I assume that the exit rates of innovative and mature firms are the same, $\psi_{in} = \psi_{ma}$, and that the probability of any other type transition is zero.

Firms grow their productive capacity by creating new organization capital. They use labor L_{git} and existing organizational capital, to create $BK_{it}^\beta L_{git}^{1-\beta}$ new organization capital, with $B > 0$ and $\beta \in (0, 1)$. Firms face three kinds of firm-level uncertainty. First, the stock of organization capital is subject to unanticipated multiplicative shocks ϵ_{it} , distributed according to $\log \epsilon_{it+1} \sim N(-0.5\sigma_K^2, \sigma_K^2)$. Second, firms receive an exogenous exit shock with probability ψ_{it} every

period. Since organization capital is firm specific, its scrap value is zero, and shareholders are left with nothing in case of exit. Third, and unique to innovative firms, is that they may unexpectedly become mature firms.

In sum, the law of motion for the organization capital of firm i can be written as:

$$K_{it+1} = \begin{cases} \left((1 - \delta_{it})K_{it} + BK_{it}^\beta L_{git}^{1-\beta} \right) \varepsilon_{it+1}, & \text{with prob. } 1 - \psi_{it} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

2.3 Startups

Entrepreneurs create new firms from “blueprints”. There is a large stock of blueprints available for free, each characterized by a predetermined firm type j . When a blueprint is implemented, it becomes a new firm with the corresponding firm type θ_j and starts producing at period $t + 1$ with a random initial stock of organization capital \tilde{K}_{it+1} . The distribution of the initial stock of organization capital is type and time invariant. It is distributed as the product of independent log-normal and Pareto random variables:

$$\tilde{K}_{it+1} = \tilde{K}_{it+1}^{(1)} \cdot \tilde{K}_{it+1}^{(2)}, \quad (3)$$

where $\log \tilde{K}_{it+1}^{(1)} \sim N(\tilde{\mu}, \tilde{\sigma})$ and $\tilde{K}_{it+1}^{(2)} \sim \text{Pareto}(\tilde{\xi})$.

Implementation of blueprints is costly and requires the use of labor. The aggregate supply of new entrants of type θ_j is denoted N_t^j and requires L_{st}^j units of labor, according to

$$N_t^j = \nu_j^\gamma (L_{st}^j)^{1-\gamma}, \quad (4)$$

where the constant $\gamma \in (0, 1)$ captures congestion in entry, and the constant $\nu_j \geq 0$ determines the scale of entry for each type.

Entrepreneurs do not internalize their impact on congestion and are free to enter, implying a zero profit condition on entry. Let S_t^j be the expected value of a new firm of type j in period t , then free entry in the startup sector implies an equilibrium condition for each type:

$$N_t^j S_t^j = W_t L_{st}^j. \quad (5)$$

Substituting the aggregate supply of new firms into the equilibrium conditions gives an expression for the number of entrants,

$$N_t^j = \nu_j (S_t^j / W_t)^{\frac{1}{\gamma} - 1}. \quad (6)$$

The supply of new firms of duration type j is determined by the ratio of firm value to current wages. The power $1/\gamma - 1$ determines the elasticity of entry with respect to firm value. When the value of new firms relative to wages is high, more labor is allocated to startups activity, and as a result more firms enter.

2.4 Aggregate Shocks

The aggregate TFP process A_t follows a trend-stationary path $A_t = e^{\mu t} Z_t$, with trend growth μ and a stationary AR(1) component Z_t ,

$$\log Z_{t+1} = \phi_z \log Z_t + \sigma_z e_{t+1}, \quad (7)$$

where ϕ_z is the persistence of TFP, σ_z is the volatility of TFP and e_{t+1} is an i.i.d standard normal random variable. e_{t+1} is the only aggregate shock in the model, and so I simply refer to it as “the” aggregate shock process.

Wages follow the trend component of TFP,

$$W_t = e^{\mu t}, \quad (8)$$

which keeps the ratio of the TFP to wages stationary and equal to Z_t .

2.5 Pricing Kernel

The pricing kernel M_{t+1} captures the state prices in the economy. This means that the return R_{t+1} on every asset must satisfy $1 = E_t[M_{t+1}R_{t+1}]$. I specify an affine pricing kernel, such that there is a constant risk free rate R_f and that the only priced risk factor is the shock to TFP e_{t+1} . The market price of risk (MPR) η_t determines the sensitivity of state prices to realizations of e_{t+1} ,

$$\log M_{t+1} = -\log R_f - \frac{1}{2}\eta_t^2 - \eta_t e_{t+1}. \quad (9)$$

Some properties of this pricing kernel are important to mention. First, we can verify that this pricing kernel implies a constant discount of risk-free assets at any maturity, since $E_t[M_{t+1}] = 1/R_f$. Second, any source of uncertainty that is independent of the unexpected change in the aggregate TFP will be priced in a risk-neutral way. This includes all idiosyncratic risk, implying that the innovation process is not distorted by uncertainty at the firm level. Third, as long as $\eta_t > 0$, the pricing kernel assigns a greater weight to states where the realization of e_{t+1} are low. Investor therefore value more assets that pay when times are (unexpectedly) worse. Furthermore, the conditional risk premium on assets with a payoff that is correlated with the priced risk factor is proportional to the MPR. To see that, consider an asset which pays $\exp(xe_{t+1} - 0.5x^2)$ next period for some constant x (mean payoff equal to 1). Then, its expected returns must satisfy

$$E_t[R_{t+1}] = 1/E_t[M_{t+1} \exp(xe_{t+1} - 0.5x^2)] = R_f \exp(x\eta_t),$$

implying a conditional expected excess return of $x\eta_t$. Here x measures the quantity of priced risk in the asset. Lastly, we can apply the Hansen-Jaganathan bounds to the conditional Sharpe ratio of any asset:

$$\frac{E_t(R_{t+1}) - R_f}{\text{VAR}_t(R_{t+1})} \leq \frac{\text{VAR}_t(M_{t+1})}{E_t(M_{t+1})} = \sqrt{\exp \eta_t^2 - 1} \approx \eta_t,$$

which means that the MPR also serves as an upper bound on the risk-return relationship in the model.

The price of risk follows an AR(1) with persistence ϕ_η and mean $\bar{\eta}$,

$$\eta_{t+1} = \phi_\eta \eta_t + (1 - \phi_\eta) \bar{\eta} - \sigma_\eta e_{t+1}, \quad (10)$$

where σ_η is the volatility of η_t . The shock to the price of risk is perfectly negatively correlated with the shock to TFP. This implies a countercyclical risk premia: when TFP unexpectedly goes down, the MPR goes up, and for a given risky cash flow the risk premium goes up and value goes down. The assumption of perfect correlation between shocks to TFP and MPR provides discipline to the model and allows recovery of all latent variables from observations of output growth alone. This in turn is used below to generate testable predictions on asset returns that indicate model success. However, this assumption is not necessary for the theoretical mechanism in the model and can be relaxed to achieve better fit to financial data.

2.6 The Firm Problem

Firms take the aggregate TFP, MPR and wages as given, as well as their stock of organization capital and duration type, and maximize shareholder's value by allocating labor to production and innovation. Let subscript t denote the aggregate state, and V_t the value function of firms, then the firm problem can be written as a Bellman equation,

$$V_t(K_{it}, \theta_{it}) = \max_{L_{yit}, L_{git}} \Pi_{it} + E_t[M_{t+1} V_{t+1}(K_{it+1}, \theta_{it+1})], \quad (11)$$

subject to the definition of profits Π_{it} and the evolution of capital (equations (1) and (2)).

The solution to the firm problem is a type specific allocations of labor to production and innovation that are proportional to the firm stock of organization capital and depends on TFP and the MPR (see Section 3.1 for detailed solution). Let l_{yt}^j and l_{gt}^j be the allocations of type j firms to production and innovation per unit of organization capital. This allocation determines the type j expected growth of capital, G_t^j , and the profits per unit of capital normalized by wages, π_t^j .

2.7 Aggregate Dynamics

The allocation of labor within firms is linear in organization capital, and the expected growth rate of firms of type j is conditionally independent of size. Therefore we can aggregate the stock of existing firms by type. The aggregate stock of capital of each type j , defined as $K_t^j = \int K_{it} I(\theta_{it} = \theta_j) di$, become state variables. The aggregate dynamics is represented by a

state-space system:

$$K_{t+1}^j = N_t^j E[\tilde{K}] + \sum_{j'} P(\theta_j|\theta_{j'}) G_t^{j'} K_t^{j'}, \quad (12)$$

$$L_t = \sum_j (l_{yt}^j + l_{gt}^j + \lambda) K_t^j + L_{st}^j, \quad (13)$$

$$Y_t = (1 - \alpha)^{\frac{1}{\alpha} - 1} W_t Z_t^{\frac{1-\omega}{\alpha}} \sum_j K_t^j, \quad (14)$$

where expected firm growth G_t^j , entry N_t^j and allocations of labor l_{yt}^j , l_{gt}^j , L_{st}^j are the solution to the firm and startup problems. In this system, the wages, the TFP and the MPR are exogenous. The labor allocations, the firm output and the firm growth rates are endogenous to the system.¹¹ Next period stock of organization capital is equal to the sum of capital in new firms and the remaining capital of incumbents, taking into account the probability of transition between innovative and mature firms. The aggregate labor demand is equal to the demand of labor in production, innovation within firms, management and startup implementation. Output is the outcome of the allocation of labor to production and the total organization capital stock in the economy.

If the expected capital growth rate is less than one for all firms in all states of the world, then the aggregate stocks of organization capital and the aggregate labor are stationary, and the output is trend stationary. This condition is however too strict in two respects. First, since the expected growth is determined endogenously and the domain of shocks is not restricted, this condition may be violated in some states. Second, in the quantitative implementation of the model, the expected growth of innovative firms is typically greater than one in all but extremely low economic conditions. A weaker condition is sufficient to guarantee mean stationarity: the growth rate of remaining firms in each type is less than 1 on average, or $E[\log P(\theta_j|\theta_j) + \log G_t^j] < 0$. This means that even if innovative firms grow at a high rate, as long as their net expected growth rate is less than the transition rate into mature type the stock of capital achieves a stationary distribution.

2.8 Discussion: Model Assumptions

Here I briefly discuss and motivate to assumptions of the model on wages and on the firm entry technology.

¹¹ This paper follows Zhang (2005) and İmrohoroğlu and Tüzel (2014) among others in studying a production economy under an exogenous countercyclical market price of risk. Since I focus on the business sector, this assumption seems appropriate. One could in principle make state prices endogenous by assuming a closed economy and constructing a general equilibrium model where the intertemporal consumption choices of households determine state prices. Such a model would have to generate similar countercyclical risk premia to capture the dynamics of stock returns and so would generate similar implications for the dynamics of innovation.

Wage process. Real wages in the model follow the predetermined trend component of TFP growth. In the data, output and hours are highly correlated, while fluctuations in wages are weakly correlated with output (see for example [Stock and Watson \(1999\)](#), [King and Rebelo \(1999\)](#) and [Swanson \(2004\)](#)). I therefore choose to capture the correlation of output and hours and abstract from an explicit wage determination mechanism.

The deterministic wage and elastic labor supply in my model amplify the response of employment and output to fluctuations in TFP. In this, I follow a considerable literature that studies what appears to be over sensitivity of employment to TFP (see for example [Shimer \(2005\)](#) and [Hall \(2005\)](#)). The trend growth in wages implies a stationary relationship between TFP and real wages, and since innovation is conducted using labor, delivers a stationary number of new firms and innovation in the model.

Startup entry. The canonical models of firm entry typically have either a fixed supply of new entrants ([Hopenhayn, 1992](#)), or a fixed cost of entry ([Klette and Kortum, 2004](#)). In my model, entry requires labor and exhibits decreasing returns to scale, putting it right between these two extremes. [Luttmer \(2007\)](#) motivates decreasing returns in entry with an unequal distribution of scarce entrepreneurial skills. Workers are endowed with one unit of labor, that can be used in production or transformed into entrepreneurial activity. The best entrepreneurs are hired first. As entry increases, less skilled entrepreneurs are hired leading to decreasing returns in terms of labor units. Another motivation, suggested by [Sedláček and Sterk \(2016\)](#), is that entry requires costly matching between ideas and investors. As the market for ideas becomes tight, search costs increase leading to decreasing returns.

