Markup Shocks and Asset Prices

May 30, 2023

Abstract

We explore the asset pricing implications of shocks that allow firms to extract more rents from consumers. These markup shocks directly impact the representative household's marginal utility and the firms' cash flow. Using firm-level data, we construct a measure of aggregate markup shocks and show that the price of markup risk is negative, that is, a positive markup shock is associated with high marginal utility states. Markup shocks generate differences in risk premia due to their heterogeneous impact on firms. Firms with larger exposures to markup shocks are less risky and have lower expected returns. We rationalize these findings in a general equilibrium model with markup shocks.

Keywords: Production-based asset pricing, imperfect competition, time-varying markups, cross-section of stock returns, recursive preferences.

1 Introduction

The last two decades saw a sharp increase in the average price markup in many developed economies.[1] Higher markups increase the wedge between the price charged to customers and the marginal cost of production, allowing firms to boost profits at the expense of consumers. While the roots and consequences of the secular trends in markups is still debated,[2] exogenous markup shocks have proven quite successful in explaining business cycle fluctuations and have been adopted as a standard feature of many dynamic stochastic general equilibrium models (e.g., see Smets and Wouters (2003) and Ireland (2004)). Surprisingly, the asset pricing implications of markup shocks and their impact on the cross-section of firms have received little attention in the literature. The goal of this paper is to explore those implications – both theoretically and empirically.

We first formalize our predictions on the asset pricing implications of markup shocks in a real business cycle model, augmented with imperfect competition and markup shocks. The model also features endogenous long-run risks and recursive preferences to match the asset market data, jointly with macroeconomic dynamics quantitatively. Long-run risks in consumption and dividend growth are generated by a non-stationary process that features an exogenous component (e.g., Croce (2014)) and an endogenous component that depends on the accumulation of aggregate capital (e.g., Romer (1990)). 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 price of risk for shocks that impacts long-run consumption growth, such as productivity or markup shocks.

Markup shocks affect the firm's elasticity of demand and thus alter the trade-off between the optimal quantity and price chosen by the firm. A positive markup shock gives the firm more market power, which leads to an increase in the output price and a reduction in the quantity supplied. When markups are persistent, the reduction in future output leads to a persistent drop in the firm's investment and hired labor. At the aggregate level, the fall in investment forecasts

[1] De Loecker et al. (2020) find that the revenue-weighted average markup in the United States climbed from about 1.2 in 1980 to 1.6 in 2016. De Loecker and Eeckhout (2018) find similar results for Europe.

[2] Recent examples of papers in this literature include Gutiérrez and Philippon (2017), Farhi and Gourio (2018) Greenwald et al. (2019), Crouzet and Eberly (2019), Corhay et al. (2020b), Barkai (2020) and Gutiérrez et al. (2021).

low future consumption growth. In an economy with recursive utility and a preference for early resolution of uncertainty, it is bad news for the representative household. In short, the model predicts that the price of markup risk is negative.[3]

Markup shocks might not impact firms equally in the cross-section. For example, a new information system technology that allows firms to collect more data on consumers can increase the potential for price discrimination. This positive markup shock can benefit larger firms relatively more than smaller firms because of the enormous potential for data collection. In this scenario, large firms would have higher positive markup exposures while small firms would have lower or even negative markup exposures. We explore the asset pricing implications of allowing for differential firm-level exposures to the aggregate markup shock and find that it generates substantial cross-sectional differences in asset prices. Firms with a high (low) markup exposure earn, on average, a lower (higher) expected return. Thus, our model predicts that markup risk is an important driver of the cross-section of stock returns.

The asset pricing implications of firms' *exposure* to markup shocks contrast with those related to the average *level* of a firm's markups. We highlight this difference by allowing for heterogeneity in both the average level and the sensitivity of markups. We show that although a higher markup is associated with a higher expected return, a higher markup exposure is associated with a lower expected return. This can be explained as follows. A higher level of price markup endogenously increases the firm's exposure to aggregate productivity risk, because monopolistic rents make firms' profits more pro-cyclical. In other words, the level of markup positively impacts the firm's productivity beta. Because aggregate productivity carries a positive price of risk, high markup firms are more risky.[4] Rather, this paper identifies a new source of priced risk that exogenously impacts firms' markup and carries a negative price. Our paper thus provides a new channel through which markups can affect asset prices.

