Competition, markups, and predictable returns

Alexandre Corhay * Howard Kung† Lukas Schmid‡§

August 2019

This paper jointly examines the link between competition and expected returns in the time series and in the cross section. To this end, we build a general equilibrium model where markups vary because of firm entry with oligopolistic competition. When concentration is high, markups are more sensitive to entry risk. We find that higher markups are associated with higher expected returns over time and across industries, in line with the data. The model can also quantitatively account for the persistent rise in aggregate risk premia and macroeconomic volatility associated with the secular increase trend industry concentration since the mid 1980s.

Keywords: Production-based asset pricing, imperfect competition, time-varying risk premia, stock return predictability, recursive preferences

*Rotman School of Management, University of Toronto. alexandre.corhay@rotman.utoronto.ca

†London Business School & CEPR. hkung@london.edu

‡Fuqua School of Business, Duke University & CEPR. lukas.schmid@duke.edu

§We thank Ravi Bansal, Diego Comin, Ian Dew-Becker, Jack Favilukis, Francisco Gomes, Ralph Koijen, Kai Li, Lars Lochstoer, Erik Loualiche, Dino Palazzo, Thomas Philippon, Mike Simutin, Amir Yaron, Lu Zhang and seminar participants at Boston College, Collegio Carlo Alberto, Copenhagen Business School, Duke University, Durham University, HKUST, London Business School, Stockholm School of Economics, University of Alberta, University of Geneva, University of British Columbia, UC Berkeley, UCLA, USC, University of Hong Kong, University of Piraeus, Peking University, CKGSB, ASU Sonoran Conference, Econometric Society World Congress, European Finance Association, NBER Capital Markets, NBER Asset Pricing, WFA, Red Rock Finance Conference, FIRN Asset Pricing Research Group Annual Meeting, and Society for Economic Dynamics for helpful comments and discussions.

1 Introduction

The dramatic rise in industry concentration over the last two decades is a ubiquitous phenomenon permeating a broad swath of sectors in the economy. For example, Grullon, Larkin, and Michaely (2018) document that over 75% of industries experienced an increase in concentration since the late 1990s and that the average increase in concentration is 90%.[1] Profit margins and markups also surged during this period, suggesting that larger firms are enjoying greater market power (e.g., Gutiérrez and Philippon (2016)). The increasing consolidation of market power among the largest firms is naturally a source of consternation among antitrust authorities concerned with economic efficiency. Given a complex web of interactions between firm strategy and market power, the economy-wide impact of these industry dynamics is actively debated. Our paper tries to uncover the aggregate consequences of declining competition using a novel array of asset market evidence.

We find that the acceleration in market power coincided with a rising equity premium and increasing macroeconomic uncertainty over the past two decades. Consistent with these recent trends, we also find a strong positive relation between market power and expected returns over time and across firms in the post-war sample. In the time series, measures of market power, such as markups and net business formation, are significant predictors of future excess market returns and consumption volatility. In the cross section, industries with higher average markups earn higher expected returns. Our empirical findings suggest that the recent observed aggregate trends are a pronounced manifestation of a systematic link between market power and risk premia spanning the past seven decades.

To explain these aforementioned relations between competition and risk, we build a dynamic general equilibrium model of heterogeneous industries, where market power varies over time and across industries due to the interplay of firm entry and oligopolistic competition (e.g., Jaimovich and Floetotto (2008)). Firms behaving as oligopolists are cognizant of the effect of their pricing strategy on industry demand. The extent to which firm decisions impact the industry price level depends on the mass of competitors. When there is less competition, firms enjoying greater market power charge higher markups, but can be aggressively undercut by new entrants, exposing them more to entry risk. As entry depends on aggregate and industry business conditions, a positive link between markups and risk premia materializes over time and across firms, consistent with our empirical evidence.

[1]To measure concentration, Grullon, Larkin, and Michaely (2018) use the Herfindahl-Hirschman index (HHI).

The model also incorporates endogenous long-run risks and recursive preferences to quantitatively fit asset market data jointly with industry and macroeconomic dynamics. Long-run risks in consumption and dividend growth are generated by the accumulation of intangible capital through spillover effects as in Kung (2015) and Kung and Schmid (2015). We assume a representative household with Epstein and Zin (1989) utility and a preference for an early resolution of uncertainty. In equilibrium, this preference specification implies a large and positive price of risk for long-run consumption growth uncertainty that is reflected in a sizable equity premium. The model is formally evaluated by estimating key parameters with asset pricing and macroeconomic data using a simulated method of moments (SMM) approach.

We use the estimated model to quantitatively examine how a steady erosion of industry competition – like the pervasive trend witnessed since the 1980s – impacts macroeconomic uncertainty and risk premia. Industry competitiveness depends on the net entry of firms, where entry requires that firms incur a fixed startup cost. A free entry condition equates the startup costs of a marginal entrant to the expected present value of profits. In the model counterfactual, we capture the decline in competition by calibrating a persistent increase in entry costs, disciplined by matching the rise in average markups over the past two decades. The increasing entry costs can be interpreted as a reduced-form way of capturing rising barriers to entry (e.g., weakening antitrust policy or increasing technological barriers). Given the positive equilibrium relation between market power and macroeconomic uncertainty, the calibrated decline in competition generates a quantitatively significant increase in consumption volatility and the equity premium, in accordance with the observed trends.

From an asset pricing perspective, our general equilibrium model provides a microfoundation for the long-run consumption and volatility risks specified in Bansal and Yaron (2004). Our benchmark model only incorporates stationary and homoskedastic shocks. As such, the rich aggregate growth and volatility dynamics are an equilibrium outcome of the model. Long-run growth risks arise from the accumulation of intangible capital in the presence of spillover effects from innovation decisions by firms. With recursive utility, persistent variation in aggregate growth leads to a sizable unconditional equity premium. Volatility risks originate from firm entry in the presence of oligopolistic competition. Persistent changes in the volatility of aggregate growth produces a predictable conditional equity premium.

The endogenous growth and volatility mechanisms outlined above are strongly supported in the data. The endogenous growth margin links expected growth rates to aggregate innovation. Kung (2015) and Kung and Schmid (2015) verify that measures of innovation are significant predictors of future consumption growth rates, especially at long horizons, in accordance with the model. The extensive margin links predictable variation in both macroeconomic uncertainty and risk premia to competition. We find that measures of competition, such as markups, forecast consumption volatility and excess stock returns, with a degree of predictability that aligns with the model. The time-series and cross-sectional linkages in competition, risk, and expected returns in our model hinge on the nonlinear equilibrium relation between markups and the mass of competitors. We uncover a negative and convex relation between markups and net business formation with reduced-form estimates that are consistent with the relation implied by the estimated model.

Our paper closely relates to equilibrium models examining the interaction between product market competition and risk premia. Loualiche (2016) studies the asset pricing implications of a general equilibrium model with endogenous entry and heterogeneous industries. Barrot, Loualiche, and Sauvagnat (2018) extend this framework to an open economy setting with trade. Both papers find that industries more exposed to entry risk, either domestic or international, earn higher expected returns. Bustamante and Donangelo (2017) find a positive relation between markups and expected returns in an industry equilibrium model where firms compete strategically. Bena, Garlappi, and Grüning (2015) show how heterogeneous innovation by incumbents and entrants generates time-varying risks in a growth model with entry. We complement these papers, but differ along the following dimensions. First, we show how the entry margin with oligopolistic competition provides a risk propagation mechanism that can *jointly* explain time-series and cross-sectional return predictability depending on markups. Second, we show that our risk propagation mechanism can quantitatively explain how the acceleration in industry concentration observed over the past few decades is mirrored by a persistent increase in both macroeconomic uncertainty and the equity premium.

The endogenous growth margin builds on the framework from Kung (2015) and Kung and Schmid (2015). These two papers illustrate how the presence of spillovers from intangible capital provide a long-run growth propagation mechanism that generates equilibrium long-run risks. Our paper differs from this strand of literature by illustrating how the presence of oligopolistic competition provides an equilibrium explanation for excess stock return predictability. More broadly, our paper is related to the literature examining the asset pricing implications of technological innovation, such as Garleanu, Panageas, and Yu (2012), Kogan, Papanikolaou, and Stoffman (2013), Garleanu, Panageas, Papanikolaou, and Yu (2016), and Lin, Palazzo, and Yang (2017), which abstract from the dynamics of imperfect competition.

We relate to papers examining endogenous mechanisms for stock return predictability in production-based asset pricing models. Gourio (2007), Belo, Lin, and Bazdresch (2014), Favilukis and Lin (2015), and Favilukis and Lin (2016) emphasize frictions in labor markets. Rigidities in adjusting wages generate an operating leverage effect that produces countercyclical risk premia, forecastable by variables related to labor market conditions. Gomes and Schmid (2016) and Favilukis, Lin, and Zhao (2017) explicitly model financial leverage in general equilibrium and find that credit spreads forecast stock returns through countercyclical leverage. Dew-Becker (2014) and Kung (2015) generate excess return predictability through exogenous time-varying risk aversion and stochastic volatility in productivity, respectively. In our paper, time-varying risk premia arise through endogenous firm entry in the presence of oligopolistic competition. Our novel theoretical channel allows us to empirically identify and test a new set of predictive variables related to time-varying measures of market power.

Our mechanism based on the competitive interactions among firms connects our paper to the growing literature that examines the asset pricing implications of firm linkages within and across sectors and networks, such as Ai, Croce, and Li (2013), Clementi and Palazzo (2016), Lyandres and Palazzo (2016), Herskovic (2015), Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016), Loualiche (2016), or Opp, Parlour, and Walden (2014). In contrast to these papers, we focus on the time-series and cross-sectional predictability of stock returns.

More broadly, our paper relates to the literature on production-based asset pricing in general equilibrium. These include production-based models with habits (e.g., Jermann (1998)), long-run risks (e.g., Ai (2010), Kaltenbrunner and Lochstoer (2010), Kuehn (2007a), Kuehn (2007b), Croce (2014), Ai, Croce, Diercks, and Li (2015)), and disasters (e.g., Gourio (2012) and Kuehn, Petrosky-Nadeau, and Zhang (2014)). Our paper extends this literature by incorporating oligopolistic competition into this class of models, and illustrating a novel equilibrium relation between the dynamics of expected returns and competition.

2 Model

We build a dynamic stochastic general equilibrium (DSGE) model that links competition to risk premia over time and across firms. Households are characterized by a representative agent and markets are complete. We assume that the household sector accumulates the physical and intangible capital stocks in the economy. Production consists of two tiers. The top tier contains a continuum of heterogeneous industries of unit measure producing differentiated *industry goods* that are packaged together into *final goods* for consumption. In the bottom tier, there is an endogenous mass of intermediate firms within each industry that produce differentiated *products* that are aggregated into industry goods. Each industry is characterized by an oligopolistic structure with free entry (e.g., Jaimovich and Floetotto (2008) and Eckel and Neary (2010)). Entry into an industry entails a fixed startup cost. We introduce heterogeneity across industries through differences in operating costs.

2.1 Households and preferences

We assume a representative household with Epstein-Zin preferences defined over aggregate consumption, $\overline{\mathcal{C}}_t$, and labor, $\overline{\mathcal{L}}_t$:

$$\overline{\mathcal{U}}_t = u\left(\overline{\mathcal{C}}_t, \overline{\mathcal{L}}_t\right) + \beta\left(E_t\left[\overline{\mathcal{U}}_{t+1}^{1-\theta}\right]\right)^{\frac{1}{1-\theta}}, \tag{1}$$

where $\theta \equiv 1 - \frac{1-\gamma}{1-1/\psi}$, γ captures the degree of relative risk aversion, ψ is the elasticity of intertemporal substitution, and β is the subjective discount rate.[2] We assume that the utility kernel is additively separable in consumption and leisure:

$$u\left(\overline{\mathcal{C}}_t, \overline{\mathcal{L}}_t\right) = \frac{\overline{\mathcal{C}}_t^{1-1/\psi}}{1-1/\psi} + \overline{\mathcal{Z}}_t^{1-1/\psi}\chi_0\frac{(1-\overline{\mathcal{L}}_t)^{1-\chi}}{1-\chi}, \tag{2}$$

where χ captures the Frisch elasticity of labor and χ_0 is a scaling parameter.[3] We assume that $\psi > \frac{1}{\gamma}$, so that the agent has a preference for early resolution of uncertainty following the long-run risks literature (e.g., Bansal and Yaron (2004)).