The model assumes that entrepreneurs have information on ex-ante firm type, but not on firm initial scale or on future shocks. Allowing startups to know their type before entry is essential to the mechanism that changes the composition of startups over the business cycle. The available information to investors is the topic of a long literature starting with [Jovanovic \(1982\)](#). In recent literature, researchers study entrepreneurship surveys and find that entrepreneurs possess considerable information on their firm type. [Hurst and Pugsley \(2011\)](#) document that entrepreneurs vary in their expectations of growth, and their expectations for innovative activities such as applying for a patent or a trademark. These expectations also correspond to the ex-post actions of the firms, providing support for the information assumptions in this model.

3 Mechanism

This section provides more detailed discussion of the firm problem and its implications to aggregate dynamics.

3.1 The Firm Problem

Firms solve the Bellman equation (11), reproduced here:

$$V_t(K_{it}, \theta_{it}) = \max_{L_{yit}, L_{git}} \Pi_{it} + E_t[M_{t+1}V_{t+1}(K_{it+1}, \theta_{it+1})],$$

subject to the definition of profits and the evolution of capital (equations (1) and (2)). Because production and innovation are both homogeneous of degree one in capital and labor, we can normalize all variables by capital and wages. Let $v_t(K_{it}, \theta_{it}) = V_t(K_{it}, \theta_{it})/K_{it}W_t$ be the normalized value function, $\pi_{it} = \Pi_{it}/K_{it}W_t$ the normalized profits, $l_{yit} = L_{yit}/K_{it}$ and $l_{git} = L_{git}/K_{it}$ the normalized allocations of labor, and $G_{it} = E_t[K_{it+1}/K_{it}]$ the expected growth of firm i . Then we can rewrite the Bellman equation as

$$v_t(K_{it}, \theta_{it}) = \max_{l_{yit}, l_{git}} \pi_{it} + e^\mu G_{it} E_t[M_{t+1}v_{t+1}(\theta_{it+1}, K_{it+1})], \quad (15)$$

subject to the definition of normalized profits,

$$\pi_{it} = Z_t l_{yit}^{1-\alpha} - \lambda - l_{yit} - l_{git}, \quad (16)$$

and the technological constraint on expected growth,

$$G_{it} = (1 - \psi_{it})[1 - \delta_{it} + B l_{git}^{1-\beta}]. \quad (17)$$

Now capital does not appear outside of the value function in the Bellman equation, or any of the constraints. Thus I safely eliminate it altogether from the problem. In fact, the solution of the problem only depends on aggregate shocks and on the duration type j . Let v_t^j , π_t^j , l_{yt}^j , l_{gt}^j , G_t^j denote type j normalized firm value, profit, allocation of labor to production and innovation and expected growth rate respectively, then

$$v_t^j = \max_{l_{yt}^j, l_{gt}^j} \pi_t^j + e^\mu G_t^j \sum_{j'} P(\theta_{j'}|\theta_j) E_t[M_{t+1}v_{t+1}^{j'}]. \quad (18)$$

The last Bellman equation gives rise to first order conditions with respect to labor allocations. The allocation to production solves an intratemporal static problem independent of type,

$$l_{yt}^j = (1 - \alpha)^{\frac{1}{\alpha}} Z_t^{\frac{1}{\alpha}}, \quad (19)$$

and the allocation of labor to innovation solves an intertemporal problem,

$$l_{gt}^j = \left((1 - \beta)(1 - \psi_j) e^\mu \sum_{j'} P(\theta_{j'}|\theta_j) E_t[M_{t+1}v_{t+1}^{j'}] \right)^{\frac{1}{\beta}}. \quad (20)$$

Substituting the first order conditions back into the Bellman equation, we can solve for the value, labor allocation and growth of firms, by iterating over (20) and

$$v_t^j = \alpha(1 - \alpha)^{\frac{1}{\alpha}-1} Z_t^{\frac{1}{\alpha}} - \lambda + \frac{\beta}{1 - \beta} l_{gt}^j + (1 - \psi_j)(1 - \delta_j) e^\mu \sum_{j'} P(\theta_{j'}|\theta_j) E_t[M_{t+1}v_{t+1}^{j'}]. \quad (21)$$

The value of new firms can then be derived from the solution to the firm problem

$$S_t^j = E_t[M_{t+1}v_{t+1}^j] \cdot E[\tilde{K}] \cdot e^{\mu t} \quad (22)$$

3.2 Assets in Place and Growth Options

The value of firms can be decomposed into two components. The value of the *assets in place* v_{at}^j , consists of the value of the cash flows derived from the current organization capital. The value of the *growth options* v_{ot}^j consisting of the market value of the opportunities of the firm to innovate in the future.¹² Together they add up to the value of the firm,

$$v_t^j = v_{at}^j + v_{ot}^j. \quad (23)$$

The value of assets in place is stated recursively as

$$v_{at}^j = \alpha(1 - \alpha)^{\frac{1}{\alpha}-1} Z_t^{\frac{1}{\alpha}} - \lambda + (1 - \psi_j)(1 - \delta_j)e^\mu \sum_{j'} P(\theta_{j'}|\theta_j) E_t[M_{t+1}v_{at+1}^{j'}]. \quad (24)$$

Subtracting the value of assets in place from equation (21) gives an expression for the value of growth options,

$$v_{ot}^j = \frac{\beta}{1 - \beta} l_{gt}^j + (1 - \psi_j)(1 - \delta_j)e^\mu \sum_{j'} P(\theta_{j'}|\theta_j) E_t[M_{t+1}v_{ot+1}^{j'}]. \quad (25)$$

This expression highlights a feature of the model: the value of growth options is proportional to a discounted value of future labor allocations to innovation. Equation (20) states that those allocations are a convex function of the value of the firm. Therefore the shareholders evaluate the value of the growth options as leveraged claims to future firm value.

The value of growth options is more accounts for a larger share of innovative firms value than of traditional firms value. To see that, first consider that for a given stock of organization capital the value of an innovative firm is higher than the value of a traditional firm due to its higher durability. This implies, according to the first order condition for innovation (equation (20)), that innovative firms allocate more labor to innovation. In contrast, the cash flow from the existing organization capital is independent of firm type. Hence the share of growth option related value from every future period is higher for innovative firms. Following that logic, and with parameters values that are described below, the value of the growth options makes up a considerably larger share of the total value of innovative firms than of traditional firms.

3.3 Entry and Business Cycles

A recession in the model is induced by consecutive negative shocks e_t . In recessions, the TFP goes down and the MPR goes up, reducing output, employment and the value of all firms.

¹² The distinction between the value of assets in place and the value of grow option was first suggested by [Berk, Green, and Naik \(1999\)](#). In their model however, firms make a binary decision whether to implement a new project, whereas here firms make a decision regarding the quantity of innovation. In both models, the future growth decisions by firms generate leveraged claims on future values, and are thus more exposed to aggregate risk than the value of assets in place.

Entry, which depends on the value of firms, falls for all firm types. But because entry of each type of firm is determined by a separate technology, the relative value of firms affects the composition of entrants.

Intuitively, there are two opposing forces affecting the relative value of firms. The fall in TFP reduces the profits of all firms, but mean reverts in the long run. Traditional firms have a large share of their value in short-term profits, and so are strongly affected by the temporary fall in profits. Innovative firms that prioritize innovation over short-term profits, have a larger share of their value in long-term profits that will be paid when TFP mean reverts. They are thus less affected by the shocks. Therefore, holding innovation constant, the value of traditional firms declines more than the value of innovative forms in response to a fall in TFP. This effect would be the dominant effect in the risk-neutral environment that is often used in the industry dynamics literature.

Second, innovative firms have a larger share of their value in the value of growth options than traditional firms. Since the growth options are leveraged claim to the future value of firms, they are more exposed to future movement in MPR. This makes them generally more risky, since MPR is more volatile. When MPR rises, the value of growth options falls more than the value of assets in place for all firms, and the values of innovative firms falls more than the values of traditional firms due to their higher exposure to growth options.¹³ The composition of entrants in recession depends on which effect is quantitatively dominant.

4 Quantitative Assessment

This section describes the calibration of the model and evaluates its quantitative success.

4.1 Quantitative Strategy

I quantify the model at a quarterly frequency. Since the model mechanism relies on a credible valuation of risky assets and on dynamics of young firms, I supplement a standard set of growth and business cycle moments with moments of financial returns and firm size and age distribution derived from public-data on cohorts of all US firms. In particular, the calibration jointly targets a rich set of data features that include (1) high expected equity premium and high equity return volatility (2) low and stable risk-free rate (3) declining exit rate and growth rate with the age of young firms (4) volatile expenditure shares on innovation within firms.

To construct target moments I use data from several sources. I use data from Center of Research in Security Prices (CRSP) for equity returns and the risk-free rate. I use data on all US firms that were born between 1980 and 2014 from the Business Dynamics Statistics (BDS)

¹³ The growth-option value also responds to the rise TFP. However, the magnitude of this response is similar to the response of the assets-in-place value, and so moves the value of innovative and traditional firms in a similar way.

provided by the Census Bureau to construct the firm size by age distributions and moments of firm dynamics. To construct target moments for firm expenditure and profit shares I use accounting data on public firms from Compustat. Finally, I use several time series from the Bureau of Labor Statistics (BLS), including the consumer price index, labor productivity and real output to construct standard business cycle moments. When using time series, I restrict the sample period to 1980-2014 if not mentioned otherwise.

After calibrating the model, I assess it's quantitative success using three measures:

1. **Model fit.** The success of the model at matching the firm size by age distribution and firm dynamics, that are the main source of over-identifying restrictions in the calibration.
2. **Cyclical dynamics of cohorts of firms.** I simulate the model and measure the success of the model at generating cyclical properties of cohorts of young firms observed in the data.
3. **Recovery of shocks from the data.** I use the state space system implied by the calibrated model to back out the shocks and latent variables of the model from output data only. Then, I compare the time series of realized equity returns and firm entry dynamics that are implied by the model to observed data time series.

4.2 Calibration

Several parameters of the model have a direct and reasonable interpretation in the data. These parameters are presented in Table 1. Trend growth of TFP and wages μ corresponds to the mean growth rate of labor productivity in the data. I set $\mu = 0.47\%$, to get an annual growth rate of 1.88% that is observed in labor productivity timeseries from BLS. The real risk-free rate R_f is set to match the average three month treasury yield collected by CRSP and deflated by CPI, which is equal to 1.28% in the sample period (1980-2014).

Table 1: Directly calibrated parameters

parameter	value	source/target	data	model
μ	0.47%	mean labor productivity growth	1.88%	1.88%
$\log R_f$	0.32%	risk-free rate	1.28%	1.28%
α	0.30	ratio of gross profits to sales	0.30	0.30
δ_{ma}	0.075	Corrado, Hulten & Sichel (2009)		
ϕ_z	0.979	King & Rebelo (1999)		
ϕ_η	0.904	log price-to-dividend autocorrelation	0.67	0.71

The organization capital share in production α is equal in the model to the gross profit margin of all active firms (the ratio of gross profits to sales). Let gross profits be equal to $GP_{it} = A_t K_{it}^\alpha L_{yit}^\alpha - W_t L_{yit}$. Then, by substituting in the first order condition for production labor demand (equation (19)) and dividing by output I get that the ratio of gross profits to output is equal to the capital share. I use data from income statement of all public firms available from Compustat to calculate the gross profit margin. I calculate the mean gross profit margin for all firms with positive sales within each year, weighted by sales. The mean gross-profit margin has been steadily rising in the sample period, from from 24.6% in 1980 to a peak of 34.4% in 2007. I set $\alpha = 0.30$ to be in the middle of that range.

There is no consensus in the economic literature on the appropriate depreciation rate for intangible capital in general, and organization capital in particular. [Corrado, Hulten, and Sichel \(2009\)](#) collect the available evidence on depreciation rates of various kinds of intangible capital in US national accounts. They suggest values ranging from 60% per year for “brand equity” (the result of activities such as advertising) to 20% year for research and development. [Gourio and Rudanko \(2014\)](#) study another type of intangible capital – customer capital – and propose a depreciation rate of 15% per year. [Eisfeldt and Papanikolaou \(2013\)](#) also set depreciation rate of organization capital to 15% in annual terms. In my model, depreciation rates vary by duration type. However, mature firms own the bulk of all organization capital and so should correspond to the “macro” rate of depreciation. I set the depreciation rate of mature organization capital to 7.5% percent per quarter, equivalent to 27% per year, which is in the range suggested in the literature. The depreciation rates of traditional and innovative firms is set by matching cross-sectional moments described later.