To test these new predictions, we construct an empirical series for the aggregate price markup shocks. Our procedure follows a bottom-up approach. First, we estimate markup surprises at the

[3]While the combination of long-run risk and a preference for early resolution of uncertainty allows us to get reasonable asset pricing implications. Our results are robust to using alternative utility functions, such as habits (Campbell and Cochrane (1999)). In addition, our results are robust to allowing for sticky price in a standard New-Keynesian model á la Galí (2015).

[4]See, for example, Bustamante and Donangelo (2017), and Corhay et al. (2020a).

firm level. The strategy to estimate firm-level markups builds on the production approach pioneered by Hall (1988). This approach obtains firm-level markups by exploiting the cost-minimization of a variable input for production. The price markup is computed as the revenue share of the variable input, multiplied by the output elasticity of the variable input. Obtaining the elasticity, however, requires the estimation of a production function, which we estimate using publicly available data from Compustat following the procedure in De Loecker et al. (2020). We then compute markup surprises for each firm and year using a predictive model that uses a series of characteristics shown to explain firm markups. The aggregate markup shock is constructed by aggregating firm-level shocks each year.

Our constructed series of aggregate markup shocks has business cycle properties consistent with the predictions of most business cycle models. In particular, we show that a surprise increase in markups is associated with a decline in economic activity as measured by future consumption, GDP, and investment. Higher markups are also associated with higher future inflation. Notably, while an increase in markups is associated with bad news for the economy, we show that firm valuations increase in response to the shock. This pattern appears because markups increase the present value of future rents of the firm. These results highlight an important contrast between the business cycle properties of standard productivity shocks and markup shocks and suggest that markup shocks contain important information about aggregate risk not contained in traditional aggregate productivity measures.

We find strong empirical evidence that the price of markup risk is negative, as predicted by the model. We estimate a two-factor linear asset pricing model with the market factor and the constructed markup shock. While the price of market risk is positive, the estimated factor loading on the markup shock is negative and significantly different from zero. The economic magnitude is also large. Our estimation implies the price of markup risk to be around -1, which corresponds to about 15% of the estimated price of market risk. The two-factor model significantly improves the performance of the one-factor model. Indeed, adding the markup shock reduces the mean pricing absolute pricing errors by 50%. In short, we provide novel empirical evidence that markup shocks are an important source of priced risk.

Markup shocks have heterogeneous effects in the cross-section of firms. We estimate exposures to the aggregate markup shocks using asset pricing data, i.e., markup betas, for various portfolios formed based on standard characteristics in the literature, such as book-to-market, size, investment, and profitability. We show that the exposures to markup shocks differ a lot across firms, and they can explain the documented return spreads for many standard portfolios. For example, large firms have a larger markup beta than smaller firms. Since the price of markup risk is negative, this implies that larger firms have lower expected returns than smaller firms, which is consistent with the documented size premium. Similarly, value firms load negatively on the markup shock while growth firms load positively, which explains the value premium. We also build portfolios by sorting firms based on their stock returns' markup betas and find that the portfolio that goes long high-markup betas and short low-markup betas earn a significant return of about -3.50% per annum. Thus, the exposure to markup shocks can explain a wide range of portfolio sorts previously documented in the literature and is consistent with the model predictions.

In the model, we argue that the fundamental force driving the impact of markup shocks in the cross-section of firms stems from individual firms' markup exposures to the aggregate markup shock. In order to bring additional corroborating evidence for this key economic channel, we directly estimate markup exposures by regressing firm-level markup surprises on the aggregate markup shock series. Thus, we do not rely on asset price data to estimate the exposure to the markup shock. We then build portfolios by sorting firms based on their fundamental markup exposures. In line with the results from return-based markup betas, we find that the portfolio with high markup exposure earns a significantly lower return than the portfolio with low markup exposure. These findings suggest that the dynamics of markups at the firm level, especially their exposure to the markup shocks, is an important characteristic driving the cross-section of asset prices.

Finally, we test the model predictions on the relation between the level of expected markup and risk, and contrast it with our novel channel working through the exposure to markup shocks. In particular, we run cross-sectional Fama-MacBeth regressions using the markup beta, the expected markup, and a battery of controls. Consistent with the model predictions and the previous results, firms with higher markup betas earn lower future returns. We also find that consistent with the existing literature (e.g., Corhay et al. (2020a)), a higher expected markup is associated with higher future returns. Importantly, we show that both effects remain statistically significant when considered together. Our results suggest that it is important to account for both the level of markups and the exposure to markup shocks to fully understand the relation between a firm's market power and assets prices.