[2] We use calligraphic font with an overline for aggregate-level variables, calligraphic font without an overline for industry-level variables, and regular font for firm-level variables.

[3] We scale the second term by the trend component of aggregate productivity, $\overline{\mathcal{Z}}_t^{1-1/\psi}$, to ensure that utility for leisure does not become trivially small along the balanced growth path.

The household accumulates the stock of physical capital for each industry, $\mathcal{K}_t$, by making investment $\mathcal{I}_t$, according to the following law of motion:

$$\mathcal{K}_{t+1} \quad = \quad (1-\delta_k)\mathcal{K}_t + \Phi_t^k\left(\frac{\mathcal{I}_t}{\mathcal{K}_t}\right)\mathcal{K}_t, \tag{3}$$

where δ_k is the depreciation rate and $\Phi_t^k(\cdot)$ captures convex adjustment costs.

The industry stock of intangible capital is interpreted as the knowledge capital specific to a particular industry. Increasing the stock of intangible capital makes production more efficient. The household also accumulates the stock of intangible capital, $\mathcal{Z}_t$, for each industry by making research and development (R&D) investments, $\mathcal{S}_t$, through the following law of motion:

$$\mathcal{Z}_{t+1} = (1-\delta_z)\mathcal{Z}_t + \Phi_t^z\left(\frac{\mathcal{S}_t}{\mathcal{Z}_t}\right)\mathcal{Z}_t, \tag{4}$$

where δ_z is the depreciation rate for intangible capital and $\Phi_t^z(\cdot)$ captures convex adjustment costs for intangible capital. The accumulation of intangible capital is also referred to as process innovation.

We assume the following functional form for capital adjustment costs as in Jermann (1998):

$$\Phi_t^m(x) = \left(\frac{\alpha_{1,m}}{1-\frac{1}{\zeta_m}}(x)^{1-\frac{1}{\zeta_m}} + \alpha_{2,m}\right)e^{\epsilon_{m,t}}, \tag{5}$$

for $m = k, z$, where $\epsilon_{m,t} \sim N(0, \sigma_m)$ represents an i.i.d. shock to the marginal efficiency of investment common across all industries (e.g., Justiniano, Primiceri, and Tambalotti (2010)). The investment shocks are important for explaining the relative volatility of investment to consumption. The parameters, $\alpha_{1,m}$, and $\alpha_{2,m}$, are chosen so that there are no adjustment costs in the steady state.

The household provides capital and labor services to each industry in competitive factor markets, and receives the rental rate $\mathcal{R}_t^m$, for $m = k, z$ for capital services and the wage rate $\overline{\mathcal{W}}_t$ for labor services. The household owns all firms and receives the aggregate payout, $\overline{\Pi}_t$. The optimization problem of the household and the corresponding optimality conditions are contained in Appendix A.1.

2.2 Production technology

Final goods, $\overline{\mathcal{Y}}_t$, are created through a two-tier production structure. First, a continuum of industry goods, $j \in [0,1]$, are bundled together using a constant elasticity of substitution (CES) aggregator:

$$\overline{\mathcal{Y}}_t = \left(\int_0^1 \mathcal{Y}_t(j)^{\frac{\nu_1 - 1}{\nu_1}} \, dj \right)^{\frac{\nu_1}{\nu_1 - 1}}, \tag{6}$$

where $\nu_1 > 0$ is the elasticity of substitution between industry goods.

Within a particular industry, a CES aggregator bundles together a mass, $\mathcal{N}_t$, of differentiated products, X_t, according to:

$$\mathcal{Y}_t = \left(\int_0^{\mathcal{N}_t} X_t(i)^{\frac{\nu_2 - 1}{\nu_2}} \, di \right)^{\frac{\nu_2}{\nu_2 - 1}}, \tag{7}$$

where $\nu_2 > 0$ is the elasticity of substitution between products. We focus on a case where the elasticity of substitution is higher within than across industries (i.e., $\nu_2 > \nu_1$).

Each product is produced by a particular firm in an industry using the following technology:

$$X_t = K_t^{\alpha} \left(\overline{A}_t Z_t^{\eta} \overline{\mathcal{Z}}_t^{1-\eta} L_t \right)^{1-\alpha}, \tag{8}$$

where K_t and Z_t are the firm-specific physical and intangible capital inputs, respectively, L_t is the labor input, and $\overline{\mathcal{Z}}_t$ is the aggregate stock of intangible capital.[4] The parameter $1-\eta$ captures the degree of technological spillovers from intangible capital. The only exogenous forcing process in the benchmark model is the stationary and homoskedastic aggregate productivity shock, $\overline{A}_t$, that affects all intermediate firms across all industries symmetrically, and evolves as an AR(1) process in logs:

$$\overline{a}_t = (1-\rho)a^{\star} + \rho \overline{a}_{t-1} + \sigma \epsilon_t, \tag{9}$$

where $\epsilon_t \sim iid\ N(0,1)$.

[4] The aggregate stock of intangible capital is $\overline{\mathcal{Z}}_t \equiv \int_0^1 \int_0^{\mathcal{N}_t(j)} Z_t(i,j) \, di dj$.

2.3 Industry and product demand

A competitive representative final goods firm produces the final goods used for consumption by first packaging products into industry goods according to Eq. (7), and then combining the industry goods into final goods according to Eq. (6). The profit maximization program for the final goods firm yields a conditional demand schedule for industry goods:

$$\mathcal{Y}_t = \overline{\mathcal{Y}}_t \left(\frac{\mathcal{P}_t}{\overline{\mathcal{P}}_t} \right)^{-\nu_1}, \tag{10}$$

and for products within an industry:

$$X_t = \mathcal{Y}_t \left(\frac{P_t}{\mathcal{P}_t} \right)^{-\nu_2}, \tag{11}$$

where P_t is the price of a product, $\mathcal{P}_t$ is an industry price index, and $\overline{\mathcal{P}}_t$ is the final goods price index.[5] Details of the final goods firm's problem are contained in Appendix A.2.

Choosing $\overline{\mathcal{P}}_t = 1$ as our numeraire and combining Eqs. (10) and (11) yields the conditional demand for a product, in terms of the final goods:

$$\Theta_t(P_t) \equiv \overline{\mathcal{Y}}_t \left(P_t\right)^{-\nu_2} \left(\mathcal{P}_t\right)^{\nu_2-\nu_1}. \tag{12}$$

A tractable way of incorporating strategic interactions among firms in a general equilibrium setting is for firms to behave as if they are "large" in their industry but "small" relative to the overall economy (e.g., Jaimovich and Floetotto (2008) and Eckel and Neary (2010)). In each industry, we assume that firms play a static game of Bertrand price competition with differentiated products, where the pricing strategies are chosen after the entry decision. As the number of competitors is taken as given in each period by active firms, our model only features strategic interactions among active firms within an industry, but not with entrants. The intermediate firms are small relative to aggregate economy due to the continuum of industries of unit mass in the top layer of the final goods production function. Thus, firms account for their pricing decisions on the industry goods price, but their actions do not affect aggregate factor prices, such as the rental

[5] The price index for an industry is given by $\mathcal{P}_t \equiv \left(\int_0^{\mathcal{N}_t} P_t(i)^{1-\nu_2} di\right)^{\frac{1}{1-\nu_2}}$ and the aggregate price index is defined as $\overline{\mathcal{P}}_t \equiv \left(\int_0^1 \mathcal{P}_t(j)^{1-\nu_1}\, dj\right)^{\frac{1}{1-\nu_1}}$.

rates of capital and wages, nor national income.

2.4 Equilibrium markups

Intermediate firms behaving as oligopolists internalize the impact of their price strategies on industry demand. The extent to which firm actions impact demand depends on the intensity of competition. An active firm has a larger impact on demand when market power is concentrated among fewer firms. Therefore, the price elasticity of demand is a function of the mass of competitors, such that greater competition leads to a more elastic demand curve and lower markups. We show that the endogenous relation between markups and competition provides a risk propagation channel that generates a positive relation between market power and risk premia both in the time series and in the cross section.

An intermediate firm maximizes profits, D_t:

$$\max_{P_t,K_t,Z_t,L_t} D_t \equiv P_t X_t - \overline{\mathcal{W}}_t L_t - \mathcal{R}_t^k K_t - \mathcal{R}_t^z Z_t - f_t, \tag{13}$$

subject to the demand constraint $X_t \leqslant \Theta_t(P_t)$, where $f_t \equiv f\ \mathcal{K}_t/\mathcal{N}_t$ is a fixed operating cost of production that is specific to an industry and the key source of heterogeneity across industries. The operating costs are scaled by the average size of a firm to ensure balanced growth. The objective of the firm is to maximize profits subject to the demand constraint, $X_t \leqslant \Theta_t(P_t)$. A full characterization of the optimization program is outlined in Appendix A.3.

The optimality condition with respect to price is given by:

$$X_t = \Lambda_t^d \overline{\mathcal{Y}}_t \left[-\nu_2 P_t^{-\nu_2-1} \mathcal{P}_t^{\nu_2-\nu_1} + (\nu_2-\nu_1) P_t^{-\nu_2} \mathcal{P}_t^{\nu_2-\nu_1-1} \frac{\partial \mathcal{P}_t}{\partial P_t} \right], \tag{14}$$

where Λ_t^d is the Lagrange multiplier on the demand constraint and $\partial \mathcal{P}_t / \partial P_t$ represents the derivative of the industry price index with respect to the product price. The partial derivative term represents how firms account for the impact of their actions on the industry price. This term also captures the externality of firm price actions on industry product demand, which characterizes the nature of strategic interactions in Bertrand competition with differentiated products. The extent of the price externality depends directly on $\mathcal{N}_t$, creating a direct link between the pricing strategy and competition.

The presence of strategic interactions in our model endogenously links market power to the mass of competitors. In a symmetric equilibrium within an industry, but allowing for potential heterogeneity across industries, the price elasticity of demand, $\tau_t \equiv \tau(\mathcal{N}_t)$, is a function of the mass of competitors:

$$\tau(\mathcal{N}_t) = -\nu_2 + (\nu_2 - \nu_1)\mathcal{N}_t^{-1}. \tag{15}$$

The price markup, φ_t, is defined as the ratio of product price to marginal costs. The profit maximization condition equating marginal revenue to marginal cost implies the following inverse relation between the price markup and the price elasticity of demand:

$$\varphi_t = \frac{1}{1+\tau_t^{-1}}. \tag{16}$$

Thus, the price markup, $\varphi_t \equiv \varphi(\mathcal{N}_t)$, is also a function of the mass of competitors:

$$\varphi(\mathcal{N}_t) = \frac{-\nu_2\,\mathcal{N}_t + (\nu_2 - \nu_1)}{-(\nu_2 - 1)\,\mathcal{N}_t + (\nu_2 - \nu_1)}. \tag{17}$$

Assuming that the elasticity of substitution is higher within an industry than across (i.e., $\nu_2 > \nu_1$), a larger mass of competitors implies a more elastic demand curve for products. The reduced market power leads firms to lower markups.

When $\mathcal{N} \to \infty$, so that there are no strategic interactions (i.e., $\partial\mathcal{P}_t/\partial P_t \to 0$), the constant price elasticity of demand and markup policy obtains, $\lim_{\mathcal{N}_t\to\infty} \xi_t = -\nu_2$ and $\lim_{\mathcal{N}_t\to\infty} \varphi_t = \nu_2/(\nu_2 - 1)$. Therefore, the presence of strategic interactions, arising from our assumption of oligopolistic competition within industries, provides a key link between markups and competition in our model.