There are two persistence parameters in the model: one for TFP and another for the MPR. For consistency with the business cycles literature I set the TFP persistence ϕ_z to the [King and Rebelo \(1999\)](#) rate of 0.979. In the model, the MPR is the main determinant of the price-to-dividends ratio. I set the persistence of the MPR ϕ_η equal to 0.904 to match the persistence of the price to dividend ratio of S&P 500 according to CRSP calculation.

The rest of the parameters are presented in [Table 2](#). The exit rates are set to match the mean exit rate by age in BDS data. I simulate the exit rate of firms with two levels of exit rates, one for the traditional firms and one for the innovative and mature firms, and allow the entry shares to adjust to minimize the distance to the mean exit rate by age in the data, for 9 age groups. I then pick the exit rates that minimize the equally weighted squared distance. Exit rates are 1.4% per quarter for long-lived and high-growth firms and 10.4% for short-lived firms.

Parameters of firm investment technology (B, β, λ) and the process for the market price of risk $(\bar{\eta}, \sigma_\eta)$ are jointly set to match moments of firm growth rate, expenditure on innovation and equity returns. The firm growth rate target is the mean employment growth of 20-25 year old firms. I match that with the mean growth rate of mature firms, since they make up the vast majority of firms at this age class. Similarly, I target the mean and variance of

Table 2: Matched parameters

parameter	value	target/source	data	model
$\psi_{in} = \psi_{ma}$	1.4%	exit rate of 20-25 year-old firms	4.8%	5.5%
ψ_{tr}	10.4%	exit rate of in first year	17.9%	17.8%
B	0.174	mean growth rate of 20-25 year-old firms	0.8%	0.8%
λ	0.037	mean EBIT/GP of public firms	0.37	0.38
β	0.74	std. EBIT/GP of public firms	0.040	0.039
$\bar{\eta}$	0.26	mean excess return (all market value weighted)	7.84%	7.84%
σ_{η}	0.086	std. excess return (all market value weighted)	16.0%	16.0%
δ_{tr}	0.056	firm size by age distribution (63 moments),		
δ_{in}	0.028	see main text for description		
$P(\theta_{ma} \theta_{in})$	0.092			
ν_{tr}	10×10^3			
ν_{ma}	2×10^3			
ν_{in}	3×10^3			
$\tilde{\mu}$	1.32			
$\tilde{\sigma}$	0.76			
$\tilde{\xi}$	1.43			
σ_k	0.027			
σ_z	0.33%	volatility of hp-detrended output	1.62%	1.65%
γ	0.58	volatility of number of entrants	9.8%	9.7%

the ratio of earnings before interest and taxes (EBIT) to gross profits. I compute the sales weighted mean of the ratio for public firms for every year in the sample period. I then take the mean over all years (0.37) and the standard deviation (0.04) as targets. I match them with the unconditional mean and standard deviations of the equivalent measure for mature firms, computed as Π_{it}/GP_{it} . Equity returns in the model are returns on a diversified portfolio of mature firms. Since equity in the data is levered, I calculate the model implied equity returns at a constant leverage of 0.35 (ratio of debt to firm value), as suggested by [Belo, Collin-Dufresne, and Goldstein \(2015\)](#).

I set firm specific parameters, including depreciation rates for traditional and innovative firms (δ_{tr}, δ_{in}), transition rate between innovative and mature firm types $P(\theta_{ma}|\theta_{in})$, entry technology parameters ($\nu_{tr}, \nu_{in}, \nu_{ma}$), parameters of the initial organization capital distribution ($\tilde{\mu}, \tilde{\sigma}, \tilde{\xi}$) and the parameter of the idiosyncratic random shock to capital σ_k , by matching the distribution of firm size by age distribution. I construct the target moments by normalizing each cohort by the number of firms one year after entry and calculate the relative number of

firms in each size by age cell. I consider 9 age groups and 7 size groups.¹⁴ I then take the mean over all cohorts, and multiply all numbers by the mean number of firms one year after entry (this way the number of firms in the model roughly matches the number of firms in the US economy). I then simulate the size distribution of firms when shocks are set to their mean, and minimize the equally weighted percentage difference from the data moments.¹⁵ Overall, I set 10 parameters by matching 63 moments.

A final step sets the values of two key parameters: the volatility of TFP and the elasticity of entry. Since they affect the calibrated values of all other parameters, I set them by guessing the value and then verifying the match in simulated data. I set the volatility of TFP $\sigma_z = 0.33\%$ by matching the simulated and observed volatility of simulated hp-filtered real output, that is 1.64%. Finally, I set the elasticity of entry $\gamma = 0.58$ by matching the simulated and observed volatility of total entry, which is equal to 9.8% in the data.

4.3 Model Fit to Firm Distribution and Firm Dynamics

I compare unconditional simulated model moments to data moments from BDS. I simulate the model for 5000 quarters, discard of the first 1000 quarters, then use the simulation to construct the simulated moments. In the simulation, I keep track of each cohort full size and type distributions.

I first compare the size distribution of firms by age group. The BDS data is binned at predetermined age and size groups. For the comparison, I use 9 age groups (0,1,2,3,4,5,6-10,11-15,16-20) and 7 employment size groups (1-4,5-9,10-19,20-49,50-99,100-249,250+). The simulation is at a quarterly frequency, so I first construct the number of firms in every age group by summing up over all ages that fall into a bin. For instance, age “0” in the data corresponds to firms that entered in the previous 4 quarters, and are consequently 0-3 quarters old, and age “16-20” corresponds to all firms that entered between 65 and 84 quarters ago. I compute the distribution on a fine grid, by adding the measure of firms at a given size and given age group over the simulated sample, then dividing by the total number firms in that age group.

I present the fit of the distribution visually in figure 1. The graph shows the share of firms that have more than a given number of employees (the complimentary cumulative distribution function). Each line in the graph represents the simulated unconditional size distribution of firms at different age bin. Markers represent the data moments. For clarity, I present the fit for

¹⁴ Age groups (0,1,2,3,4,5,6-10,11-15,16-20) and employment size groups (1-4,5-9,10-19,20-49,50-99,100-249,250 and above).

¹⁵ Discrepancies could arise between the moments generated by this simulation and moments generated by a simulation with random stochastic draws. To deal with that, I simulate the model with the random stochastic draws and verify that the moments are not too different. Indeed, the difference between the two methods of simulation is not big, and is concentrated at the largest age group, where it deviates by 7 percent compared to the steady state moments.

6 of the age group. The share of large firms increases with the age of cohorts. This is happening throughout the size distribution, and in particular in the largest size group (more than 250 employees). The model implied distributions have Pareto tails that become heavier with the age of the cohort. The model does a remarkable job at fitting 63 detailed firm distribution moments, despite having only three duration types and no age or size dependent technology, and with only 10 degrees of freedom.

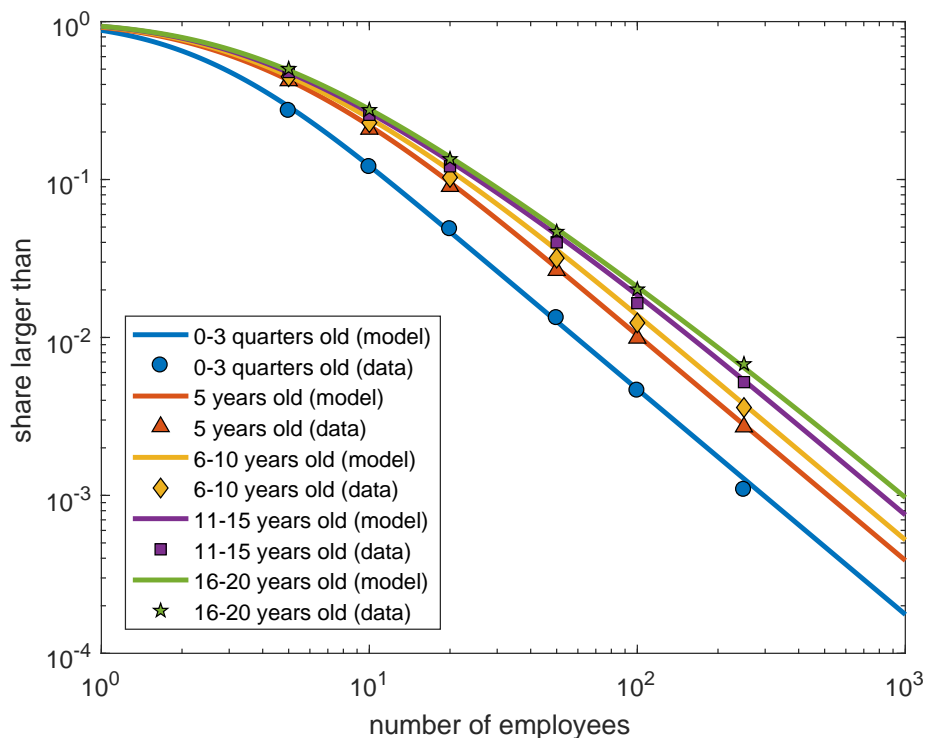


Figure 1: Fit of size distribution

Notes: Horizontal axis is the employment size of firms. Vertical axis is the complimentary distribution function (share of firms that are larger). Markers show size distribution of all firms that employ at least one worker for different age groups. Data moments are based on Business Dynamics Statistics (BDS) data provided by the Census Bureau. Lines are the model implied distribution for the same age groups. See main text for details.

I then compare exit rates and growth rates. I construct the exit rates from the data using the following formula:

$$exit_{a,t} = firmsdeath_{a,t} / (firms_{a,t} + firmsdeath_{a,t}),$$

where $firms_{a,t}$ is the number of firms of age group a at year t , and $firmsdeath_{a,t}$ is the number of firms that were in the data that exited in the last 12 months, and would have been at age group a otherwise. I then take the unweighted average over all years in the sample as data

moments, which makes sure I do not overweight large cohorts.

I construct the employment growth rates of continuing firms using the following formula:

$$growth_{a,t} = njcc_{a,t} / (emp_{a,t} - njcc_{a,t}),$$

where $emp_{a,t}$ is total employment of firms of age a at year t and $njcc_{a,t}$ is the net jobs created by continuing firms that are now at age group a . I also take their unweighted average as the data moments. I construct the simulated moments in the exact same way.

Table 3 presents the comparison of data and model in exit rates and employment growth rates. As opposed to the firm size by age distribution, these moments are not directly targeted, yet the simulated exit rates and growth rates closely follow the exit rates and growth rates from the data. Exit rates and growth rates fall quickly in the first five years, and settle after cohorts are 10 years old. This pattern is captured in the model by the combination of heterogeneous exit rates and endogenous growth rate of the three duration types. A noticeable exception to the fit is in the growth rate of 1 year old firms (between the first observation of the firm and the second). I explain this discrepancy with the possibility of measurement error of firm employment at the first observation.¹⁶

4.4 Firm Cohort Cyclical Properties

The model makes predictions on the cyclical properties of cohorts of young firms: cohorts of firms that enter in recessions have fewer firms, and a smaller share of those firms are innovative firms. The first prediction can be observed in the data. I regress the natural logarithm of the number of one year old firms on real GDP growth one year earlier, when the cohort of firms entered, and on a linear time trend. Table 4 Panel A first column reports the result. Total firm entry is pro-cyclical. Cohorts that enter when output growth is 1% lower have 2.4% fewer firms.

The second prediction is harder to test since the number of innovative firms is not directly observed.¹⁷ Instead we can observe the relative size distribution of firms in those cohorts. Cohorts that have a larger share of innovative firms will have a larger share of large firms when observed one year after entry. Table 4 Panel A second and third column test this prediction. The number of large firms, with more than 100 employees, one year after entry goes down by 5.1% when GDP growth is 1% lower, more than the total number of firms. The difference between the coefficients is also the coefficient when the dependent variable is the share of large firms, and is significant at 0.01 level despite having only 34 observations.

¹⁶ This explanation has some support in the documentation of BDS. For instance, in FAQ it says: “The BDS also excludes most very large single unit births (age 0 firms with only one establishment and more than 2499 employees) both from the entry measures (job creation) and from current employment in the birth year and all future years. “ Systematic exclusion of firms based on size at entry but not when 1 year old may lead to upward bias in estimates.

¹⁷ The type of firm could potentially be inferred from panel data at the firm level.