Our analysis is intentionally agnostic about the fundamental forces driving the exogenous markup process and why firms have different exposure to that shock. The objective is to have the simplest framework to help us understand the asset pricing implications of shocks that increase the wedge between the price and the marginal cost – whatever their sources. The literature has come up with several potential channels that could endogenize our assumptions. Some examples include intangible capital (e.g., Crouzet and Eberly (2019) and Crouzet and Eberly (2021)), the ability to collect and exploit data (e.g., Begenau, Farboodi, and Veldkamp (2018), and Eeckhout and Veldkamp (2021)), firms' monopsony power (e.g., Manning (2013)), or barriers to entry (e.g., Gutiérrez, Jones, and Philippon (2019)). Providing micro-foundations goes beyond the scope of this paper, but we hope that our findings will spur additional research aimed at understanding the fundamental sources behind markup shocks and their impact on firms.

This paper relates to the burgeoning literature in finance studying the impact of markup power on asset prices in equilibrium models with production and strategic interactions, e.g., Aguerrevere (2009); Bustamante and Donangelo (2017); Loualiche (Forthcoming); Corhay, Kung, and Schmid (2020a); Dou, Ji, and Wu (2021a); Dou and Ji (2021); Dou, Ji, and Wu (2021b); Doshi and Kumar (2021).[5] In these studies, market power matters for risk because it amplifies the exposure to aggregate productivity risk. In contrast, our paper identifies a new source of priced risk – unrelated, but complementary to productivity risk – that directly impacts price markups in the cross-section of firms. We show that accounting for this risk is key to fully understand the impact of market power on asset prices.

Closely related to our paper is Cho, Grotteria, Kremens, and Kung (2021) who use a Campbell and Shiller (1988) decomposition to decompose the market-to-book ratio into long-run expectations

[5] Corhay (2017), Chen, Dou, Guo, and Ji (2020), and Dou, Ji, and Wu (2021a) allow for default and find that accounting for credit risk is important to understand the impact of competition on asset prices. Clara (2018) examines the implications of demand elasticities in the presence of nominal rigidities on asset prices. Clara, Corhay, and Kung (2021) study multi-product firms.

about a firm's cash flows and long-run expectations about its asset discount rates. They find that future markups play an important role in explaining the cash-flow components in the cross-section of firms. Our paper is similar in that we show the importance of future markups for the cross-section of asset prices but the focus and mechanisms are different. Cho, Grotteria, Kremens, and Kung (2021) show the importance of future expected markups at the *firm level* to explain the observed dispersion in book-to-market ratios via a cash-flow channel. In contrast, we study the asset pricing implications of the *common* component that drives firm-level markup shocks. We show that this shock is priced and that heterogeneous exposures to it can explain portfolios sorted on various firm-level characteristics.

More broadly, our work is closely related to a growing literature that studies asset prices in production economies (e.g., Cochrane (1991), Jermann (1998) Zhang (2005)). For the most part, this literature has focused on aggregate shocks that affects aggregate productivity (e.g., Favilukis and Lin (2013), Kung and Schmid (2015)), investment-specific shocks (e.g., Kogan and Papanikolaou (2014), Knesl (2018)), shocks to fiscal and monetary policies (e.g., Belo and Yu (2013), Croce, Nguyen, Raymond, and Schmid (2019), Bretscher, Hsu, and Tamoni (2020a)), uncertainty shocks (e.g., Croce (2014), Bretscher, Hsu, and Tamoni (2020b)), shocks to external financing (e.g., Belo, Lin, and Yang (2019)), or shocks to capital depreciation (e.g., Ai, Li, and Tong (2022)). In contrast to this literature, this paper explores the consequences of shocks that affects the wedge between the output price and the marginal cost on the cross sectional variation of stock returns.