The remaining first-order conditions yield conditional factor demands for physical capital, intangible capital, and labor:

$$\mathcal{R}_t^k = \frac{\alpha}{\varphi_t}\frac{P_t X_t}{K_t} \tag{18}$$

$$\mathcal{R}_t^z = \frac{\eta(1-\alpha)}{\varphi_t}\frac{P_t X_t}{Z_t} \tag{19}$$

$$\overline{\mathcal{W}}_t = \frac{(1-\alpha)}{\varphi_t}\frac{P_t X_t}{L_t} \tag{20}$$

Therefore, greater competition depresses markups and increases demand for factor inputs. As described in the section below, the degree of competition in an industry is determined endogenously through the entry and exit of firms.

2.5 Entry and exit

Setting up a new firm in an industry entails the fixed cost $\mathcal{X}_t \equiv \kappa \cdot \mathcal{Z}_t$, where $\kappa > 0$. The cost is scaled by the aggregate trend in technology to ensure that the entry costs do not become trivially small along the balanced growth path. These costs are funded by the households each period, and in return, the households are entitled to the future cash flows. We assume that a newly created firm today at time t will start producing its new product in the following period, at time $t+1$. The evolution of firms in an industry is given by:

$$\mathcal{N}_{t+1} = (1-\delta_n)(\mathcal{N}_t + \mathcal{E}_t),$$

where $\mathcal{E}_t$ is the number of new entrants and δ_n is the constant fraction of products, randomly chosen each period, that become obsolete. The creation of new firms is also referred to as product innovation.

A free entry condition endogenously determines the number of intermediate firms in a particular industry:

$$(1-\delta_n)E_t[\overline{\mathcal{M}}_{t,t+1}V_{t+1}] = \mathcal{X}_t, \tag{21}$$

where $\overline{\mathcal{M}}_{t,t+1}$ is the stochastic discount factor (defined in Appendix A.1) and $V_t = D_t + (1-\delta_n)E_t[\overline{\mathcal{M}}_{t,t+1}V_{t+1}]$ is the market value of an intermediate firm. Therefore, changes in expected profit opportunities and discount rates lead to fluctuations in the number of entering firms. The mass of firms, $\mathcal{N}_t$, is approximated as a continuous variable, defined on the positive real number line, to ensure that the zero profit condition binds in equilibrium.

2.6 Aggregation and equilibrium

Given that firms face an identical environment within each industry, we focus on a symmetric Nash equilibrium within each industry, where firms make identical decisions. We model heterogeneity across industries by assuming that there are two industry types, sector L and sector H, equally distributed across the unit interval, that are distinguished by the level of fixed operating costs, with $f_H > f_L$. Without loss of generality, we assume that the low cost sector are the industries located on the interval, $[0, 1/2]$, and the high cost sector are the ones located on the interval, $(1/2, 1]$. The differences in fixed operating costs generate heterogeneity in market power across the L and H sectors.

Given the symmetric equilibrium within industries, aggregate output is a weighted average of the output from the two industy sectors: $\overline{\mathcal{Y}}_t \equiv \left(\frac{1}{2}\mathcal{Y}_{L,t}^{\frac{\nu_1-1}{\nu_1}} + \frac{1}{2}\mathcal{Y}_{H,t}^{\frac{\nu_1-1}{\nu_1}}\right)^{\frac{\nu_1}{\nu_1-1}}$. The aggregate resource constraint is given by:

$$\overline{\mathcal{Y}}_t = \overline{\mathcal{C}}_t + \overline{\mathcal{I}}_t + \overline{\mathcal{J}}_t + \overline{\mathcal{S}}_t + \overline{\mathcal{F}}_t, \tag{22}$$

where $\overline{\mathcal{I}}_t \equiv \frac{1}{2}\mathcal{I}_{L,t} + \frac{1}{2}\mathcal{I}_{H,t}$ is aggregate investment, $\overline{\mathcal{J}}_t \equiv \mathcal{X}_t(\mathcal{N}^E_{L,t} + \mathcal{N}^E_{H,t})$ captures aggregate startup costs for the creation of new firms, $\overline{\mathcal{S}}_t \equiv \frac{1}{2}\mathcal{S}_{L,t} + \frac{1}{2}\mathcal{S}_{H,t}$ represents aggregate expenditures in R&D, and $\overline{\mathcal{F}}_t \equiv \frac{1}{2}f_{L,t} + \frac{1}{2}f_{H,t}$ is the aggregate fixed operating costs.

3 Economic mechanisms

Our model provides a microfoundation for persistent volatility and growth risks that endogenously relate to industry competition. Volatility risks arise from the entry and exit of firms in the presence of oligopolistic competition. Long-run risks emerge from the accumulation of intangible capital in the presence of aggregate knowledge spillovers. These economic mechanisms generate strong linkages between market power and risk premia that are evaluated quantitatively in Sections 4 and 5.

3.1 Endogenous volatility risks

The assumption of oligopolistic competition within industries gives rise to strategic interactions among firms. As described in Section 2.4, the presence of strategic interactions links industry markups, φ, to the mass of firms, $\mathcal{N}$, through the following nonlinear equilibrium relation:

$$\varphi(\mathcal{N}_t) = \frac{-\nu_2\,\mathcal{N}_t + (\nu_2 - \nu_1)}{-(\nu_2 - 1)\,\mathcal{N}_t + (\nu_2 - \nu_1)}. \tag{23}$$

When the substitutability of products is higher within than across industries ($\nu_2 > \nu_1$), as our benchmark parametrization entertains, the markup-competition relation is negative and convex

($\varphi'(\mathcal{N}_t) < 0$, $\varphi''(\mathcal{N}) > 0$ for $\mathcal{N} > 1$). When the substitutability of goods is higher across than within industries, the markup-competition relation is positive and concave ($\varphi'(\mathcal{N}) > 0$, $\varphi''(\mathcal{N}) < 0$ for $\mathcal{N} > 1$). Both cases imply that the sensitivity of markups (i.e., curvature of $\varphi(\cdot)$) to entry is larger when there is less competition. Note that when $\nu_1 = \nu_2$, a constant markup policy obtains ($\varphi_t = \nu_1/(\nu_1 - 1) = \nu_2/(\nu_2 - 1)$).

The left panel of Figure 1 plots markups as function of $\mathcal{N}$ for the benchmark parametrization, $\nu_2 > \nu_1$. When $\mathcal{N}$ is small, the curvature of the markup relation is highest. With less competition, active firms have a stronger impact on industry demand through the strategic effect, $\partial\mathcal{P}/\partial P$, from Eq. (14), but are susceptible to a greater erosion of market power and industry price impact by a marginal entrant. Conversely, the markup relation is flatter when $\mathcal{N}$ is large as the strategic effects diminish with more competition. This nonlinear markup relation generates a negative link between markup volatility and $\mathcal{N}$, both over time and across industries, that transmits to real activity. The factor demands for physical capital, intangible capital, and labor inputs, are directly affected by markups, as characterized in Eqs. (18)–(20), respectively. When $\mathcal{N}$ is smaller, firms are more exposed to entry risk, facing increasing uncertainty about factor demands thereby amplifying real activity.

The mass of competitors, $\mathcal{N}$, depends on both aggregate and industry business conditions through the free entry condition:

$$(1 - \delta_n)E_t[\overline{\mathcal{M}}_{t,t+1}V_{t+1}] = \mathcal{X}_t, \tag{24}$$

where $V_t = D_t + (1 - \delta_n)E_t[\overline{\mathcal{M}}_{t+1}V_{t+1}]$ is the present value of current and future expected profits. Therefore, competition endogenously varies over time and across industries, generating predictability in risk premia over time and across firms.

3.1.1 Time-series implications

In economic expansions (high productivity periods), profits are higher and the marginal utility of households is lower, both of which contribute to higher valuations across all industries. Higher valuations attract more firms, increasing $\mathcal{N}$ up until the free entry condition is satisfied. A larger mass of firms reduces exposure to entry risk, lowering macroeconomic uncertainty (e.g., conditional volatility of aggregate consumption, output, dividends, investment, and labor hours, etc.). Lower aggregate risk reduces the equity premium. Thus, the model generates predictable variation in macroeconomic uncertainty and risk premia *over time* that are both negatively related to the degree of competition. In our benchmark parametrization ($\nu_2 > \nu_1$), markups are negatively related to $\mathcal{N}$, which implies a positive relation between average industry markups and the conditional equity premium.

3.1.2 Cross-sectional implications

Competition varies across industry sectors, L and H, because of heterogeneity in operating costs, f, which act as a barrier to entry. Higher operating costs lower profits and valuations, which dampens firm entry. In particular, the higher operating costs in sector H attract fewer new firms than in sector L (i.e., $\mathcal{N}_H < \mathcal{N}_L$). When the substitutability of products is higher within than across industries ($\nu_2 > \nu_1$), fewer competitors in H implies that active firms face a more inelastic demand curve that induces higher average markups compared to L. Weaker competition also implies that industries in sector H are more exposed to entry risk through markups, producing cash flows that are both more volatile and covary more with aggregate conditions. Consequently, firms in H earn higher expected returns than those in L. Therefore, the model also generates predictable variation in uncertainty and risk premia *across industries* that are both negatively related to the degree of competition.

3.2 Endogenous long-run risks

The accumulation of intangible capital in the presence of positive spillover effects generates equilibrium long-run risks. The technological spillover effects (characterized in Eq. (8)), are specified such that the accumulation of intangible capital exactly offsets the effects of diminishing marginal returns to production at the aggregate level, leading to sustained long-run growth. Long-term growth prospects are determined endogenously by the accumulation of intangible capital. Kung (2015) and Kung and Schmid (2015) show that this innovation margin provides a powerful growth propagation mechanism at low frequencies. The persistence in stationary shocks are transmitted to expected growth rates through the accumulation of intangible capital. A positive technology shock raises the marginal product of intangible capital, increasing demand for innovation. An increase in intangible capital raises trend growth persistently due to the technological spillover effects. Following Bansal and Yaron (2004), we assume that the representative agent has a preference for an early resolution of uncertainty (i.e., $\gamma > 1/\psi$), implying that the agent is averse to uncertainty about long-term growth prospects. In equilibrium, the persistent growth dynamics contribute to a sizable unconditional equity premium.

4 Data and estimation

This section characterizes the data and quantitative evaluation of the model. Two empirical measures of markups are constructed based on equilibrium conditions from the model. Key structural parameters are estimated using a simulated method of moments (SMM) procedure.

4.1 Data

The quarterly data series for per capita aggregate consumption, investment, the physical capital stock, R&D expenditures, and output are obtained from the Bureau of Economic Analysis (BEA). The consumer price index (CPI), labor hours, the labor share, and the stock of intangible capital are from the Bureau of Labor Statistics (BLS). Stock return data are from Center for Research in Security Prices (CRSP). Nominal variables are deflated using the CPI. We use the sample period, 1948:Q1 - 2016:Q4, to compute aggregate asset pricing and macroeconomic statistics. This sample period is chosen given the data availability of our business formation measure described below.

An aggregate measure of the mass of competitors, $\mathcal{N}$, is constructed using the Net Business Formation (NBF) index as in Chatterjee and Cooper (2014). In our model, the aggregate measure of firms captures the common time-varying component of competition shared across all industries. The NBF index is sampled monthly for the period, 1948 - 1995, which is the longest available time series of the extensive margin. We extend the NBF series using the Number of Establishment Births (NEB) in the following steps. First, we convert the monthly NBF series to quarterly by averaging the observations within each quarter to align with the NEB, which is available from 1992:Q1 - 2016:Q4. Second, we project the NEB on the NBF for the period in which both data series are available (the correlation between the two series is 0.64). Third, we build a proxy for the post-1995 period using the fitted values from the projection. Note that the mass of competitors, $\mathcal{N}$, is approximated as a continuous variable in the model, while in the data, the measure is discrete. When comparing the dynamics of the model and data measures of $\mathcal{N}$, we compute log deviations from trend so that both series are in the same units. All macroeconomic variables are detrended using the Christiano and Fitzgerald (2003) bandpass filter.