Table 3: Exit rates and employment growth rates of continuing firms by firm ages

age	exit (%)		growth (%)	
	model	data	model	data
1	17.9	17.9	6.8	12.7
2	15.2	13.7	5.3	5.3
3	12.9	11.8	4.1	4.2
4	11.0	10.5	3.2	3.4
5	9.5	9.5	2.5	2.8
6-10	7.1	7.7	1.5	1.8
11-15	5.7	6.1	0.9	1.2
16-20	5.5	5.3	0.8	0.9
21-25	5.5	4.8	0.8	0.8

Notes: Model and data implied exit rates and growth rates of continuing firms by the age of the firm. Numbers are in percentage of existing firms/employment. Data exit rates and growth rates are based on the Business Dynamics Statistics data provided by the Census Bureau for the period 1980-2014. Model exit and growth rates are calculated for TFP and MPR set to mean. See main text for details.

I evaluate the quantitative model by replicating this regression with simulated data. I use the same simulated sample described above and regress each annual observation on the growth of output over the previous years.¹⁸ Table 4 Panel B presents the results. The model is successful in capturing the magnitude of the effects, despite the regression coefficients not being a target moments (the only target moments that captures entry dynamics is the volatility of entry, which is used to set the elasticity coefficient γ in the startups technology). Simulated coefficient on large firms and the difference are within standard confidence intervals distance from the data coefficients. The simulated coefficient on total entry (Panel B first column) is lower and significantly different, but on the same order of magnitude as the data coefficient.

4.5 Estimated Firm Types

Before moving into the aggregate implications of the model, I first look at some of the characteristics of the different firm types. Table 5 summarizes the main characteristics. Panel A shows the dynamic properties of different firm types. Traditional firms exit at a high rate of 36 percent per year, implying a firm life expectancy of 2.8 years. This captures the high turnover in young firms. Innovative and mature firms exit at a rate of 5.5 percent per year, which implies a firm life expectancy of 18 years. Traditional firms also allocate less labor to

¹⁸That is growth in annual GDP is $(Y_{t-1} + Y_{t-2} + Y_{t-3} + Y_{t-4}) / (Y_{t-5} + Y_{t-6} + Y_{t-7} + Y_{t-8}) - 1$.

Table 4: Cyclical properties of firm cohorts

dependent variable: log number of one year old firms						
	<i>Panel A: data</i>			<i>Panel B: simulation</i>		
	all	large	diff.	all	large	diff.
real GDP growth at entry	2.39	5.10	2.71	1.69	5.06	3.38
(standard errors)	(0.32)	(1.18)	(1.07)	–	–	–
linear time trend	yes	yes	yes	–	–	–
number of observations	34	34	34	1000	1000	1000
R^2	0.37	0.29	0.13	0.21	0.36	0.44

Notes: Regression results for firm-cohort data and firm-cohort model simulations. Dependent variables are the logged total number of firms (all), the logged number of firms with more than 100 employees (large), and the difference (diff.) when the cohort is one year old. “real GDP growth at entry” is the logged annual real GDP growth at the cohort’s year of entry (t-1). Panel A shows regression results based on the Business Dynamics Statistics data provided by the Census Bureau for the period 1980-2014 and BLS data for GDP growth. Data regression includes a linear time trend. Standard errors (in parentheses) are calculated using Newey-West estimator with 5 lags. Panel B shows regression results based on 1000 years of simulated data. See main text for details.

innovation and shrink by 2 percent per year on average, while innovative firms grow at a rate of 28 percent per year. Mature firm size is relatively stable, and they grow at a rate of 0.8 percent per year. The differential growth rates are reflected in the profits of firms. Traditional firms collect almost 20 percent of their revenue as profits. In the other extreme, innovative firms keep only 1 percent of their revenue on average, and are often generating negative profits. When their growth potential declines and they become mature firms, the profit shares jumps to 11 percent. It is also interesting to see their impact on the economy. Innovative firms make up 37 percent of entering cohorts on average, but make up only 16 percent of the population of firms.

Panel B shows the properties of unlevered returns. Returns are calculated for a diversified portfolio of firms of each type and take into account exit. The mean excess returns on traditional firms is 2.5 percent, which is substantially lower than the returns on innovative and mature firms. The excess returns on innovative firms is 6 percent per year and on mature firms 5 percent per year. The standard deviations of returns change proportionately to the mean excess returns, so that Sharpe ratios are almost the same.

Table 5: Estimated firm types

	firm type		
	traditional	innovative	mature
<i>Panel A: firm dynamics</i>			
annual exit rate	35.6%	5.48%	5.48%
mean employment growth rate	-2.10%	28.2%	0.80%
mean profit share of output	19.5%	1.16%	11.1%
mean entry shares	44.4%	37.6%	17.9%
mean population shares	18.4%	15.7%	65.9%
<i>Panel B: unlevered returns</i>			
mean excess return	2.55%	6.06%	5.10%
s.d. returns	5.07%	12.6%	10.5%
unconditional Sharpe ratio	0.50	0.48	0.49

4.6 The Term Structure of Equity

Recent literature has challenged existing asset pricing models by providing evidence on the term-structure of equity risk premia. In an influential paper, [van Binsbergen, Brandt, and Koijen \(2012\)](#) study the pricing of dividend-strips: claims to dividends at a specified interval in the future. They show that dividend strips on one and two year claims on S&P 500 have had higher mean excess returns than the underlying index, which is unlikely in typical asset pricing models. [van Binsbergen, Hueskes, Koijen, and Vrugt \(2013\)](#) and [van Binsbergen and Koijen \(2017\)](#) provide additional evidence on other stock indices and more maturities that the excess returns on dividend strips is declining with maturity.

Here, I show that the model is consistent with downward sloping risk premia for mature firms. I base my argument on leverage dynamics, based on a similar argument by [Belo, Collin-Dufresne, and Goldstein \(2015\)](#). For notation clarity, I suppress the firm and type indexes i and j .

The starting point is a claim to a firm's future profits starting at period $n + 1$, that is a claim on $\{\Pi_{t+n+1}, \Pi_{t+n+2}, \dots\}$. Let $S_{t,n}$ be the value of a future profits claim of maturity n .¹⁹ Naturally, the value of $S_{t,0}$ is equal to the value of the firm net of time t profits, $S_{t,0} = V_t - \Pi_t = \hat{V}_t$. The spot value of future profits claims of maturity n can be calculated by recursion,

$$S_{t,n} = E_t[M_{t+1}S_{t+1,n-1}]. \quad (26)$$

¹⁹ This is a theoretical definition that can be seen as the spot price of a futures or forward contract on the value of the firm.

Let $s_{t,n} = S_{t,n}/K_t W_t$ be the normalized price of the claim. Then the initial condition is $s_{t,0} = v_t - \pi_t = \hat{v}_t$ and the recursion equation is

$$s_{t,n} = e^\mu G_t E_t[M_{t+1} s_{t+1,n-1}], \quad (27)$$

where G_t is the expected growth of a mature firm at time t . The expected one period returns on this claim $X_{t,n}^s$ is then

$$X_{t,n}^s = e^\mu G_t \frac{E_t[s_{t+1,n-1}]}{s_{t,n}}. \quad (28)$$

The claim for future profits can be used to construct the values of other asset types and their expected returns. Let $p_{t,n}^\pi$ be the normalized value of a claim to profits π_{t+n} . Then, $p_{t,n}^\pi = s_{t,n-1} - s_{t,n}$. The expected returns $X_{t,n}^\pi$ on a profit claim can be expressed in terms of $s_{t,n}$ and its expected return $X_{t,n}^s$,

$$X_{t,n}^\pi = \frac{X_{t,n-1}^s s_{t,n-1} - X_{t,n}^s s_{t,n}}{s_{t,n-1} - s_{t,n}}. \quad (29)$$

The solid line in Figure 2 shows the risk premia for profit strips, which are their unconditional expected excess returns $E[X_{t,n}^\pi]$. The horizontal axis shows the maturity of the strip in years and the vertical line the annualized expected returns. The term structure of risk premia for profit strips in the quantified model is increasing in maturity. For reference, I draw the unconditional expected returns on levered equity as the dotted horizontal line. The return on equity is higher than the returns on maturities less than 10 years. The intuitive reason is that firms respond to shocks by adjusting innovation, making short term profits less volatile and future profits more volatile.

This may seem at odds with the evidence from dividend strip prices. However, the evidence in the literature is on dividends and not on profits. Since the model is set in complete markets, firms are indifferent between issuing debt and equity, and can choose any payout policy. [Belo, Collin-Dufresne, and Goldstein \(2015\)](#) show in a different setup, that a simple dividend payout policy that keeps the leverage ratio stationary is consistent with the data, and can generate the downward sloping term-structure of risk premia in standard asset pricing models.

I specify the following payout policy, based on a leverage target \mathcal{L} . Each period, firms repay a fraction $1 - \tau$ of their outstanding debt B_{t-1} , and issue a constant fraction of their ex-profits value in risk-free debt, $\mathcal{L}\hat{V}_t$. The law of motion for debt is then,

$$B_t = \mathcal{L}\hat{V}_t + \tau B_{t-1}, \quad (30)$$

and the dividend payments D_t follow

$$D_t = \mathcal{L}\hat{V}_t - (R_f - \tau)B_{t-1} + \Pi_t. \quad (31)$$

The dividend payment at $t + n$ can be expressed as a weighted sum of future values of the firm,

$$D_{t+n} = \mathcal{L}\hat{V}_{t+n} - (R_f - \tau) \sum_{j=1}^n \tau^{j-1} \mathcal{L}\hat{V}_{t+n-j} - (R_f - \tau)\tau^n B_{t-1} + \Pi_{t+n}. \quad (32)$$

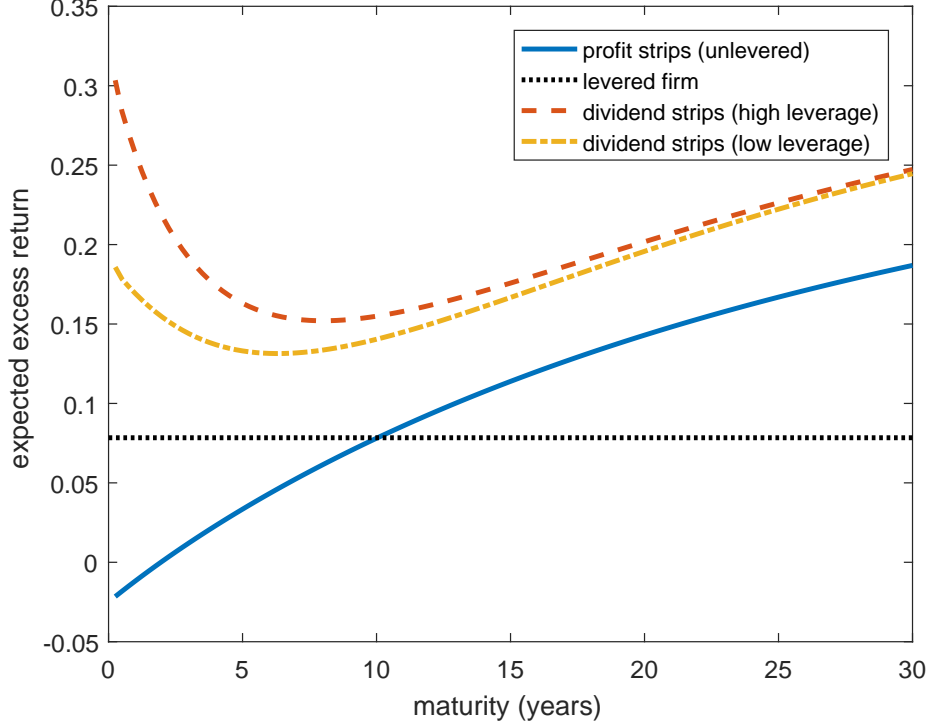


Figure 2: Term structure of risk premia – profit and dividend strips

Notes: Unconditional risk premia (annualized expected returns minus the risk free rate) of profit and dividend strips by maturity. Strips are claims to one quarter of profits or dividends. Horizontal axis shows the time to maturity of the strip in years. Solid line shows the risk premia on profit strips. Dotted horizontal line shows the unconditional risk premia on firm equity (7.84 percent). Dashed and dashed-dotted lines show the risk premia on dividend strip at high (leverage ratio 0.4) and low (leverage ratio 0.3) leverage states of the firm.