Several studies, however, have documented the importance of aggregate shocks – different from standard productivity shocks – that jointly affect firms' market power and asset prices. Loualiche (Forthcoming) documents that aggregate entry risk is priced in the cross-section of industry portfolios. He finds that industries that are more (negatively) exposed to entry risk command positive risk premia and estimate the price of entry risk to be negative. This happens because a positive entry shock increases new firms' entry, which increases competition and displaces incumbents' monopoly rents. Barrot, Loualiche, and Sauvagnat (2019), and Bretscher (2020) document a similar effect for the risk of output (input) import competition, respectively. In all these papers, the displacement shock carries a negative price of risk but leads to a *decrease* in firms' market power. In other words, the implied "markup shocks" would carry a positive price of risk. In contrast, we show that markup shocks carry a negative price of risk, suggesting that we are capturing a different effect. In addition, while these displacement shocks explain the cross-section of industry returns, we show that markup shocks price a wide variety of portfolios sorted on various firm-level characteristics. This suggests that our markup shock impacts firms at a more granular level than industries. Therefore, we see our results as providing a different but complementary channel.

We also relate to the literature in macroeconomics that explore the importance of markup shocks to explain business cycle fluctuations (e.g., Smets and Wouters (2003), Ireland (2004), Smets and Wouters (2007)). These papers show that markup shocks are important to explain inflation dynamics but have little explanatory power on macroeconomic quantities such as GDP. The representative firm assumption inherent in those studies might understate the impact of markup shocks. In fact, we show that markups shocks have large heterogeneous effects in the cross-section of firms. Given the importance of markups for the transmission of monetary policy, extending our framework into a Heterogeneous Agent New Keynesian model (e.g., Kaplan et al. (2018)) appears an interesting avenue for future research. Also related is Palomino (2012) who shows that the price of risk for markup shocks depends on the optimal monetary policy. The paper explores asset prices but only focuses on the term structure of interest rates. In contrast, we explore the impact of markup risk for the cross-section of stock returns.

The paper is organized as follows. Section 2 presents a dynamic general equilibrium model with exogenous markup shocks. Section 3 presents the calibration and the main results on the pricing of markup shocks. Section 4 describes the construction of our empirical measure of markup shock and tests several model predictions. Section 5 concludes.

2 Model

This section presents a general equilibrium model to study the impact of markup shocks on asset prices. The model has two sectors. The production sector produces consumption goods using labor and capital. The household sector is characterized by a representative household who makes consumption and saving decisions. They own all the assets in the economy. Markup shocks affect the firm's ability to extract monopolistic rents from households, which jointly affects the household marginal utility and firm valuations.

2.1 Production

Production consists of two layers. The top layer aggregates industry goods to produce the final consumption good. The bottom layer consists of industries with firms that produce differentiated products. Each industry is characterized by a monopolistic competition structure. Firms are subject to two types of aggregate shocks – productivity shocks that affect the marginal productivity of labor and markup shocks that impact the wedge between the output price and the marginal cost of production. We introduce heterogeneity across firms through differences in the average level of markups and their exposure to the aggregate markup shocks.

2.1.1 Final goods

Technology The final consumption good Y_t is produced using a two-tier production structure as in Rotemberg and Woodford (1992)). First, a continuum of industry goods of measure one are packaged using a constant elasticity of substitution (CES) aggregator:

$$Y_t = \left(\int_0^1 Y_{j,t}^{\frac{\tau-1}{\tau}} \, dj \right)^{\frac{\tau}{\tau-1}}, \tag{1}$$

where $Y_{j,t}$ is the output of industry j at time t and $\tau > 0$ is the elasticity of substitution across industry goods.

In turn, each industry good $Y_{j,t}$ is a CES aggregator that bundles a continuum of differentiated products of measure one according to:

$$Y_{j,t} = \left(\int_0^1 b_{ij} \, y_{ij,t}^{\frac{1}{\mu_{j,t}}} \, di \right)^{\mu_{j,t}}, \tag{2}$$

where $y_{ij,t}$ is the output of firm i in industry j and at time t, and b_{ij} denotes the share of each product in the industry-level CES aggregator. The shares across all firms and industries add up to unity $\int_0^1 b_{ij} di = 1$.