We define an industry by the four-digit Standard Industrial Classification (SIC) code. Annual industry output and total wage expenses used to compute the industry-level labor shares are obtained from NBER-CES Manufacturing Industry Database, available for the period, 1958 - 2011. For the years after 2011, we forward-fill the missing quarterly observations until the end of 2016 using the last available observation in 2011. The Herfindahl-Hirschman index (HHI) for each industry is computed according to:

$$\mathcal{H} \equiv \sum_{i=1}^{\mathcal{N}} S_i^2,$$

where S_i denotes the market share of sales for firm i in its respective industry. The HHI measures are obtained directly from the US Census of Manufacturers, sampled quinquennially starting in 1992. This measure is computed using the fifty largest firms within each four-digit SIC industry in the manufacturing sector, including both private and public firms.[6] Following prior studies (e.g., Haushalter, Klasa, and Maxwell (2007) and Ali, Klasa, and Yeung (2008)), we assume that the values for the 1992, 1997, 2002, 2007, and 2012 index are valid for the five-year windows centered around those dates. We forward-fill missing quarterly observations between the five-year intervals using the last available observation.

To correct for firm-specific characteristics, we collect accounting data from Compustat starting in 1970.[7] Two years of data are needed to estimate the characteristic-based risk correction. Therefore, for the cross-sectional asset pricing moments, we use the sample period 1972:Q1-2016:Q4 for the industry-level markup measure using the labor share and 1990:Q1 -2016:Q4 for the industry-level markup measure constructed using HHI (both described in the section below).

Appendix B provides additional details for each data series, including the corresponding sources for the data.

[6]Since 1997, the US census provides the HHI series at the six-digit NAICS level. We convert the HHI measure to four-digit SIC levels using the same methodology as Ali, Klasa, and Yeung (2008) and use the concordance tables available on the U.S. Census website to link NAICS to SIC codes.

[7]The variables we use from Compustat are: Total Assets (AT), Income before Extraordinary Items (IB), Total Long-term Debt (DLTT), Cash Holdings (CH), Cost of Goods Sold (COGS), Selling, General and Administrative Expenses (XSGA), Total Sales (SALE), and Book Equity (CEQ).

4.2 Markup measures

The price markup is defined as the ratio between the price of output and marginal cost. Using equilibrium conditions from the model, we construct two measures of markups from supply- and demand-side approaches, both at the aggregate- and industry-level.

4.2.1 Supply-side approach

The supply-side approach (e.g., Hall (1986)) infers markups from cost minimization. If the production function is differentiable and factor prices are taken as given, a firm minimizing costs equates the marginal cost of each production input to the marginal product, including that of the labor margin. The optimality condition with respect to labor from cost minimization is given by:

$$\varphi_t = \frac{P_t}{MC_t} = \frac{P_t MPL_t}{\overline{\mathcal{W}}_t}, \tag{25}$$

where MC_t denotes marginal costs and MPL_t denotes the marginal product of labor. Given the Cobb-Douglas-type production function (Eq. 8), $MPL_t = (1-\alpha)Y_t/L_t$. Therefore, we can express Eq. (25) in terms of the labor share, $S_{L,t} \equiv (\overline{\mathcal{W}}_t L_t)(P_t Y_t)$:

$$\varphi_t = \frac{1-\alpha}{S_{L,t}}, \tag{26}$$

which can be expressed in terms of log-deviations from steady-state (denoted by a tilde):

$$\widetilde{\varphi}_t = -\widetilde{s}_{L,t}. \tag{27}$$

As discussed in Section 2.4, the presence of oligopolistic competition implies that markups vary with business formation according to Eq. (23). Given the equilibrium condition above, the labor share dynamics are therefore tied to the competitive environment.

There is strong evidence refuting the assumption that firms take wages as given (i.e., the marginal wage coincides with the average wage). We can make an adjustment to the measure above by assuming that the total wage bill, $\Theta(L) \equiv \mathcal{W}(L) \cdot L$, is nonlinear in labor hours as in Bils (1987) and Rotemberg and Woodford (1999). In this specification, the markup relates to the marginal wage $\mathcal{W}'(L)$:

$$\varphi_t = \frac{(1-\alpha)}{S_{L,t}\Gamma_t}, \tag{28}$$

where $\Gamma \equiv \frac{\mathcal{W}'(L)L}{\mathcal{W}(L)}$ is the ratio of the marginal wage to the average wage bill. Log-linearizing Eq. (28) about the deterministic steady state yields:

$$\widetilde{\varphi}_t = -\widetilde{s}_{L,t} - \omega_L \widetilde{l}_t, \tag{29}$$

where ω_L is the elasticity of Γ_t with respect to L_t in the steady state, i.e., $\omega_L \equiv \partial \log(\Gamma) / \partial \log(L)$. Given a symmetric equilibrium within each industry (assumed in the benchmark model), we can use Eq. (29) to relate markups to the industry labor share. Bils (1987) estimates the elasticity parameter, ω_L, to be equal to 1.4, implying a convex wage bill with respect to hours. Using this estimate for ω_L and the industry labor shares measured at the four-digit SIC-level, we use Eq. (29) to construct industry markups with the supply-side approach. Similarly, assuming a symmetric equilibrium across industries as an approximation (e.g., Rotemberg and Woodford (1999)), average markups can also be linked to the aggregate labor share according to Eq. (29) from which we obtain our aggregate measure of supply-side markup.

4.2.2 Demand-side approach

The demand-side approach requires an assumption for the demand system and for the competitive environment. In our benchmark model, product demand is given by Eq. (12) and we assume Bertrand competition. Assuming a symmetric equilibrium within each industry allows us to relate markups to business formation according to Eq. (23) that is reproduced below:

$$\varphi(\mathcal{N}_t) = \frac{-\nu_2 \mathcal{N}_t + (\nu_2 - \nu_1)}{-(\nu_2 - 1)\mathcal{N}_t + (\nu_2 - \nu_1)}. \tag{30}$$

The section below describes how the elasticity parameters, ν_1 and ν_2, are estimated. In a symmetric equilibrium within an industry, the HHI is inversely related to industry business formation, $\mathcal{H} =$

$1/\mathcal{N}$.[8] We can rewrite Eq. (30) as a function of industry concentration, $\mathcal{H}$:

$$\varphi(\mathcal{H}) = \frac{-\nu_2\,\mathcal{H}^{-1} + (\nu_2 - \nu_1)}{-(\nu_2 - 1)\,\mathcal{H} + (\nu_2 - \nu_1)}. \tag{31}$$

Since HHI controls for heterogeneity in industry size, we use Eq. (31) rather than Eq. (30) in constructing our demand-based measure at the industry level.

Assuming a symmetric equilibrium across industries as an approximation allows average markups to be related to aggregate business formation according to Eq. (30). Our aggregate markup measure from the demand-side is constructed by using the aggregate measure of business formation.

To distinguish between the two measures in the tables, the supply-based measure is henceforth referenced as φ^s while the demand-based measure is labeled as φ^d.[9]

4.3 Model estimation

We estimate eight structural parameters ($\Theta = [\zeta_k, \zeta_z, \sigma_k, \sigma_z, \sigma, a^\star, \beta, \nu_2]'$) using the simulated methods of moments (SMM) approach. This procedure chooses values for $\hat{\Theta}$ in order to minimize the distance between a vector of identifying moments from the data and the corresponding moments generated from model simulations:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}}\,[\hat{m} - m(\Theta)]'\,W\,[\hat{m} - m(\Theta)]\,, \tag{32}$$

where W is a weighting matrix, $\hat{m}$ is a vector of empirical moments, and $m(\Theta)$ is the vector of model-implied moments obtained by assuming a value of Θ for the structural parameters.[10] Although the estimator, $\hat{\Theta}$, is consistent for any positive-definite matrix, W, we follow the literature and set W equal to the inverse of the covariance matrix of the moments. This specification allocates more weight to moments estimated with greater precision and it has the advantage of minimizing the asymptotic variance of $\hat{\Theta}$. To estimate the covariance matrix of the moments, we use the influence function approach of Erickson and Whited (2000).

The eight structural parameters, Θ, are identified using eight target moments described in

[8] The HHI is defined as $\mathcal{H} \equiv \sum_{i=1}^{\mathcal{N}} S_i^2$, where S_i is defined as the market share. In a symmetric equilibrium within an industry, $S_i \equiv 1/\mathcal{N}$ for all i. Therefore, $\mathcal{H} = \sum_{i=1}^{\mathcal{N}}(1/\mathcal{N})^2 = \mathcal{N}(1/\mathcal{N})^2 = 1/\mathcal{N}$.

[9] The demand- and supply-based measures are strongly positively correlated with the markup measure from De Loecker and Eeckhout (2017) who use a production-based approach (correlations of 0.817 and 0.706, respectively).

[10] Model moments are computed over 100 samples of length similar to the data, with a burning-in period of 100 quarters.

Panel A of Table 1. The parameter estimates with the corresponding standard errors are reported in Panel B. The identification of each parameter in relation to the target moments is described below.

The adjustment cost parameter, ζ_k, determines the elasticity of the investment rate with respect to the shadow value of physical capital. A lower elasticity parameter (ζ_k) makes investment rates smoother and more persistent. A sufficiently low value of ζ_k is required to match the first autocorrelation of the investment rate. The parameter, σ_k, governing the standard deviation of the *i.i.d.* investment-specific shocks are important for explaining the high volatility of investment growth relative to the volatility of consumption growth. Similarly, we identify ζ_z and σ_z with the first autocorrelation of the R&D investment rate and the volatility of R&D investment growth relative to the volatility of consumption growth, respectively. As the investment-specific shocks are i.i.d. and homoskedastic, they do not drive the asset pricing results. The primary role of these shocks is to explain the unconditional investment volatility. Section 5.2 discusses how the main asset pricing results, including return predictability, are robust to the exclusion of the investment-specific shocks.

The volatility parameter for aggregate productivity shock, σ, is important for determining the quantity of aggregate risk. We use the volatility of consumption growth to identify the parameter, σ. Average growth in the economy is endogenous and depends on the deep parameters of the model. The unconditional mean of the stationary aggregate productivity shock, $a^\star$, is a scale parameter that positively affects the marginal product of factor inputs, including that of intangible capital. A higher value of $a^\star$ increases the demand for intangible capital, which increases the average R&D investment rate. Due to spillover effects from the accumulation of intangible capital, higher average R&D investment permanently raises trend growth. Therefore, we identify the parameter, $a^\star$, using average output growth as a target moment. The time preference parameter, β, is important for determining the average level of the riskfree rate. A lower β implies that the agent is more impatient and would like to borrow more, which raises the interest rate in equilibrium. We identify β using the average riskfree rate.