The value of a dividend strip of maturity $n \geq 1$, $P_{t,n}^d$ can then be written as as a weighted sum of future claims and previous debt,

$$P_{t,d}^d = \sum_{j=0}^n q_{j,n} S_{t,j} - (R_f - \tau)(\tau/R_f)^n B_{t-1} \quad (33)$$

where $q_{j,n}$ are constant functions of \mathcal{L} and τ (see Appendix A for exact expression). The normalized value $p_{t,n}^d = P_{t,n}^d/K_t W_t$ is similarly expressed as

$$p_{t,n}^d = \sum_{j=0}^n q_{j,n} s_{t,j} - (R_f - \tau)(\tau/R_f)^n e^{-\mu} G_{t-1}^{-1} b_{t-1}, \quad (34)$$

where $b_t = B_t/K_t W_t$ is the normalized debt. Finally, the expected returns on dividend strips of maturity n can be found using the expression

$$X_{t,n}^d = \frac{\sum_{j=0}^{n-1} q_{j,n-1} X_{t,j+1}^s s_{t,j+1} - (R_f - \tau)(\tau/R_f)^{n-1} (\mathcal{L} s_{t,0} + \tau e^{-\mu} G_{t-1}^{-1} b_{t-1})}{p_{t,n}^d}. \quad (35)$$

I plot the expected excess returns on dividends strips for $1 - \tau = 1/15$ and $\mathcal{L} = 0.021$ to get an average debt maturity of 3.75 years, and an unconditional mean leverage ratio $E[B_t/\hat{V}_t] = 0.35$ as suggested by [Belo et al. \(2015\)](#). I fix two debt levels b_{t-1} to capture high and low leverage cases. In high debt the leverage is 0.4 on average, and in low debt it is 0.3 on average, where averages are taken over 1000 years simulation. The dashed and dashed-dotted lines show the term structure of unconditional risk premia on dividend strips. The high leverage risk premia (dashed) has a steep downward sloping path in the first 7 years. The low leverage risk premia (dashed-dotted) is less steep and starts climbing earlier. This exercise illustrates that the model is consistent with a downward sloping term structure of risk premia.

4.7 Extracting Shocks from Output Data

In this section I use the state space system of the quantified model to recover the latent variables in the model from a time series of output growth. Given the the aggregate state vector $(Z_{t-1}, \eta_{t-1}, \{K_{t-1}^j\})$ and the output growth $\Delta \log Y_t$ there is a unique solution for the random shock e_t , and hence a solution for the state vector $(Z_t, \eta_t, \{K_t^j\})$, and other outcomes including employment L_t , entry $\{N_t^j\}$ and realized returns on assets.

I use the real GDP growth for the US over the period 1979Q1:2016:Q4. I remove a linear trend from the natural log of GDP, then take first difference to get the time series $\Delta \log \hat{Y}$. I start by guessing that TFP, MPR and the stocks of capital are at their unconditional means at 1978Q4. Given the estimated state vector $(\hat{Z}_{t-1}, \hat{\eta}_{t-1}, \{\hat{K}_{t-1}^j\})$ and the law of motion for aggregate organization capital stocks, I construct $\hat{K}_t = \sum_j \hat{K}_t^j$. Then the estimated random shock \hat{e}_t is equal to

$$\sigma_z \hat{e}_t = (1 - \phi_z) \hat{Z}_{t-1} + \alpha (\Delta \log \hat{Y}_t - \Delta \log \hat{K}_t). \quad (36)$$

Equation (36) does not contain a term for the trend growth μ because trend growth has already been removed from the data time series for $\Delta \log \hat{Y}$.

Figure 3 shows the estimated process for TFP. The TFP process falls sharply in every recession, reaching lows of -0.015 in both the 1982 recession and the Great Recession. It reached a peak of 0.01 at 2000Q2. The standard deviations of \hat{e}_t are 0.62, smaller than the assumed standard normal distribution.

Next, I use the estimated state variables to construct time series for hours, entry and returns. These implied time series can be compared to observable time series to evaluate the success of the model.

4.8 Hours, Entry and Returns

I construct the implied time series of hours using the expression for total labor,

$$L_t = \sum_j (l_{yt}^j + l_{gt}^j + \lambda) K_t^j + L_{st}^j.$$

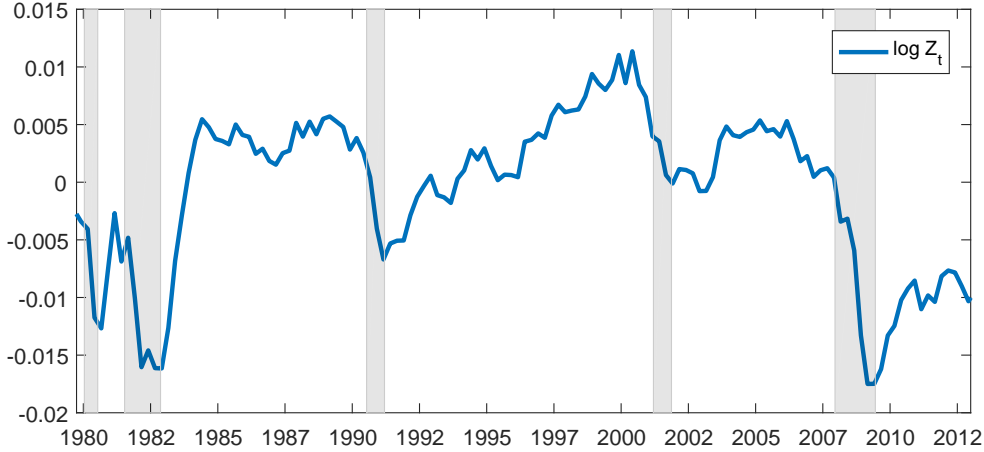


Figure 3: Estimated TFP process

Notes: Estimates of TFP ($\log \hat{Z}_t$) that exactly match log-linearly detrended output for the period 1979:Q1 to 2016:Q4. Gray bars represent NBER recessions. See main text for details on procedure.

The data target for hours is the product of the civilian employment-population ratio, and the average weekly hours in nonfarm business sector, both from the BLS.

Figure 4 presents the percentage deviation from the mean of the two time series. The hours implied by the model are successful at capturing the shape and the timing of fluctuations in labor supply, both at business cycle frequency and over the medium and long terms. The hours implied by the model are somewhat more volatile than the hours in the data. These features come mainly from the assumption that real wages follow a deterministic trend, and the high correlation between output and hours in the data.

To capture the cyclical properties of cohorts of startups, I compare the model implications for the number of one year old firms that are small (less than 100 employees) and large (more than 100 employees). This is the time series visual equivalent to the regression evidence above. For each cohort I simulate the entrants full size distribution then calculate the number of firms in each size category. I the implied number of firms to a time series constructed using the cohort data in the BDS, and based on the start year of the firm.

Figure 5 presents the percentage deviation of the number of small one year old firms (Panel A) and the number of large one year old firms (Panel B), implied by the model and in the data. The model implied entry is falling in recession and rising in recoveries. The model also captures the magnitude of the fluctuations.

Lastly, I compute the realized returns on a diversified portfolio of mature firms. The implied unlevered returns \hat{R}_{t+1} are equal to the trend growth e^μ times the expected firm organization capital growth \hat{G}_t times the ratio of the normalized value \hat{v}_{t+1} to time t normalized ex-profit

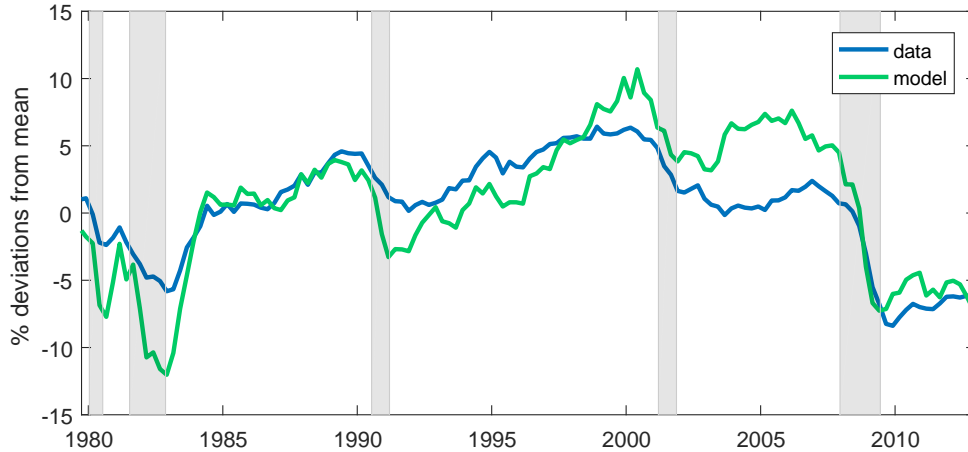


Figure 4: Total hours, model vs. data

Notes: Estimated and observed total hours worked at quarterly frequency logged and demeaned. Dark blue line shows the hours from the data. Total hours series constructed as the product of civilian employment-population ratio and average weekly hours in nonfarm business sector, both from BLS. Light green line shows model implied hours worked based on recovered shocks. See details in main text.

value $\hat{v}_t - \hat{\pi}_t$,

$$\hat{R}_{t+1} = e^{\mu} \hat{G}_t \hat{v}_{t+1} / (\hat{v}_t - \hat{\pi}_t).$$

The levered excess returns $\hat{X}R_{t+1}$ are computed using with a fixed leverage $\bar{\mathcal{L}} = 0.35$, so that $\hat{X}R_{t+1} = (\hat{R}_{t+1} - R_f) / (1 - \bar{\mathcal{L}})$.

For comparison I use CRSP realized value weighted returns on the market at quarterly frequency minus 3-month treasury yield. For visual clarity, I filter both time series with 4 quarter standard moving average.

Figure 6 shows the model implied and the data time series of realized excess returns. The model implied realized equity returns exhibit long booms between recessions and sharp busts in every recession in the sample period, with similar magnitudes as in the data. This is an unexpected success, because the parameters are only chosen to match unconditional moments, and the shocks are extracted without taking into account any financial time series. It emphasizes the importance of countercyclical risk premia in explaining aggregate fluctuations in both asset prices and aggregate quantities.

4.9 Decomposition of Output and Hours

Fluctuations in output come from the direct effect of TFP, which moves output for a given quantity of organization capital, and the indirect effect of TFP and MPR through innovation, which changes the aggregate growth rate of organization capital. I Decompose the fluctuations

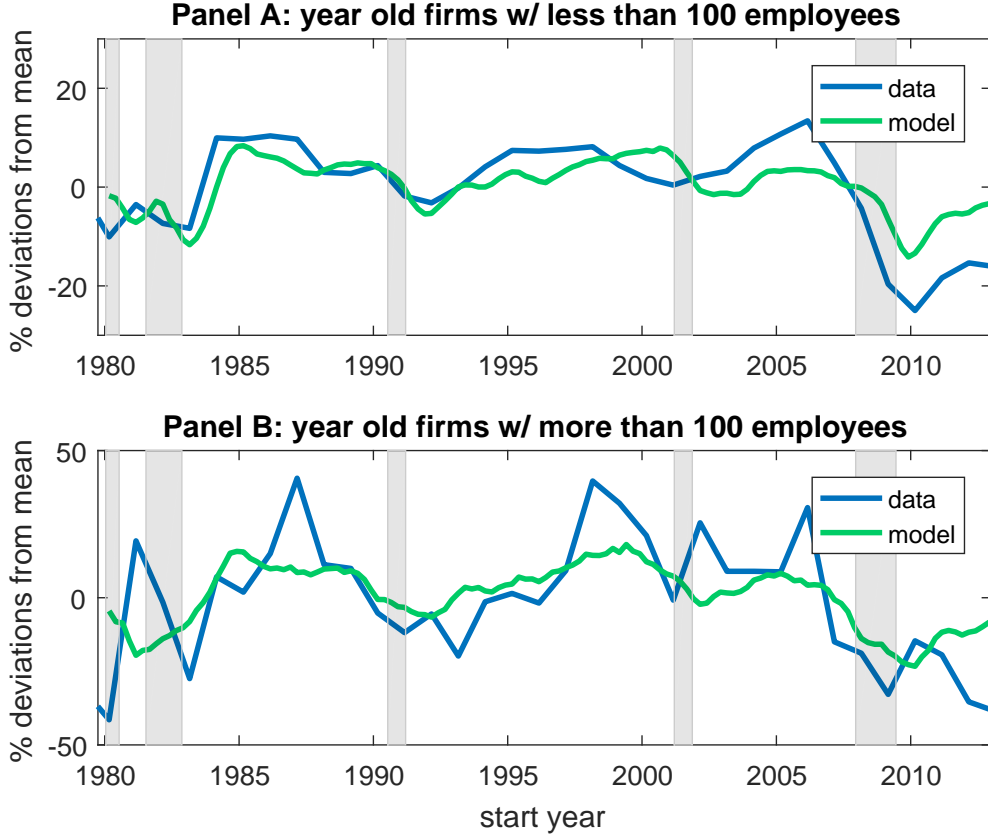


Figure 5: Firm entry, model vs. data

Notes: Estimated and observed number of firms in one year old cohorts. Horizontal axis shows the year of entry. Vertical axis is the deviation from mean in percentage points. Dark blue line shows the observed number of one year old firms in the Business Dynamics Statistics data. Light green shows the number of one year old firms in the model with estimated shocks as described in the main text. Gray bars are NBER recessions. Panel A presents the time series for firms with less than 100 employees. Panel B presents the time series for firms with more than 100 employees.

in output to the direct effect of TFP and the indirect effects through innovation by keeping one channel and shutting of the other two.