We introduce markup shocks in the model by assuming that parameter $\mu_{j,t}$ is time-varying due to exogenous shocks (e.g., see Justiniano et al. (2010)). As we show later, $\mu_{j,t}$ directly affects the wedge between the price charged by the firm and the marginal cost of production. As such, innovations to $\mu_{j,t}$ can be interpreted as markup shocks.[6]

Product demand Solving the cost minimization problem of the final goods producer yields the following inverse demand function for each industry good:

$$Y_{j,t} = Y_t \left(\frac{P_{j,t}}{P_t} \right)^{-\tau} \tag{3}$$

where $P_{j,t}$ is the market clearing price of industry good j and P_t is the aggregate price index, which we take as our numéraire, i.e., $P_t = 1$.[7]

Similarly, solving the cost minimization problem for industry k leads to the following inverse demand function for each product i in industry j:

$$y_{ij,t} = Y_{j,t} \left(\frac{p_{ij,t}}{P_{j,t}} \right)^{-\frac{\mu_{j,t}}{\mu_{j,t}-1}} b_{ij}^{\frac{\mu_{j,t}}{\mu_{j,t}-1}}. \tag{4}$$

Therefore the demand function for product i in industry j is given by:

$$y_{ij,t} = Y_t P_{j,t}^{\frac{\mu_{j,t}}{\mu_{j,t}-1}-\tau} p_{ij,t}^{-\frac{\mu_{j,t}}{\mu_{j,t}-1}} b_{ij}^{\frac{\mu_{j,t}}{\mu_{j,t}-1}}. \tag{5}$$

Note that the demand for a firm's product depends inversely on the price charged by the firm, that is, the demand schedule decreases in the output price $p_{ij,t}$. The demand elasticity is given by $\frac{\mu_{j,t}}{\mu_{j,t}-1}$. The larger this parameter, the more elastic demand is. The limiting case where $\mu_{j,t} \to 1$ implies an infinitely elastic demand, which corresponds to a perfectly competitive economy. Any specifications where $1 < \mu_{j,t} < \infty$ will grant the firm some degree of market power over the sale of its product.

[6] Parameter $\mu_{j,t}$ is also related to the elasticity of substitution across goods in industry j. Indeed, one can show that the elasticity of substitution $\omega_{j,t}$ is related to $\mu_{j,t}$ in the following way: $\omega_{j,t} = \frac{\mu_{j,t}}{\mu_{j,t}-1}$.

[7] One can show that the market clearing price in industry j is given by: $P_{j,t} = \left(\int_0^1 b_{ij}^{\frac{\mu_{j,t}}{\mu_{j,t}-1}} p_{ij,t}^{\frac{1}{1-\mu_{j,t}}} dj \right)^{1-\mu_{j,t}}$. Refer to the online appendix for details.

2.1.2 Intermediate firms

Technology The firm is equipped with Cobb-Douglas production technology. In particular, the individual firm's output is produced using pre-installed capital $k_{ij,t}$ and a fully flexible variable input $v_{ij,t}$ as follows:

$$y_{ij,t} \quad = \quad k_{ij,t}^{\alpha^k} \left(Z_t \; v_{ij,t}\right)^{\alpha^v}, \tag{6}$$

where α^k and α^v represent the share of capital and variable input in the production function, respectively. Additionally, we assume the production technology is constant returns to scale, such that $\alpha^k + \alpha^v = 1$. In a typical production function, the variable input is labor. However, in general, it can be any production inputs that the firm can freely adjust, such as the cost of material, electricity, etc.

The aggregate productivity Z_t is a non-stationary process that is the same for all intermediate firms in the economy. The growth rate of aggregate productivity Δz_t follows a mean-reverting process:

$$\begin{aligned} \Delta z_t =& \Delta \overline{z} + (1 - \phi_x) x_t + \phi_x \; \hat{ik}_t \\ x_t =& \rho_x x_{t-1} + \sigma_x \epsilon_t^x, \end{aligned} \tag{7}$$

where $\Delta \overline{z}$ is the productivity growth rate in the steady state and the hat-superscript denotes the log-deviations from the steady state. The time-varying component in aggregate productivity growth is a weighted average of two terms: a latent exogenous state variable x_t as in Croce (2014) and an endogenous component that depends on the aggregate investment rate, $\hat{ik}_t$. This specification is a parsimonious way to introduce long-run risk in the spirit of Bansal and Yaron (2004) without specifying a fully-fledged endogenous growth model à la Romer (1990). In the quantitative section, we calibrate ϕ_x to match salient features of the growth cycle as documented in Kung and Schmid (2015).