The parameter governing the substitutability across products, ν_2, is identified using the average elasticity of markups with respect to business formation, $\xi_t \equiv \frac{\partial \varphi_t/\varphi_t}{\partial \mathcal{N}_t/\mathcal{N}_t}$. In the homogeneous industry case, an analytical expression can be derived for ξ_t that directly relates to ν_2:

$$\xi_t = -\frac{(\nu_2 - \nu_1)\mathcal{N}_t/\varphi_t}{[-(\nu_2 - 1)\mathcal{N}_t + (\nu_2 - \nu_1)]^2}. \tag{33}$$

Conditional on a value of ν_1, the moment, $E[\xi_t]$, pins down ν_2. An estimate of $E[\xi_t]$ is obtained projecting the log of the supply-based measure of markups on the log of business formation.[11]

We calibrate the parameter determining the substitutability across industry goods, ν_1, as it cannot be independently identified from ν_2. The reason why we do not estimate ν_1 (and calibrate ν_2) is that ν_1 is not well-identified. In the homogeneous industry case, the elasticity, ξ_t, does not depend on ν_1 in the steady-state after substituting out the steady-state mass of firms:

$$\xi = -\frac{\frac{(1-\varphi)/\varphi}{-(\nu_2-1)\varphi+\nu_2}}{\left[-\frac{(\nu_2-1)(1-\varphi)}{-(\nu_2-1)\varphi+\nu_2} + 1\right]^2}$$

Fixing a level of φ, the steady-state elasticity, ξ, only depends on ν_2. Since ξ_t varies around its steady-state value in the ergodic distribution, the average elasticity is not informative for identifying ν_1, but it is, however, informative for ν_2.[12]

Conditional on a value of ν_1, matching the negative estimated value for $E[\xi_t]$ requires that the estimated value for ν_2 exceed the value of ν_1. The parameter relation, $\nu_2 > \nu_1$, is also consistent with estimates from micro-data. Broda and Weinstein (2010) estimate the within- and across-industry elasticities (corresponding to ν_2 and ν_1, respectively) using bar-code-level product data, and find that the elasticities of substitution are unequivocally higher within than across industries over the entire distribution of product share ratios.

A parametrization satisfying $\nu_2 > \nu_1$ implies a negative and convex relation between markups and business formation that we verify directly in the data (Panel D in Table 1). We approximate the markup relation using a second-order polynomial approximation in business formation. The estimated coefficient on the linear term is negative and the coefficient on the squared term is positive (both statistically significant), implying a negative and convex relation, consistent with the model predictions. As described in Section 3.1, this nonlinear markup relation generates

[11] We cannot use our demand-based measure to estimate this elasticity as the construction of the demand-based measure requires values for ν_1 and ν_2.

[12] In our benchmark model featuring heterogeneous industries, there is no analytical expression for ξ_t. However, we can show that the insights from the homogeneous industry case extend to benchmark model numerically. Indeed, $E[\xi_t]$ is not very responsive to changes in ν_1, while it is sensitive to changes in ν_2.

time-series and cross-sectional return predictability.

4.4 Calibrated parameters

The remaining thirteen parameters, $\Gamma \equiv [\nu_1, \psi, \gamma, \chi, \alpha, \eta, \delta_k, \delta_z, \delta_n, \kappa, \rho, f_H, f_L]$, are pinned down by values from the existing literature or set to match steady-state evidence. These values are reported in Panel C of Table 1.

As discussed above, the elasticity of substitution across industry goods, ν_1, is not well-identified through aggregate or industry moments, given that the steady-state of the model is not sensitive to changes in ν_1. We calibrate this parameter to a value similar to that of Jaimovich and Floetotto (2008).

The preference parameters controlling risk tastes, ψ (intertemporal elasticity of substitution) and γ (coefficient of relative risk aversion), are calibrated to 1.75 and 10, respectively, and are in the standard range of values in the long-run risks literature (e.g. Bansal and Yaron (2004)). This preference configuration implies that the representative agent prefers an early resolution of uncertainty (i.e. $\psi > 1/\gamma$), which implies that the price of risk for low-frequency consumption growth uncertainty is positive. The labor elasticity parameter, χ, is set to 3. This value implies a Frisch elasticity of labor supply of 2/3, which is consistent with estimates from the microeconomics literature (e.g. Pistaferri (2003)).

The capital share, α, is set to 0.33 and the quarterly depreciation rate for physical capital, δ_k, is set to 2%. These are standard values in the macroeconomics literature and designed to match steady-state evidence. The quarterly depreciation rate of the R&D capital stock, δ_z, is set to 3.75%, and is the value that the BLS uses in its construction of the R&D capital stock. The degree of technological appropriability, η, is calibrated to 0.1, in order to match the steady-state level of R&D intensity (i.e., the ratio of the R&D flows to the R&D stock), as in Kung (2015). The exogenous firm exit shock, δ_n, is set to 2%, in line with Bilbiie, Ghironi, and Melitz (2012). The entry cost parameter κ is chosen to generate a steady state market-to-book ratio of 1.75, as in the data. The persistence parameter, ρ, corresponding to the aggregate productivity shock is calibrated to an annualized value of 0.95, designed to match the dynamics of R&D intensity, as in Kung (2015).

The industry-specific cost parameters, f_L and f_H, are chosen to replicate two moments: (i) an average aggregate markup of 35% and (ii) an average cross-sectional difference between the high and low markup industries inline with the data, where the targeted difference is obtained by sorting industries into terciles based on their price markup.

4.5 Model fit

The first part of Panel A in Table 2 shows that the model provides a good fit of macroeconomic dynamics, as targeted in the estimation and calibration. The last part of Panel A reports the aggregate asset pricing statistics. The model generates a sizable equity premium and a smooth riskfree rate. The endogenous growth margin generates long-run risks in macroeconomic quantities due to the spillover effects from the accumulation of intangible capital that are characterized in Section 3.2. The calibration of the preference parameters satisfying $\gamma > 1/\psi$, implies a large and positive market price of risk for low-frequency consumption risks in equilibrium, allowing the model to generate a large unconditional equity premium. Assuming $\psi > 1$ helps to explain a smooth riskfree rate that accords with the dynamics observed in the data.

Kung (2015) and Kung and Schmid (2015) provide strong empirical support for the endogenous growth margin. They show that the accumulation of intangible capital is a strong predictor of future consumption growth, particularly at long horizons (e.g., five years). Panel B of Table 2 provides a structural interpretation of the frequency decomposition from Dew-Becker and Giglio (2016) through the lens of our model. We run regressions of the high- and low-frequency component of consumption growth on aggregate measures of business formation and intangible capital and report the corresponding R^2s from these projections in the model and the data. The high- (low-) frequency component is obtained by using a bandpass filter and isolating frequencies between 1.5 and 20 years (greater than 20 years). Consistent with our model predictions, business formation is more strongly related to high-frequency movements while the accumulation of intangible capital is more salient for explaining low-frequency movements. In the context of our model, the presence of positive spillovers from the accumulation of intangible capital provide a growth propagation mechanism that transmits stationary level shocks to expected growth rates.

5 Main quantitative results

This section presents the main quantitative results. Our estimated model can explain the positive relation between market power and risk premia observed both in the time series and in the cross section. In counterfactual simulations, we show how a persistent increase in barriers to entry can account for a rise in industry concentration, macroeconomic volatility, and risk premia witnessed over the past three decades.

5.1 Competition and risk premia

Our model generates an equilibrium link between competition and risk premia due to the combination of (i) endogenous firm entry and (ii) oligpolistic competition. Firms behaving as oligopolists are aware of the effects of their pricing actions on industry demand. The extent to which firm actions affect the price elasticity of demand depends on the intensity of competition, leading to a nonlinear relation between industry markups and the mass of firms according to Eq. (23). The model estimation finds a parametrization that satisfies $\nu_2 > \nu_1$, which implies a negative and convex relation ($\varphi'(\mathcal{N}_t) < 0$, $\varphi''(\mathcal{N}) > 0$). As described in Section 3.1, this markup relation produces a positive link between markups and risk premia. As firm entry varies depending on aggregate and industry conditions, the model produces positive co-movement in markups and risk premia over time and across industries. Absent the entry margin or strategic interactions arising from oligopolistic competition, markups and risk premia would be constant in our framework.

5.1.1 Time-series relations

The intensity of competition depends on aggregate economic conditions through the free entry condition linking the present value of expected profits to the fixed startup costs according to Eq. (24). Figure 2 plots impulse response functions to an aggregate productivity shock. Economic expansions (periods of high productivity) are characterized by higher profits, which attracts new firms, generating procyclical business formation. Given the negative relation between markups and the mass of firms, markups are therefore countercyclical. The second part of Panel A in Table 2 verifies that the cyclical properties of business formation and markups implied by our model are consistent with the data. We construct the aggregate markup using both our demand- and supply-based approaches, which exhibit similar dynamic properties.

The convexity in the markup relation generates conditional heteroskedasticity in aggregate growth rates that depend on the state of industry competition. Economic downturns are associated with lower valuations, triggering a reduction in the mass of firms. When competition is weak, firms charge higher markups, but are more susceptible to having their enhanced market power eroded by a new entrant. The increased exposure to entry risk is reflected by a markup relation exhibiting higher curvature with respect to changes in competition as depicted in Figure 1. The sensitivity of markups transmits directly to real activity through the factor demands for capital and labor inputs. Figure 2, therefore, shows impulse response functions to an aggregate productivity shock conditional on high and low markups. It illustrates how macroeconomic fluctuations are amplified when competition is weak (high φ) due to the higher sensitivity of markups.

Table 3 provides empirical support for the link between competition and macroeconomic risk produced by our model. Panel A shows that first and second moments of conditional consumption volatility generated by the model are consistent with the empirical counterparts. Conditional consumption volatility is computed in two steps. First, realized consumption growth is fitted to an $AR(1)$ process. Second, conditional consumption volatility is obtained by estimating an $EGARCH(1,1)$ on the estimated innovations from the first step. Panels B, C, and D report consumption volatility forecasting regressions at a horizon of four quarters using the two measures of markups and the price-dividend ratio. Consistent with the model predictions, an increase in markups forecasts higher future consumption volatility, while the relation is negative with the price-dividend ratio.

The endogenous time-varying macroeconomic risks translate to predictability in excess stock returns. Weaker competition is associated with higher macroeconomic uncertainty, which magnifies the covariance between the stochastic discount factor and aggregate dividends, implying a larger equity premium. Panels A and B in Table 4 illustrate that our markup measures are strong predictors of future excess stock returns at horizons of one, three, and five years. Higher markups are associated with higher subsequent returns, especially at longer horizons. The model can also replicate the return forecasting evidence with the price-dividend ratio, reported in Panel C. Assuming an intertemporal elasticity of substitution greater than one, a preference for an early resolution of uncertainty, and a sufficiently inelastic supply of capital is important for the endogenous volatility dynamics to obtain a negative slope coefficient in the forecasting regressions with the price-dividend ratio. Table 1 of the Appendix shows that our markup measures remain significant predictors of returns after bias-correcting the coefficients according to Stambaugh (1999).

Table 5 reports results from a horse race of our supply- and demand-based markup measures with two macroeconomic variables that also predict returns, the labor share and the *cay* factor from Lettau and Ludvigson (2001). The reported coefficients and R^2 are from excess return forecasting regressions at a horizon of five years. Both of our markup measures improve forecasting power and the slope coefficients remain statistically significant with the inclusion of the labor share and the *cay* factor. As the labor share is a stronger predictor of returns at higher frequencies, we also run a horse race at the one year horizon with the labor share, reported in Table 2 of the Appendix, showing that our markup measures remain significant predictors. These horse race results suggest that markups contain important information about the conditional equity premium beyond that of standard macroeconomic variables. Our model connects the predictive power of markups to the extensive margin through oligopolistic competition.

5.1.2 Cross-sectional relations

The negative and convex markup relation with respect to business formation also produces a positive link between markups and risk premia in the cross section. Cross-sectional heterogeneity arises in our model due to differences in fixed operating costs, f_H and f_L, that act as barriers to entry corresponding to sectors H and L, respectively. Sector H is therefore characterized by weaker competition and higher average markups due to the negative markup elasticity. The convexity in the markup relation implies greater sensitivity of cash flows to entry risk for firms in sector H. As entry in each industry also depends on aggregate conditions through the free entry condition, firms in sector H have higher betas with respect to macroeconomic risks compared to those in sector L, which contributes to a positive spread between high- and low-markup industries.

Panel A in Table 6 reports the cross-sectional asset pricing moments depending on competition. In the data, for each measure of markups, φ^s and φ^d, we sort firms by markups into terciles and rebalance the portfolios at a quarterly frequency. The upper and lower terciles are defined as the *high* and *low* markup industries, respectively. We form value-weighted industry portfolios to compute the asset pricing statistics. Sorting on both the supply- and demand-side measures deliver a sizable high-minus-low-markup spread of around 3-4%, consistent with the magnitudes from the model. The higher expected return in the high markup industry is also attributed to higher operating leverage. In Table 3 of the Appendix, we show that over 75% of the average spread in returns is attributed to the extensive margin.