I construct counterfactual series for output using the estimated shocks \hat{Z}_t , but without innovation and fluctuations in organization capital to measure the direct effect of TFP. The constructed time series of output is then

$$\log \tilde{Y}_t^1 = \mu t + \log \bar{K} + \left(\frac{1}{\alpha} - 1\right) \log(1 - \alpha) + \frac{1}{\alpha} \log \hat{Z}_t, \quad (37)$$

where \bar{K} is the mean aggregate stock of organization capital. The direct effect of TFP is then just the fluctuation in $\log Z_t$ scaled by $1/\alpha$ which is equal to 3.33 in the quantified model.

I measure the indirect effect of TFP through innovation in two steps. First, I construct counterfactual entry \hat{N}_t^j and firm growth \hat{G}_t^j using the policy function of firms, and based on

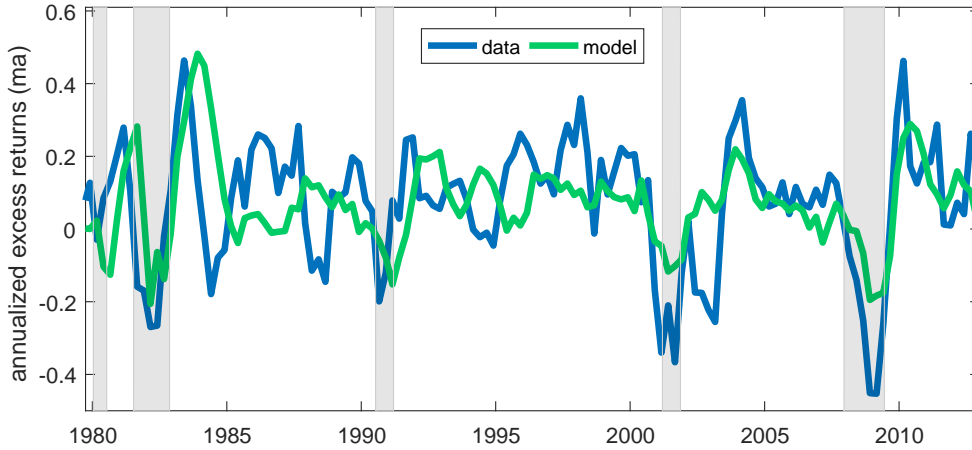


Figure 6: Realized excess returns, model vs. data

Notes: Realized excess returns on stocks (data) and levered mature firms (data). Horizontal axis shows the annualized returns minus the risk free rate. Dark blue line shows the annualized quarterly returns on value weighted portfolio of all stocks from CRSP (VWRETD) minus the three-month T-Bill yield. Light green shows the annualized quarterly returns on a diversified portfolio of mature firms minus the risk-free rate implied by the model. The two series are filtered with a standard moving average with 4 lags. Gray bars represent NBER recessions.

the estimated value of TFP, while MPR is set to mean value, $\eta_t = \bar{\eta}$. Then, I construct the counterfactual series of organization capital and output as there are no fluctuation in TFP and MPR. I measure the indirect effect of MPR in a similar way, but with the estimated MPR and TFP set to mean value in the first step.

Figure 7 presents the decomposition of detrended output into the three effects. The solid line is the linearly detrended output. The dotted line is the direct effect of TFP. The direct effect captures the high frequency movements in output. The dashed and dash-dotted lines are the effects of TFP and MPR on innovation respectively. The effects through innovation are much smoother than the direct effect, yet they still generate a large quantity of the fluctuations. The thin green line horizontal line is the residual when the three effects are taken out of the detrended output. The small size of the residual suggests that the interaction in innovation policy between TFP and MPR is not quantitatively important.

What is the share of the fluctuations that can be accounted for by fluctuations in MPR? The standard deviations of output without the effect of MPR (direct effect of TFP + effect of TFP on innovation) are 2.89 percent. The standard deviation of the detrended output is 4.64 percent. Therefore I conclude that fluctuations in MPR account for $100\% - 2.89/4.64 = 38\%$, or around two fifths, of the fluctuations in output in the sample period.

I perform a similar exercise with labor. An important difference is that I decompose the model implied labor demand and not the data on hours. Another important difference between the decomposition exercises is that innovation activity directly affects labor.

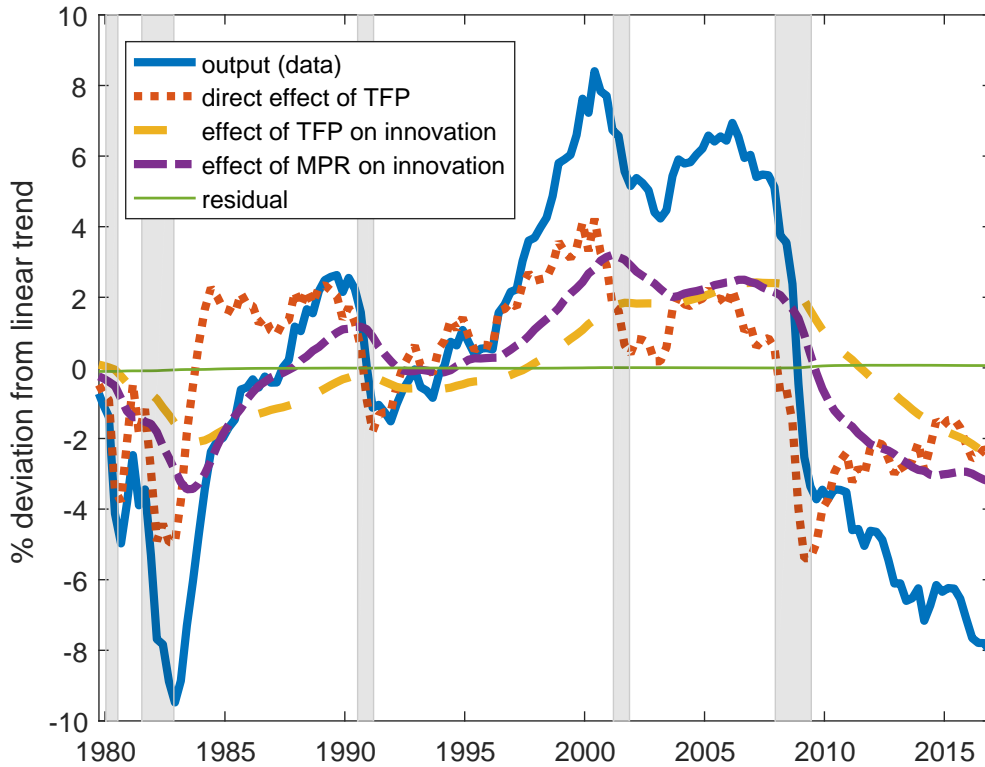


Figure 7: Decomposition of output

Notes: Decomposition of output time series. Horizontal axis is percent deviation from log-linear trend. Solid thick line shows the real GDP from BLS. Dotted line shows the direct effect of TFP through production and the allocation of labor to production. Dashed line shows the effect of TFP on through changes in innovation rates. Dashed-dotted line shows the effect of MPR through changes in innovation rates. Thin line shows the residual after accounting for the three effects. See details in the main text.

Figure 8 shows the decomposition of implied hours. The thick solid line is the implied hours from the extracted shocks. The dotted line is the direct effect of TFP. This effect is smaller than the direct effect of TFP on output because the allocation of labor to production is only one out of four different activities (production, management, innovation in firms, startup activity). The effect of MPR through innovation contains both the impact on the quantity of organization capital, and the allocation of labor to innovation.

This provides a new theory for the volatility of hours. When the market price of risk declines, firms increase the allocation of labor to innovation more than the allocation to production. This amplifies the initial shock and contribute to business cycle frequency fluctuations in employment. The amplification mechanism also works in the opposite direction: MPR fell sharply during the Great Recession leading to a decline in firm innovation activity and in

startup activity, contributing to the decline in aggregate employment and hours.

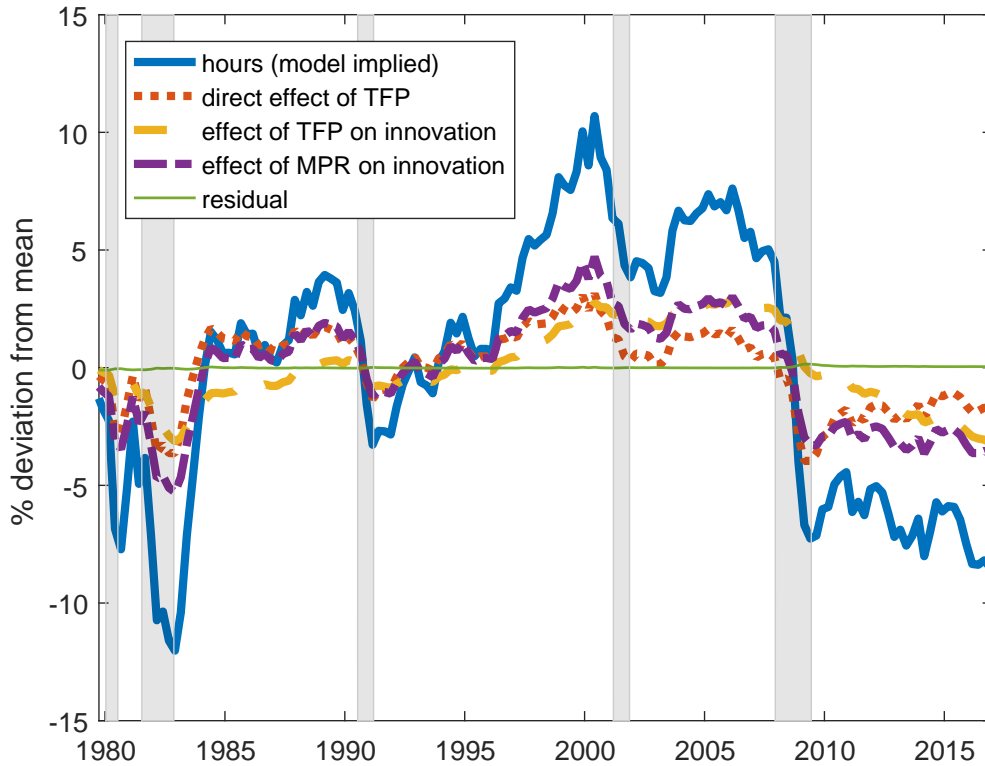


Figure 8: Decomposition of hours

Notes: Decomposition of hours time series. Horizontal axis is percent deviation from log-linear trend. Solid line shows the hours implied by the model. Dotted line shows the direct effect of TFP through the allocation of labor to production. Dashed line shows the effect of TFP through changes in innovation rates and allocation of labor to innovation. Dashed-dotted line shows the effect of MPR through changes in innovation rates and allocation of labor to innovation. Thin line shows the residual after accounting for the three effects. See details in the main text.

4.10 Application: the Great Recession

An application of the quantified model is to evaluate the long-term impact of the Great Recession. I conduct a counterfactual exercise to answer the following question: what would be the real output in 2016Q4 if innovation activity would not have declined during the Great Recession? To do that, I replace the entry and growth rate implied by the model with the mean values of the rates in the period 2007Q4:2010Q1, and reconstruct the time series of output from that period to the end of the sample at 2016Q4.²⁰

²⁰ 2007Q4 is the official start of the Great Recession. While unemployment peaked at 2009Q2, the official end of the Great Recession, aggregate employment reached a trough only at 2010Q1.