In Panel B of Table 6, we correct for firm characteristics in the data known to predict returns in the cross-section by using the methodology from Bessembinder, Cooper, and Zhang (2018) in the following steps. First, for each quarter, a cross-sectional regression of firm returns on lagged firm characteristics is estimated. Second, the estimated regression coefficients (averaged over the preceding four quarters) along with the lagged firm characteristics are used to obtain a measure of expected returns for a firm, $E[R_{it}|I_{t-1}]$. The characteristics-adjusted return is then computed as $R_{it} - E[R_{it}|I_{t-1}]$. We consider two sets of characteristics. Under the column labelled, 'Fama-French,' the adjusted return controls for momentum, size, and book-to-market factors. The next column labelled, 'Characteristics,' corrects for asset growth, profitability, asset utilization financial leverage, and operating leverage, in addition to size, book-to-market, and momentum. Our markup-sorted returns spread is robust to controlling for these characteristics.

5.1.3 Interpreting recent trends

Our model can also be used to assess the macroeconomic consequences of the secular rise in market power since the mid-1980s. To document the recent low-frequency trends, we follow Farhi and Gourio (2018) by splitting the sample starting in 1984 into two sub-periods of equal length. Table 7 documents that the increase in aggregate markups between these two periods is associated with a significant rise in both consumption volatility and the equity premium, consistent with findings in Farhi and Gourio (2018) and Greenwald, Lettau, and Ludvigson (2019). These low-frequency trend relations are also consistent with the cyclical and cross-sectional competition-risk relations documented above.

To generate a secular trend in competition in our model comparable to that observed in recent decades, we modify the benchmark model to include an aggregate shock that affects the startup costs of entry:

$$\widetilde{\chi}_t \equiv \kappa_t \cdot \mathcal{Z}_t, \tag{34}$$

where $\mathcal{Z}_t$ is the aggregate trend in technology and κ_t is the entry cost shock that evolves as:

$$\ln(\kappa_t) = (1 - \rho_\kappa)\ln(\kappa) + \rho_\kappa \ln(\kappa_{t-1}) + \sigma_\kappa \epsilon_{\kappa,t}, \tag{35}$$

where $\epsilon_{\kappa,t} \sim N(0,1)$ is an i.i.d. disturbance. Note that if σ_κ is set to zero, we return to the benchmark model. This autoregressive shock can be interpreted as a reduced-form way of capturing the low-frequency fluctuations in barriers to entry that have contributed to the acceleration in industry concentration (e.g., Gutiérrez and Philippon (2016)).

In the model counterfactual, a positive entry cost shock, κ_t, is calibrated to replicate the 9.5% increase in aggregate markups averaged across simulations of equal duration as the data counterparts. In response to the persistent increase in concentration, the model also generates a significant positive trend in consumption volatility and the equity premium in accordance with the empirical analogues. The positive entry cost shock causes a reduction in the mass of firms, which raises markups due to the negative markup elasticity with respect to business formation. The convexity in the markup relation implies that weaker competition increases the sensitivity of markups to entry risk, amplifying macroeconomic fluctuations. An increase in macroeconomic uncertainty increases covariance risk with marginal utility, raising the equity premium in equilibrium. The endogenous markup channel linking competition and risk can also provide a quantitatively relevant account of the secular trends in market power, macroeconomic uncertainty, and risk premia.

5.2 Additional specifications

Table 8 shows that our main quantitative results are robust to alternative specifications of the benchmark model. The column labelled as 'No IST' turns off the shocks to the marginal efficiency of investment, so that the model only has one shock. Given that the estimated volatility for the R&D investment shock is zero, this specification is effectively only setting the volatility of physical investment-specific shocks to zero ($\sigma_k = 0$). As the investment shocks are assumed to be i.i.d., the unconditional asset pricing moments and return predictability results are not affected in any material way. The absence of dynamics in these shocks also explains why the macroeconomic moments are mostly unchanged except for the unconditional volatility of physical investment.

The column referred to as 'Convex W' modifies the benchmark model to incorporate a reduced-form convex wage schedule as a function of labor hours that is parametrized according to the empirical specification used to correct our supply-based measure (described in Section 4.2.1). While accounting for convexity in wages is an important adjustment to the empirical supply-based measure, we abstract from such distortions in our benchmark model by assuming perfect competition in factor markets to focus on the effects of strategic interactions among firms on asset prices. Competitive labor markets imply that the wage bill is linear with respect to hours. The results presented in this column highlight how the macroeconomic dynamics and main asset pricing results are robust to this modification to wages.

The column labelled as 'Single' considers the homogeneous industry case in which fixed operating costs are the same across all industries ($f_H = f_L$). The column referred to as 'Single - No F' examines the homogeneous industry case in which there are no fixed operating costs ($f_H = f_L = 0$). The macroeconomic fluctuations and aggregate asset pricing results, including return predictability, are preserved with the main exception being that the equity premium is somewhat reduced absent operating costs (5.58% compared to 7.30%). These two specifications illustrate that the heterogeneity across industry sectors is not important for time-series return predictability nor aggregate fluctuations in our benchmark model.

Table 9 illustrates that the extensive margin is the key source of excess return predictability in our model, attributing to oligopolistic competition. We compare population forecasting regressions with the price-dividend ratio between our benchmark model and an alternative specification, labeled as 'No-Entry', in which the extensive margin is shut down by imposing a constant mass of firms across all industries. Without entry, risk premia are effectively constant, as evidenced by the slope coefficients and R^2 being zero.

6 Conclusion

This paper highlights how endogenous firm entry in the presence of oligopolistic competition provides a powerful risk propagation mechanism, emanating from a negative and convex relation between industry markups and competition. When industry concentration is high, active firms enjoy greater market power by charging higher markups, but are more exposed to the risk of being undercut by new entrants. As entry dynamics depend on aggregate and industry business conditions, a positive link between markups and risk premia manifests over time and across firms. To test the predictions of the model, two empirical measures of markups are constructed from demand- and a supply-side equilibrium conditions. We find that markups are strong positive predictors of future excess returns, both in the time series and in the cross section, in accordance with the model.

Our model is used to interpret recent trends in industry concentration, macroeconomic uncertainty, and risk premia witnessed since the mid-1980s. Ascribing the secular rise in industry concentration to increasing barriers to entry, our model can reproduce the persistent rise in markups, consumption volatility, and the equity premium. Convexity in the equilibrium markup relation implies that rising industry concentration exposes firms more to entry risk, coalescing into a positive trend in macroeconomic uncertainty and aggregate risk premia. Overall, our model provides a coherent account of the systematic relation between competition and risk premia exhibited across different frequencies and across firms.

References

Ai, H., 2010. Information Quality and Long-Run Risk: Asset Pricing Implications. The Journal of Finance 65(4), 1333–1367.

Ai, H., M. M. Croce, A. M. Diercks, K. Li, 2015. News shocks and production-based term structure of equity returns. Working Paper, University of Minnesota.

Ai, H., M. M. Croce, K. Li, 2013. Toward a quantitative general equilibrium asset pricing model with intangible capital. Review of Financial Studies 26(2), 491–530.

Ali, A., S. Klasa, E. Yeung, 2008. The limitations of industry concentration measures constructed with Compustat data: Implications for finance research. The Review of Financial Studies 22(10), 3839–3871.

Bansal, R., A. Yaron, 2004. Risks for the long run: A potential resolution of asset pricing puzzles. The Journal of Finance 59(4), 1481–1509.

Barrot, J.-N., E. Loualiche, J. Sauvagnat, 2018. The globalization risk premium. Journal of Finance, Forthcoming.

Belo, F., X. Lin, S. Bazdresch, 2014. Labor hiring, investment, and stock return predictability in the cross section. Journal of Political Economy 122(1), 129–177.

Bena, J., L. Garlappi, P. Grüning, 2015. Heterogeneous innovation, firm creation and destruction, and asset prices. The Review of Asset Pricing Studies 6(1), 46–87.

Bessembinder, H., M. J. Cooper, F. Zhang, 2018. Characteristic-based benchmark returns and corporate events. The Review of Financial Studies 32(1), 75–125.

Bilbiie, F. O., F. Ghironi, M. J. Melitz, 2012. Endogenous entry, product variety, and business cycles. Journal of Political Economy 120(2), 304–345.

Bils, M., 1987. The cyclical behavior of marginal cost and price. The American Economic Review 77(5), 838–855.

Broda, C., D. E. Weinstein, 2010. Product creation and destruction: Evidence and price implications. The American economic review 100(3), 691–723.

Bustamante, M., A. Donangelo, 2017. Product market competition and industry returns. Review of Financial Studies, Forthcoming.

Chatterjee, S., R. Cooper, 2014. Entry and exit, product variety, and the business cycle. Economic Inquiry 52(4), 1466–1484.

Christiano, L. J., T. J. Fitzgerald, 2003. The band pass filter. international economic review 44(2), 435–465.

Clementi, G. L., D. Palazzo, 2016. Entry, exit, firm dynamics, and aggregate fluctuations. American Economic Journal: Macroeconomics 8, 1–41.

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

De Loecker, J., J. Eeckhout, 2017. The rise of market power and the macroeconomic implications. Unpublished working paper. National Bureau of Economic Research.

Dew-Becker, I., 2014. Bond pricing with a time-varying price of risk in an estimated medium-scale bayesian DSGE model. Journal of Money, Credit, and Banking 46(5), 837–888.

Dew-Becker, I., S. Giglio, 2016. Asset pricing in the frequency domain: theory and empirics. The Review of Financial Studies 29(8), 2029–2068.

Eckel, C., J. P. Neary, 2010. Multi-product firms and flexible manufacturing in the global economy. The Review of Economic Studies 77(1), 188–217.

Epstein, L., S. Zin, 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. Econometrica 57(4), 937–69.

Erickson, T., T. M. Whited, 2000. Measurement error and the relationship between investment and q. Journal of Political Economy 108(5), 1027–1057.

Farhi, E., F. Gourio, 2018. Accounting for macro-finance trends: Market power, intangibles, and risk premia. Unpublished working paper. National Bureau of Economic Research.

Favilukis, J., X. Lin, 2015. Wage rigidity: A quantitative solution to several asset pricing puzzles. Review of Financial Studies 29(1), 148–192.

Favilukis, J., X. Lin, 2016. Does wage rigidity make firms riskier? Evidence from long-horizon return predictability. Journal of Monetary Economics 78, 80–95.

Favilukis, J., X. Lin, X. Zhao, 2017. The Elephant in the Room: the Impact of Labor Obligations on Credit Markets,. Unpublished working paper. University of British Columbia.

Garleanu, N., S. Panageas, D. Papanikolaou, J. Yu, 2016. Drifting apart: The pricing of assets when the benefits of growth are not shared equally. Working paper, UC Berkeley.

Garleanu, N., S. Panageas, J. Yu, 2012. Technological growth and asset pricing. The Journal of Finance 67(4), 1265–1292.

Gomes, J. F., L. Schmid, 2016. Equilibrium asset pricing with leverage and default. Working paper, University of Pennsylvania.

Gourio, F., 2007. Labor leverage, firms? heterogeneous sensitivities to the business cycle, and the cross-section of expected returns. Unpublished working paper, Boston University.

Gourio, F., 2012. Disaster risk and business cycles. The American Economic Review 102(6), 2734–2766.

Greenwald, D. L., M. Lettau, S. C. Ludvigson, 2019. How the Wealth Was Won: Factors Shares as Market Fundamentals. Unpublished working paper. National Bureau of Economic Research.

Grullon, G., Y. Larkin, R. Michaely, 2018. Are US industries becoming more concentrated?. Forthcoming, Review of Finance.

Gutiérrez, G., T. Philippon, 2016. Investment-less Growth: An Empirical Investigation. Working Paper, New York University.

Hall, R., 1986. Market structure and macroeconomic fluctuations. Brookings papers on economic activity 1986(2), 285–338.

Haushalter, D., S. Klasa, W. F. Maxwell, 2007. The influence of product market dynamics on a firm's cash holdings and hedging behavior. Journal of Financial Economics 84(3), 797–825.

Herskovic, B., 2015. Networks in production: Asset pricing implications. Journal of Finance.

Herskovic, B., B. Kelly, H. Lustig, S. Van Nieuwerburgh, 2016. The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications. Journal of Financial Economics 119(2), 249–283.

Jaimovich, N., M. Floetotto, 2008. Firm dynamics, markup variations, and the business cycle. Journal of Monetary Economics 55(7), 1238–1252.

Jermann, U. J., 1998. Asset pricing in production economies. Journal of Monetary Economics 41(2), 257–275.

Justiniano, A., G. E. Primiceri, A. Tambalotti, 2010. Investment shocks and business cycles. Journal of Monetary Economics 57(2), 132–145.

Kaltenbrunner, G., L. A. Lochstoer, 2010. Long-run risk through consumption smoothing. Review of Financial Studies 23(8), 3190–3224.

Kogan, L., D. Papanikolaou, N. Stoffman, 2013. Winners and losers: Creative destruction and the stock market. Unpublished working paper. National Bureau of Economic Research.

Kuehn, L.-A., 2007a. Asset pricing with real investment commitment. Working Paper, Carnegie Mellon University.

Kuehn, L.-A., 2007b. The importance of time-to-build for investment-based asset pricing. Working Paper, Carnegie Mellon University.

Kuehn, L.-A., N. Petrosky-Nadeau, L. Zhang, 2014. Endogenous disasters. Working Paper, Carnegie Mellon University.

Kung, H., 2015. Macroeconomic linkages between monetary policy and the term structure of interest rates. Journal of Financial Economics 115(1), 42–57.

Kung, H., L. Schmid, 2015. Innovation, growth, and asset prices. The Journal of Finance 70(3), 1001–1037.

Lettau, M., S. Ludvigson, 2001. Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. Journal of political economy 109(6), 1238–1287.

Lin, X., D. Palazzo, F. Yang, 2017. The risks of old age: Asset pricing implication of technology adoption. Working Paper, Ohio State University.

Loualiche, E., 2016. Asset pricing with entry and imperfect competition. Working Paper, University of Minnesota.

Lyandres, E., D. Palazzo, 2016. Cash Holdings, Competition, and Innovation. Journal of Financial and Quantitative Analysis 51(6), 1–39.

Opp, M. M., C. A. Parlour, J. Walden, 2014. Markup cycles, dynamic misallocation, and amplification. Journal of Economic Theory 154, 126–161.

Pistaferri, L., 2003. Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply. Journal of Labor Economics 21(3), 729–754.

Rotemberg, J. J., M. Woodford, 1999. The cyclical behavior of prices and costs. Handbook of Macroeconomics 1, 1051–1135.

Stambaugh, R., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

Table 1: Parametrization

Panel A: Moments

Description	Simulated moments	Data	Standard error
Average output growth	2.00%	2.00%	0.349%
Average risk-free rate	1.00%	1.00%	0.294%
Relative investment volatility ($\sigma_{\Delta i}/\sigma_{\Delta c}$)	4.42	4.28	0.295
Relative R&D volatility ($\sigma_{\Delta s}/\sigma_{\Delta c}$)	2.49	2.43	0.345
First autocorrelation of investment rate	0.937	0.937	0.013
First autocorrelation of R&D intensity	0.977	0.988	0.056
Consumption growth volatility (Δc)	1.40%	1.40%	0.093%
Elasticity of markup w.r.t number of firms	-0.138	-0.126	0.029

Panel B: Parameter estimates

ζ_k	ζ_z	σ_k	σ_z	$a^\star$	σ	β	ν_2
0.531	0.922	0.037	0.000	-0.005	0.016	0.991	5.868
(0.027)	(0.350)	(0.044)	(0.000)	(0.087)	(0.001)	(0.002)	(0.006)

Panel C: Calibrated parameters

ψ	1.750	η	0.100	δ_n	2.00%	ρ^4	0.950
γ	10.00	δ_k	2.00%	ν_1	1.150	f_L	0.001
χ	3.000	δ_z	3.75%	κ	1.226	f_H	0.034
α	0.330						

Panel D: Non-linearities

	Model		Data	
$\mathcal{N}$	-0.971	-3.788	-0.721	-4.466
	(0.042)	(0.385)	(0.109)	(1.015)
$\mathcal{N}^2$		2.821		3.757
		(0.380)		(1.062)
R^2	0.944	0.971	0.523	0.626

This table summarizes the parametrization of the benchmark model. Panel A reports the simulated moments used in the simulated method of moments estimation, the empirical targets, and the standard error of the empirical moments, corrected for Newey-West. The sample period is 1948:Q1-2016:Q4. Panel B reports the estimated structural parameters with the corresponding standard errors reported below in parenthesis. Panel C reports the calibrated parameters values. Panel D presents regression estimates on the relation between the price markup and business formation. In particular, we report the coefficient estimates for the following time-series regressions run in both the model and the data: $\varphi_t^s = \alpha_0 + \beta_1\mathcal{N}_t + \beta_2\mathcal{N}_t^2 + \epsilon_t$. Newey-West standard errors are reported below in parenthesis.

Table 2: Simulated Moments

Panel A. Moments

	Data	Model
I. Macro moments		
E[Δc]	2.00%	2.00%
$\sigma_{\Delta c}/\sigma_{\Delta y}$	0.62	0.48
$\sigma_{\Delta i}/\sigma_{\Delta c}$	4.28	4.42
$\sigma_{\Delta s}/\sigma_{\Delta c}$	2.43	2.49
$\sigma_{\Delta c}$	1.40%	1.40%
$\sigma_{\Delta l}$	0.64%	1.00%
II. Industry moments		
$\sigma[\varphi^s]$	2.42%	1.38%
$\sigma[\varphi^d]$	1.17%	0.71%
corr$[\varphi^s, Y]$	-0.250	-0.911
corr$[\varphi^d, Y]$	-0.696	-0.668
corr$[\varphi^s, \varphi^d]$	0.443	0.820
III. Asset pricing moments		
E$[r_f]$	1.00%	1.00%
E$[r_d - r_f]$	7.74%	7.38%
E$[pd]$	3.49	3.32
$\sigma[r_f]$	1.17%	1.66%
$\sigma[r_d - r_f]$	16.31%	4.21%
$\sigma[pd]$	0.43	0.25

Panel B. Frequency decomposition

	Data	Model
I. Business cycle		
Δn	15.10%	41.51%
Δz	11.42%	8.29%
II. Growth cycle		
Δn	6.33%	14.79%
Δz	42.37%	60.79%

This table reports several moments from the data and the model. Panel A reports a series of macroeconomic, industry, and asset pricing moments. Reported model moments are averaged across 100 simulations that are equivalent in length to the data sample. Growth rate moments are annualized percentage. Correlation coefficients are obtained after filtering the data series with a bandpass filter that isolates frequencies between 1.5 and 20 years. The risk premiums are levered following Boldrin, Christiano, and Fisher (2001). Panel B reports the R^2 from consumption growth regressions over various frequencies using the change in the number of firms (Δn) and the growth in the stock of R&D (Δz). The business (growth) cycle component of consumption growth is obtained by isolating frequencies between 1.5 and 20 years (above 20 years). The sample period is 1948:Q1-2016:Q4.

Table 3: Consumption Volatility Dynamics

	Data	Model
A. Summary Statistics		
$\mathrm{E}[\sigma_{\Delta c}]$	0.622%	0.405%
$\sigma[\sigma_{\Delta c}]$	0.195%	0.061%
$\mathrm{AC1}[\sigma_{\Delta c}]$	0.852	0.624
B. Forecasts with φ^s		
β	0.118	0.104
S.E.	0.026	0.007
R^2	0.169	0.416
C. Forecasts with φ^d		
β	0.058	0.103
S.E.	0.028	0.008
R^2	0.040	0.402
D. Forecasts with pd		
β	-0.143	-0.070
S.E.	0.026	0.011
R^2	0.242	0.214

This table presents summary statistics (Panel A) and forecasting regressions (Panels B to D) for the conditional consumption growth volatility ($\sigma_{\Delta c}$) from the data and the model. Forecasting regressions are obtained as follows: $\sigma_{\Delta ct} = \alpha + \beta x_t + \epsilon_t$, where x_t corresponds to φ_t^s (Panel B), φ_t^d (Panel C), and the price-dividend Ratio (Panel D). The series for the conditional volatility of consumption growth is computed by first fitting realized consumption growth to an AR(1) process: $\Delta c_t = \beta_0 + \beta_1 \Delta c_{t-1} + u_t$. Next, $\sigma_{\Delta ct}$ is obtained using a EGARCH(1,1) on the estimated innovations $\hat{u}_t$. All independent variables are normalized and standard errors, S.E., are corrected for heteroscedasticity. The sample period is 1948:Q1-2016:Q4. Reported model moments are averaged across 100 simulations that are equivalent in length to the data sample.

Table 4: Stock Return Predictability

	Data			Model		
	Horizon (in years)					
	1	3	5	1	3	5
A. Markup - φ^s						
$\beta^{(n)}$	1.610	3.570	5.841	0.771	2.155	3.324
S.E.	0.698	1.432	2.094	0.276	0.701	1.010
R^2	0.051	0.107	0.182	0.049	0.135	0.207
B. Markup - φ^d						
$\beta^{(n)}$	3.541	6.528	9.417	1.437	3.967	6.118
S.E.	1.382	2.364	4.266	0.539	1.356	1.939
R^2	0.055	0.079	0.102	0.047	0.125	0.191
C. Price-dividend ratio						
$\beta^{(n)}$	-0.104	-0.220	-0.353	-0.027	-0.075	-0.111
S.E.	0.041	0.088	0.093	0.014	0.034	0.049
R^2	0.072	0.135	0.217	0.027	0.072	0.100

This table reports excess stock return forecasts for horizons of one, three, and five years in the data and the model, i.e. $r^{ex}_{t,t+n} - y^{(n)}_t = \alpha_n + \beta\ x_t + \epsilon_{t+1}$, where x_t is the predicting variable. The different panels present forecasting regressions using different predicting variables: φ^s (panel A), φ^d (panel B), and the log-price-dividend ratio (panel C). The price markup measures are linearly detrended and averaged over the preceding four quarters. The forecasting regressions use overlapping quarterly observations. $\beta^{(n)}$ is the coefficient estimate obtained via OLS. S.E. is the standard error, corrected for heteroscedasticity using Newey-West with $k+1$ lags. The sample period is 1948:Q1-2016:Q4. Model moments are averaged across 100 simulations that are equivalent in length to the data sample. The risk premiums are levered following Boldrin, Christiano, and Fisher (2001).

Table 5: Horse Race

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
s_L	-3.45			0.22	-2.29		
	(3.02)			(2.37)	(2.82)		
φ^s		5.84		5.92		3.66	
		(2.10)		(2.21)		(1.71)	
φ^d			9.42		8.58		8.89
			(4.28)		(3.79)		(4.05)
cay						7.73	8.66
						(1.66)	(1.60)
R^2	0.037	0.182	0.102	0.182	0.118	0.342	0.372

This table compares the five-year excess stock return forecasts for different predictors. Each column corresponds to a different set of predicting variables. The predicting variables are the log-labor share s_L, the two measures of price markup φ^s and φ^d, and the consumption-wealth ratio cay. The price markup measures and the labor share are linearly detrended and averaged over the preceding four quarters. The forecasting regressions use overlapping quarterly observations. Standard errors are corrected for heteroscedasticity using Newey-West with $k+1$ lags and are reported below each coefficient estimate in parentheses. The sample period is 1952:Q1-2016:Q4 for cay, and 1948:Q1-2016:Q4 for s_L, φ^s, and φ^d.