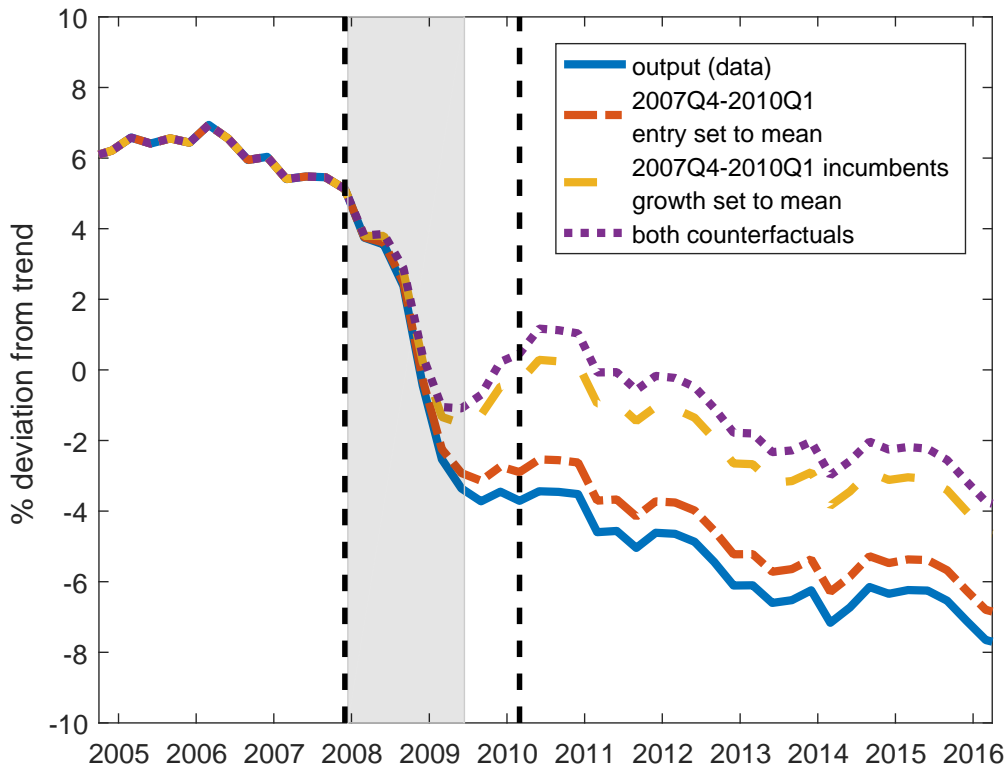


Figure 9: Counterfactual scenarios for the Great Recession

Notes: Counterfactual exercise. Horizontal axis shows percent deviation from log-linear trend. Solid line shows real output from the data (BLS). Other lines show the effect of replacing the innovation during the Great Recession with unconditional mean values: entry set to mean (dashed), growth rate of incumbents set to mean (dashed-dotted) and both effects (dotted). Vertical lines show the counterfactual experiment period. All other innovation/production rates are the same across the lines.

Figure 9 presents the results of the exercise. The solid line is the detrended output from the data. The two vertical dashed lines indicate the treated period, where the entry and growth of incumbents has been set to their long-term means. The drop in GDP accounted for by the lack of innovation in the Great Recession is substantial: 4.4% at the end of the recession and 3.6% at 2016Q4, six and a half years later. To further investigate the mechanism, I also construct counterfactuals with just the entry set to the mean (dashed) and just the innovation by incumbents set to the mean (dashed-dotted). The fall in entry accounts for 1% loss in GDP and the fall in incumbent firms innovation accounts for 2.6% loss in GDP. I interpret this as indication for an important role for the impact of financial shocks on innovation in the slow recovery after the Great Recession.

5 Related Literature

This paper contributes to three main strands of literature, each corresponding to a different key idea. First, firms are a productive resource, and the distribution of firms is important for understanding aggregate fluctuations. Second, innovation propagates shocks beyond standard business cycle frequencies, and can help explain slow recoveries. Third, countercyclical shocks to uncertainty contribute to aggregate fluctuations in output and employment. Following is a brief discussion of this paper’s contribution to each respective literature.

A large literature, following [Hopenhayn \(1992\)](#), postulates a technological role for firms. Firms enter, produce, invest and eventually exit, according to economic principles. The distribution of firms is then an aggregate state of the economy. Aggregate shocks that affect firm investment decisions lead to changes in the distribution of firms, providing a channel for transmission of aggregate shocks. Motivated by the experience of the Great Recession and the availability of new datasets, economists have recently adopted this framework to study aggregate fluctuations. [Clementi, Khan, Palazzo, and Thomas \(2015\)](#) and [Clementi and Palazzo \(2016\)](#) study propagation of aggregate TFP shocks in a model that captures size dependence in firm dynamics, but have little persistence of cohort characteristics.

The closest papers to this work are [Moreira \(2015\)](#) and [Sedláček and Sterk \(2016\)](#). [Moreira \(2015\)](#) studies the effects of the initial aggregate conditions on cohorts of young firms using establishment-level data on all US firms. She finds that establishments entering when aggregate output is high are on average larger, even several years after entry. In her model firm-level output has a persistent effect on firm-level demand. Firms that enter when aggregate demand is high produce more and then face persistently higher demand as they grow old. [Sedláček and Sterk \(2016\)](#) also document persistent cohort effects. They use the same public-use data as this paper, and show that much of the variation in cohort employment is determined at entry and is captured in the mean employment size of entering firms. They propose a model where the ex-ante type composition of entering firms responds to aggregate conditions. Exogenous demand shocks alter the efficiency of investment in “customer capital” and change the relative incentive to create different types of firms. When the demand shock is high, firms with greater growth potential enter and generate persistent differences across cohorts.

There are three main differences between this paper and [Moreira \(2015\)](#) and [Sedláček and Sterk \(2016\)](#). First, this paper documents differences in the cross-sectional firm size distribution across cohorts, while the other two focus on the mean size of firms only. The distribution of firm employment size in this paper provides both discipline to the quantitative implementation of the model and untargeted moments to evaluate the cyclical predictions of the model. [Sedláček and Sterk \(2016\)](#) do use the size distribution of old firms to discipline their type-specific model parameters, but they do not match the full firm size by age distribution nor evaluate their model’s cyclical predictions with firm size distributions.

Second, the proposed mechanism in this paper is different. [Moreira \(2015\)](#) proposes a model

in which the composition of entrants has a minor role in generating persistent differences across cohorts. Instead, a direct effect of initial output on future idiosyncratic demand generates persistence within firms, and leads to persistent differences across cohorts. [Sedláček and Sterk \(2016\)](#) propose a mechanism where the composition of entrants is essential for persistence, based on ex-ante heterogeneity in firm growth potential. But the difference in the composition of cohorts in their paper is driven by demand shocks that alter the technology of firm growth. In this paper, the composition of firm cohorts varies due to countercyclical risk premia and heterogeneous exposure to risk. Innovative firms enter when the risky growth options they own are more valuable.

Third, this paper has different observable implications on the co-movement of firm entry, output, hours and stock returns. In [Moreira \(2015\)](#), output by incumbents has a muted response to the aggregate demand shocks that move entry, since firms want to maintain their level of idiosyncratic demand. Also, in her model hours and discounts are constant. In [Sedláček and Sterk \(2016\)](#), incumbent firms are less affected by the demand shocks that drive entry because they are closer to their optimal scale. Hours, output and discounts move due to other shocks that are uncorrelated with demand shocks. Risk premia do not play any part in their mechanism. This paper proposes a theory with testable implications where the same correlated shock drives the co-movement of firm entry, output, hours and stock returns, as well as the cross-sectional distribution of firm size by age.

Another strand of literature studies propagation of shocks beyond business cycle frequency through innovation, based on the framework of [Comin and Gertler \(2006\)](#). [Comin, Gertler, Ngo, and Santacreu \(2016\)](#), [Bianchi, Kung, and Morales \(2017\)](#) and [Anzoategui, Comin, Gertler, and Martinez \(2016\)](#) propose extensions to [Comin and Gertler \(2006\)](#) that explain slow recoveries through product innovation. Their mechanisms are based on the proposition that product variety, the outcome of product innovation, is both a productive factor and an input in more product innovation. When the economy is hit with shocks that reduce the incentives to innovate, product variety falls. It is then harder to create new products, slowing down recoveries.

This paper shares a similar channel for slow recovery: firms own organization capital that is both a productive factor and an input into within-firm innovation. The key distinction is in the type of shocks that affect innovation and in the treatment of risk.²¹ [Bianchi, Kung, and Morales \(2017\)](#) explore shocks to the technology of investment as the main drivers of cyclical innovation. [Comin, Gertler, Ngo, and Santacreu \(2016\)](#) and [Anzoategui, Comin, Gertler, and Martinez \(2016\)](#) explore preference shocks to holdings of different types of assets, that change the risk-free rate. In all three papers, the model equations are log-linearized, essentially treating innovation as a riskless activity. In contrast, in this paper firms are treated as risky

²¹ Another technical difference is that in my model innovation in new firms is invariant to the existing stock of organization capital, making the aggregate stock of organization capital stationary. Other papers feature externalities that induce endogenous growth, as in [Romer \(1990\)](#).

assets. Financial uncertainty shocks change the incentives to innovate, while the innovation technology and the risk-free rate do not change. This paper also has implications for observed firm dynamics.

Other papers study the idea that shocks to uncertainty contribute to aggregate fluctuations in investment, output and employment. Bloom (2009), Gilchrist, Sim, and Zakrajšek (2014), Arellano, Bai, and Kehoe (2016) and Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2016) document cyclical variations in idiosyncratic productivity risk, and use them to explain aggregate fluctuations. Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez, and Uribe (2011), Leduc and Liu (2016), Basu and Bundick (2017) study shocks to the volatility of aggregate states. Ilut and Schneider (2014) and Bianchi, Ilut, and Schneider (2017) study shocks to Knightian uncertainty in models where households are ambiguity averse.

This paper is different from that literature in two main ways. First, I model uncertainty shocks as direct shocks to state prices, with a time invariant physical probabilities. This approach has been successfully applied in the finance literature to capture cyclical variation in bond and stock prices (see for example Ang and Piazzesi (2003) for bonds, and Ang and Bekaert (2006) for stocks). While I do not explicitly model the investors that are generating these state prices, the type of variation that I capture—including high and countercyclical risk premia, stable and low risk free rate and return predictability—have long been the objective of the consumption based asset pricing literature (see for example Campbell and Cochrane (1999), Bansal and Yaron (2004) and Gabaix (2012)). Thus, any model of investor behavior consistent with these facts should generate similar variations in state prices.

Second, in the surveyed literature, the main impact of uncertainty shocks is to delay investment in physical capital (and hiring if labor markets have frictions). When uncertainty returns to normal the economy quickly recovers. In contrast, in my model shocks to uncertainty affect entry of firms and innovation within existing firms. As discussed above, this contributes to propagation of aggregate shocks beyond standard business cycle frequencies, and can help account for slow recoveries.

Recent papers have questioned whether conventional models can match the term structure of equity risk premia. These models generate countercyclical risk premia with an upward sloping equity yield, while the empirical evidence using returns on dividend strips suggests a flat or downward sloping equity risk premia (see van Binsbergen and Koijen (2017)). Indeed, claims to firm profit strips in my model do earn an expected return that is increasing in maturity. However, the empirical evidence is on dividend strips of levered equity, and not on profits strips. I show that a simple dividend payout rule that keeps leverage stationary at the firm level, along the lines suggested by Belo, Collin-Dufresne, and Goldstein (2015), delivers a downward sloping term-structure for levered equity. This is because when firms hold a large quantity of debt when the economy is hit by a bad shock they deleverage, and thus drastically reduce their dividend payments in the short-term. This makes short-term dividends more risky than long term dividends. Interestingly, the slope of the term structure becomes steeper in

high leverage states, such as in recessions, consistent with the recent findings of [Bansal, Miller, and Yaron \(2017\)](#).

6 Conclusions

This paper studies how financial uncertainty shocks that change risk premia contribute to fluctuation in the real economy. It is motivated by the evidence that risk premia, or the effective discount rate on risky assets, are high when economic conditions are bad and low when economic conditions are good. Innovation is a real economic activity that creates risky assets, and therefore is sensitive to fluctuations in risk premia. Innovation also propagates and amplifies aggregate shocks because it simultaneously increases the productive capacity and decreases the future cost of innovation.