Table 6: Cross-sectional return spread

Panel A. Return spread

	φ^s		φ^d	
	Data	Model	Data	Model
E$[r_H - r_L]$	3.22%	3.63%	4.03%	3.63%
S.E.	(2.72%)	(0.30%)	(2.15%)	(0.30%)

Panel B. Characteristics-corrected spread

	φ^s		φ^d	
	Fama-French	Characteristics	Fama-French	Characteristics
E$[r_H - r_L]$	6.96%	6.66%	4.20%	3.67%
S.E.	(2.69%)	(2.44%)	(1.96%)	(1.49%)

This table reports the cross-sectional return spread for portfolios sorted on price markup. Panel A reports the return on the zero investment portfolio that is long the portfolio of high markup industries and short the portfolio of low-markup industries both in the data and the model. High- (Low-) markup industries are defined as the set of industries making up the highest (lowest) tercile of the cross-sectional distribution of markups. Panel B reports the same zero investment portfolio strategy, except that returns are first adjusted for firm characteristics following the methodology in Bessembinder, Cooper, and Zhang (2019). In particular, column "Fama-French" controls for the size and book-to-market characteristics. Column "characteristics" controls for size, book-to-market, momentum, asset growth, profitability, leverage, operating leverage, and asset utilization. Quarterly portfolio returns are annualized. Newey-West standard errors are reported below in parentheses.

Table 7: Secular trend and asset prices

	Data			Model		
	1984-2000	2001-2016	Diff.	1984-2000	2001-2016	Diff.
Markup	30.23%	39.77%	9.55%***	30.26%	39.78%	9.53%***
$\sigma_{\Delta c}$	1.06%	1.26%	0.20%*	1.16%	1.41%	0.25%***
$\mathrm{E}[r_d - r_f]$	2.00%	7.05%	5.05%**	7.34%	8.84%	1.50%***

This table reports the average price markup, the volatility of consumption growth $\sigma_{\Delta c}$, and the equity risk premium $\mathrm{E}[r_d - r_f]$ for two subsamples of the data - 1984-2000, and 2001-2016 - and compares it to simulated moments from the model. The equity risk premium in the data is obtained from the Gordon growth model $E[r_d - r_f] = E[D/P] + E[\Delta d] - E[r_f]$ using data from Bob Shiller website. Model moments are obtained by augmenting the benchmark model with persistent entry cost shocks and simulating the model when entry costs are persistently high (high markup case) and persistently low (low markup case). The calibration used in the simulation is the same as the benchmark model. *,**,*** indicates significance at the 10%, 5%, and 1% level for the test of difference between the two subamples.

Table 8: Additional specifications

	Benchmark	No IST	Convex W	Single	Single - no F
A. Macro moments					
$E[\Delta c]$	2.00%	2.00%	2.26%	2.00%	2.00%
$\sigma_{\Delta c}/\sigma_{\Delta y}$	0.48	0.48	0.54	0.45	0.44
$\sigma_{\Delta i}/\sigma_{\Delta c}$	4.45	2.17	4.10	4.49	4.53
$\sigma_{\Delta s}/\sigma_{\Delta c}$	2.50	2.50	2.45	2.47	2.55
$\sigma_{\Delta c}$	1.39%	1.40%	1.52%	1.40%	1.40%
$\sigma_{\Delta l}$	1.01%	1.00%	0.82%	1.13%	1.31%
B. Asset pricing moments					
$E[r_f]$	1.04%	0.99%	0.05%	1.00%	1.00%
$E[r_d - r_f]$	7.30%	7.38%	9.80%	8.18%	5.58%
$E[pd]$	3.32	3.32	3.24	3.16	3.66
$\sigma[r_f]$	1.66%	1.66%	1.97%	1.91%	2.03%
$\sigma[r_d - r_f]$	4.20%	4.20%	3.56%	3.91%	1.74%
$\sigma[pd]$	0.25	0.25	0.26	0.23	0.37
C. Predictability regressions - φ^s					
$\beta^{(1)}$	0.783	0.769	0.645	0.882	0.512
S.E.	0.263	0.258	0.196	0.246	0.144
R^2	0.049	0.049	0.065	0.077	0.076
$\beta^{(3)}$	2.188	2.151	1.872	2.466	1.438
S.E.	0.699	0.686	0.533	0.630	0.376
R^2	0.134	0.135	0.173	0.201	0.194
$\beta^{(5)}$	3.375	3.319	2.986	3.839	2.230
S.E.	1.017	0.999	0.798	0.885	0.541
R^2	0.205	0.207	0.260	0.298	0.285

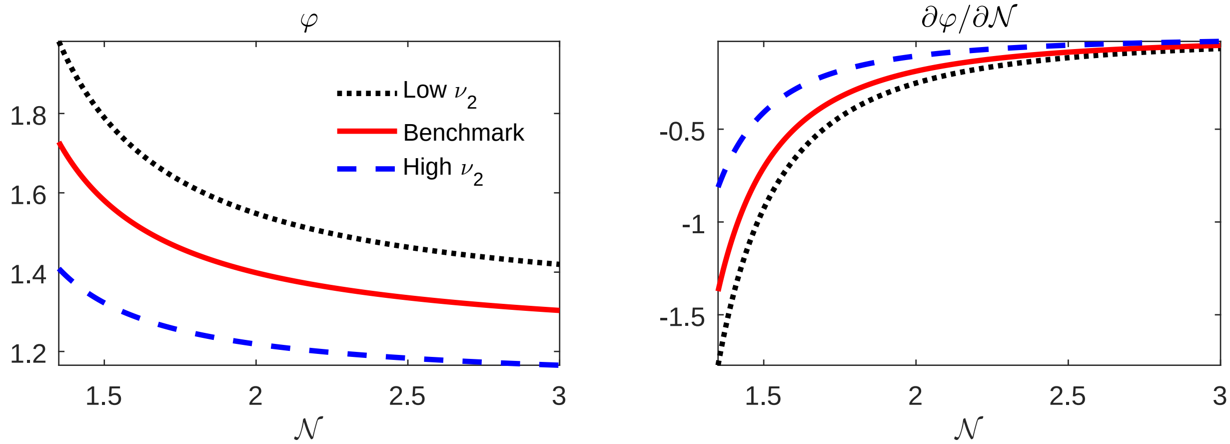
This table compares various macroeconomic, asset pricing and return predictability moments of the model to several other specifications. *No IST* is the benchmark calibration with no IST shocks, i.e. $\sigma_z = \sigma_k = 0$. *Convex W* is the same calibration as the benchmark model except that the total firm wage bill W_t^T is convex in labor, i.e, $W_t^T = W_t L_t^{1.4}$. *Single* is a re-estimated specification of the benchmark model with only one representative industry. *Single - no FC* is a re-estimated specification of the benchmark model with only one representative industry and no operating fixed cost. Reported model moments are averaged across 100 simulations that are equivalent in length to the data sample. Growth rate moments are annualized percentage. The risk premiums are levered following Boldrin, Christiano, and Fisher (2001).

Table 9: Stock Return Predictability with *pd*

	Horizon (in years)		
	1	3	5
Panel A. Model			
$\beta^{(n)}$	-0.014	-0.045	-0.072
R^2	0.009	0.031	0.049
Panel B. No-entry			
$\beta^{(n)}$	0.000	0.000	0.000
R^2	0.000	0.001	0.002

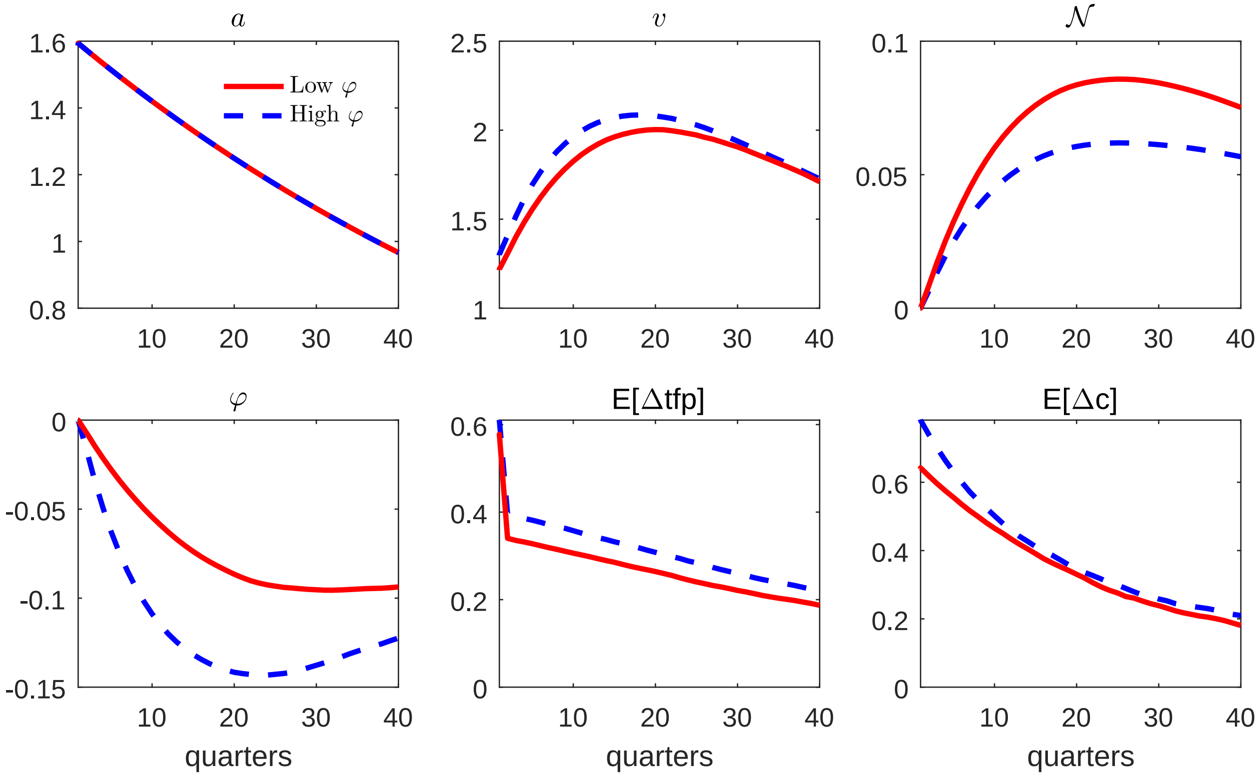
This table reports excess stock return forecasts for horizons of one, two, and three years using the log-price-dividend ratio: $r^{ex}_{t,t+n} - y^{(n)}_t = \alpha_n + \beta\ \log(P_t/D_t) + \epsilon_{t+1}$. Panel A presents population forecasting regressions for the Benchmark model. Panel B presents population forecasting regressions for an alternative specification of the model with no entry, i.e., where markup are constant. The forecasting regressions use overlapping quarterly data. The risk premiums are levered following Boldrin, Christiano, and Fisher (2001).

Figure 1: Equilibrium markup relation



This figure plots the relation between the level of price markup and the number of firms (left), and the first derivative of the price markup with respect to the number of firms (right) for three values of ν_2. The Low and High ν_2 cases are obtained by setting ν_2 to 4.5 and 10, respectively.

Figure 2: Responses conditional on markup



This figure compares the impulse response functions to a positive technology shock, conditional on the level of aggregate price markup, φ, for the exogenous technology process a, the firm value v, the aggregate number of firms $\mathcal{N}$, the price markup φ, expected total factor productivity growth $\mathrm{E}[\Delta tfp]$, and expected consumption growth $E[\Delta c]$. The High φ (Low φ) case corresponds to the average responses accross 250 draws in the highest (lowest) quintile sorted on φ_t. The data for the sorting is obtained by simulating the economy for 50 periods prior to the realization of the positive technology shock. All values on the y-axis are percentage deviation from the steady state. Growth rates variables are annualized.