The paper proposes a model that combines two main features—heterogeneity in the expected duration of firms and time-varying market price of risk—to capture the response of innovation to countercyclical risk premia. Some firms in the model have a longer expected duration, and so endogenously choose to innovate more. When risk premia are low, the value of these firms is high relative to the value of other firms, and more of them are created. If financial risk premia remain low, these firms grow quickly and generate a boom. When risk premia are high, fewer innovative firms are created and the business sector shrinks. In the quantitative implementation of the model I find that this mechanism amplifies the volatility of output and hours by 60 percent compared to a constant market price of risk.

The model provides a new narrative for the differences between the outcomes of the 2001 recession and the 2007-2009 Great Recession. In the mid to late 1990s risk premia were low and many innovative firms entered and grew quickly, creating an economic boom. When financial conditions deteriorated in 2000, innovation slowed down. But, according to the model, there were already many large innovative firms active that offset the decline in productivity and kept output and employment relatively high throughout the short 2001 recession. Also according to the model, innovative firms made up a smaller fraction of the existing US business sector at the beginning of the Great Recession. As risk premia increased, fewer new firms entered and incumbents innovated less. This led to a persistent decline in output, that a weak recovery of productivity and risk premia after the Great Recession helped propagate.

References

- Ang, A. and G. Bekaert (2006). Stock return predictability: Is it there? *The Review of Financial Studies* 20(3), 651–707.
- Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics* 50(4), 745–787.
- Anzoategui, D., D. Comin, M. Gertler, and J. Martinez (2016). Endogenous technology adoption and r&d as sources of business cycle persistence. Technical report, National Bureau of Economic Research.
- Arellano, C., Y. Bai, and P. J. Kehoe (2016). Financial frictions and fluctuations in volatility. Technical report, National Bureau of Economic Research.
- Atkeson, A. and P. J. Kehoe (2005). Modeling and measuring organization capital. *Journal of Political Economy* 113(5), 1026–1053.
- Bansal, R., S. Miller, and A. Yaron (2017). Is the term structure of equity risk premia upward sloping?
- Bansal, R. and A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance* 59(4), 1481–1509.
- Basu, S. and B. Bundick (2017). Uncertainty shocks in a model of effective demand. *Econometrica* 85(3), 937–958.
- Belo, F., P. Collin-Dufresne, and R. S. Goldstein (2015). Dividend dynamics and the term structure of dividend strips. *The Journal of Finance* 70(3), 1115–1160.
- Berk, J. B., R. C. Green, and V. Naik (1999). Optimal investment, growth options, and security returns. *The Journal of Finance* 54(5), 1553–1607.
- Bianchi, F., C. Ilut, and M. Schneider (2017). Uncertainty shocks, asset supply and pricing over the business cycle.
- Bianchi, F., H. Kung, and G. Morales (2017). Growth, slowdowns, and recoveries. Technical report, National Bureau of Economic Research.
- Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica* 77(3), 623–685.
- Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2016). Really uncertain business cycles.

- Brown, J. R., S. M. Fazzari, and B. C. Petersen (2009). Financing innovation and growth: Cash flow, external equity, and the 1990s r&d boom. *The Journal of Finance* 64(1), 151–185.
- Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy* 107(2), 205–251.
- Clementi, G. L., A. Khan, B. Palazzo, and J. K. Thomas (2015). Entry, exit and the shape of aggregate fluctuations in a general equilibrium model with capital heterogeneity. *Unpublished Working Paper*.
- Clementi, G. L. and B. Palazzo (2016). Entry, exit, firm dynamics, and aggregate fluctuations. *American Economic Journal: Macroeconomics* 8(3), 1–41.
- Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American Economic Review*, 138–160.
- Comin, D., C. M. Gertler, P. Ngo, and A. M. Santacreu (2016). Stock price fluctuations and productivity growth.
- Comin, D. and M. Gertler (2006). Medium-term business cycles. *The American Economic Review* 96(3), 523–551.
- Corrado, C., C. Hulten, and D. Sichel (2009). Intangible capital and us economic growth. *Review of income and wealth* 55(3), 661–685.
- Eisfeldt, A. L. and D. Papanikolaou (2013). Organization capital and the cross-section of expected returns. *The Journal of Finance* 68(4), 1365–1406.
- Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of financial economics* 25(1), 23–49.
- Fernández-Villaverde, J., P. Guerrón-Quintana, J. F. Rubio-Ramírez, and M. Uribe (2011). Risk matters: The real effects of volatility shocks. *The American Economic Review* 101(6), 2530–2561.
- Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly journal of economics* 127(2), 645–700.
- Garcia-Macia, D. (2017). The financing of ideas and the great deviation. Technical report, IMF working paper.
- Garleanu, N., S. Panageas, and J. Yu (2012). Technological growth and asset pricing. *The Journal of Finance* 67(4), 1265–1292.
- Gilchrist, S., J. W. Sim, and E. Zakrajšek (2014). Uncertainty, financial frictions, and investment dynamics. Technical report, National Bureau of Economic Research.

- Gilchrist, S. and E. Zakrajšek (2012). Credit spreads and business cycle fluctuations. *The American Economic Review* 102(4), 1692–1720.
- Gomes, J., L. Kogan, and L. Zhang (2003). Equilibrium cross section of returns. *Journal of Political Economy* 111(4), 693–732.
- Gourio, F. and L. Rudanko (2014). Customer capital. *The Review of Economic Studies* 81(3), 1102–1136.
- Hall, R. E. (2005). Employment fluctuations with equilibrium wage stickiness. *American economic review*, 50–65.
- Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, 1127–1150.
- Hurst, E. and B. W. Pugsley (2011). What do small businesses do? *Brookings Papers on Economic Activity* (2).
- Ilut, C. L. and M. Schneider (2014). Ambiguous business cycles. *The American Economic Review* 104(8), 2368–2399.
- İmrohoroğlu, A. and Ş. Tüzel (2014). Firm-level productivity, risk, and return. *Management Science* 60(8), 2073–2090.
- Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica: Journal of the Econometric Society*, 649–670.
- Kaplan, S. N. and A. Schoar (2005). Private equity performance: Returns, persistence, and capital flows. *The Journal of Finance* 60(4), 1791–1823.
- King, R. G. and S. T. Rebelo (1999). Resuscitating real business cycles. *Handbook of macroeconomics* 1, 927–1007.
- Klette, T. J. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of political economy* 112(5), 986–1018.
- Kogan, L. and D. Papanikolaou (2014). Growth opportunities, technology shocks, and asset prices. *The Journal of Finance* 69(2), 675–718.
- Leduc, S. and Z. Liu (2016). Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics* 82, 20–35.
- Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance* 56(3), 815–849.

- Luttmer, E. G. J. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics* 122(3), 1103–1144.
- Moreira, S. (2015). Firm dynamics, persistent effects of entry conditions, and business cycles.
- Prescott, E. C. and M. Visscher (1980). Organization capital. *Journal of political Economy* 88(3), 446–461.
- Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy* 98(5, Part 2), S71–S102.
- Schoar, A. (2010, February). *The Divide between Subsistence and Transformational Entrepreneurship*, pp. 57–81. University of Chicago Press.
- Sedláček, P. and V. Sterk (2016). The growth potential of startups over the business cycle. *Unpublished*.
- Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American economic review*, 25–49.
- Stock, J. H. and M. W. Watson (1999). Business cycle fluctuations in us macroeconomic time series. *Handbook of macroeconomics* 1, 3–64.
- Swanson, E. T. (2004). Measuring the cyclicity of real wages: How important is the firm’s point of view? *The Review of Economics and Statistics* 86(1), 362–377.
- van Binsbergen, J., M. Brandt, and R. Koijen (2012). On the timing and pricing of dividends. *The American Economic Review* 102(4), 1596–1618.
- van Binsbergen, J., W. Hueskes, R. Koijen, and E. Vrugt (2013). Equity yields. *Journal of Financial Economics* 110(3), 503–519.
- van Binsbergen, J. H. and R. S. Koijen (2017). The term structure of returns: Facts and theory. *Journal of Financial Economics* 124(1), 1–21.
- Villacorta, A. (2017). Business cycles and the balance sheets of the financial and non-financial sectors.
- Zhang, L. (2005). The value premium. *The Journal of Finance* 60(1), 67–103.

A The Value of Dividend Strips

This section solves for the value of dividend strip of maturity n using the prices of options on firm with maturities $j \leq n$. Let V_t be the value of a firm with a profit process Π_t and a debt process B_t . Let $\hat{V}_t = V_t - \Pi_t$ be the after profit value of the firm. Define $M_{t,t+n} = \prod_{j=1}^n M_{t+j}$ as the n periods pricing kernel. The value of a single profit strip $P_{t,n}^\pi$ is then

$$P_{t,n}^\pi = E_t[M_{t,t+n}\Pi_{t+n}].$$

The after profit value of the firm is the sum of all claims to future profits, $\hat{V}_t = \sum_{j=1}^{\infty} P_{t,j}^\pi$. Let $S_{t,n}$ be the value at time t of a claim to the profits of the firm beginning at period $t+n+1$. Then $S_{t,0} = \hat{V}_t$, and

$$S_{t,n} = \sum_{j=n+1}^{\infty} E_t[M_{t,t+j}\Pi_{t+j}] = E_t[M_{t,t+n} \sum_{j=n+1}^{\infty} E_{t+n}[M_{t+n,t+j}\Pi_{t+j}]] = E_t M_{t,t+n} \hat{V}_{t+n},$$

where the first equality applies the law of iterated expectations and replaces the order of integration and summation. This means that $S_{t,n}$ is also the value of a European call option on the after profit value of the firm at time $t+n$ with strike at 0. The profit strip can be written as $P_{t,n}^\pi = S_{t,n-1} - S_{t,n}$.

I specify the following payout policy, based on a leverage target \mathcal{L} . Each period, firms repay a fraction $1 - \tau$ of their outstanding debt B_{t-1} , and issue a constant fraction of their after profits value in risk-free debt, $\mathcal{L}\hat{V}_t$. The law of motion for debt is then,

$$B_t = \mathcal{L}\hat{V}_t + \tau B_{t-1},$$

Since the firm issues risk free obligations, the debt carries the risk-free rate R_f , and the dividend payments D_t follow

$$D_t = B_t - R_f B_{t-1} + \Pi_t = \mathcal{L}\hat{V}_t - (R_f - \tau)B_{t-1} + \Pi_t.$$

Similarly, the dividend payment at time $t+n$ can be written as

$$D_{t+n} = \mathcal{L}\hat{V}_{t+n} - (R_f - \tau)B_{t+n-1} + \Pi_{t+n}.$$

Iterating over B_{t+n-1} , the dividend payment becomes a weighted sum of future values of the firms, profits and previous debt,

$$D_{t+n} = \mathcal{L}\hat{V}_{t+n} - \frac{R_f - \tau}{\tau} \sum_{j=0}^{n-1} \tau^{n-j} \mathcal{L}\hat{V}_{t+j} - (R_f - \tau)\tau^n B_{t-1} + \Pi_{t+n}.$$

Now, I use the fact that $E_t[M_{t,t+n}\hat{V}_{t+j}] = R_f^{j-n} S_{t,j}$ for every $j < n$ and write the value of dividend strip with $n \geq 1$ as,

$$P_{t,n}^d = \mathcal{L}S_{t,n} - \frac{R_f - \tau}{\tau} \sum_{j=0}^{n-1} \tau^{n-j} R_f^{j-n} \mathcal{L}S_{t,j} - (R_f - \tau)\tau^n R_f^{-n} B_{t-1} + S_{t,n-1} - S_{t,n}.$$

Define the weights $q_{j,n}$ as

$$q_{j,n} = \begin{cases} -(R_f - \tau)\tau^{n-j-1}R_f^{j-n}\mathcal{L}, & \text{if } j < n - 1 \\ 1 - (R_f - \tau)R_f^{-1}\mathcal{L}, & \text{if } j = n - 1 \\ \mathcal{L} - 1. & \text{if } j = n \end{cases}$$

Then, the value of a dividend strip can be written as

$$P_{t,d}^d = \sum_{j=0}^n q_{j,n} S_{t,j} - (R_f - \tau)(\tau/R_f)^n B_{t-1}$$