The variable input is obtained in competitive markets for a unit price of P_t^v. The firm's capital is accumulated through investment $i_{ij,t}$ and only becomes productive in the next period, i.e., time-to-build. The stock of productive capital accumulates as follows,

$$k_{ij,t+1} = (1-\delta_k)k_{ij,t} + \Gamma\left(\frac{i_{ij,t}}{k_{ij,t}}\right)k_{ij,t} \tag{8}$$

where δ_k is the depreciation rate of capital and $\Gamma(\cdot)$ captures the idea that capital accumulation is subject to adjustment costs. Those costs are key to quantitatively match investment dynamics jointly with asset prices (see Jermann (1998)).

The firm's dividend is defined as the total revenue minus variable costs, minus investment:

$$d_{ij,t} \quad = \quad p_{ij,t}y_{ij,t} - P_t^v v_{ij,t} - i_{ij,t}, \tag{9}$$

where $p_{ij,t}$ is the price charged by the firm on its product and P_t^v is the competitive variable input price, taken as given by the firm.

Optimality conditions The firm's objective is to choose the level of investment and the quantity and price of its output to maximize the firm's market value, defined as the present discounted value of cash-flows

$$e_{ij,t} = \max_{i_{ij,t},v_{ij,t}} d_{ij,t} + E_t[M_{t+1}e_{ij,t+1}],$$

where $e_{ij,t}$ is the value of firm i in industry j. In doing so, the firm understands that it has local monopoly power over the sale of its product (firms are monopolistic), and thus, firms will take the demand as given, as defined in Eq. (5).

The first order condition with respect to the variable input of production yields a factor demand relation that equates the marginal cost of adding one additional unit of variable input (left-hand side) to its marginal product (right-hand side):

$$P_t^v = \frac{\alpha^v}{\mu_{j,t}}\frac{p_{ij,t}y_{ij,t}}{v_{ij,t}}. \tag{10}$$

The first order condition with respect to investment yields the familiar Euler equation that equates the marginal cost of building an additional unit of capital (left-hand side) to the present value of the expected return on investment (right-hand side):

$$\frac{1}{\Gamma'_{ij,t}} = E_t\left[M_{t+1}\left(\frac{\alpha^k}{\mu_{ij,t+1}}\frac{y_{ij,t+1}}{k_{ij,t+1}} + \frac{1-\delta-\Gamma'_{ij,t+1}\frac{i_{ij,t+1}}{k_{ij,t+1}}+\Gamma_{ij,t+1}}{\Gamma'_{ij,t+1}}\right)\right], \tag{11}$$

where $\Gamma'_{ij,t} = \dfrac{d\Gamma(i_{ij,t}/k_{ij,t})}{d(i_{ij,t}/k_{ij,t})}$, and M_{t+1} is the one-period stochastic discount factor (to be defined later). Note that $q_{ij,t} = \Gamma'^{-1}_{ij,t}$ represents the shadow value of an installed unit of capital, i.e., the marginal q.

2.2 Markup shocks

We can reorganize the first order condition with respect to the variable input, i.e., Eq. (10), as follows:

$$\mu_{ij,t} = \frac{p_{ij,t}}{\frac{P^v_t v_{ij,t}}{\alpha^v y_{ij,t}}}. \tag{12}$$

This expression equates $\mu_{ij,t}$ to the wedge between the output price charged by the firm $p_{ij,t}$ and the real marginal cost of production $\dfrac{P^v_t v_{ij,t}}{\alpha^v y_{ij,t}}$. In other words, $\mu_{ij,t}$ is the price markup of firm i in industry j.

To understand the impact of aggregate markup risk on firms' stock returns, we introduce an exogenous aggregate markup process μ_t that impacts the price markup of all firms in the economy. We assume, however, that individual firms have heterogeneous exposures to this aggregate shock. In particular, we assume that firm-level markup obeys the following process in logs:[8]

$$\log(\mu_{ij,t}) = \log(\bar{\mu}_{ij}) + \lambda^{\mu}_{ij} \times \log(\mu_t), \tag{13}$$

where $\log(\mu_t)$ is an AR(1) process with iid innovations $\varepsilon^{\mu}_t \sim N(0, \sigma_\mu)$:

$$\log(\mu_t) = \rho_\mu \log(\mu_{t-1}) + \varepsilon^{\mu}_t. \tag{14}$$

[8]In the model, firm-level markups are the same for all firms within the industry because of our CES assumption. Thus, the i subscript can be dropped. In the empirical section, however, the exact definition of an industry is not clear so we directly estimate the markup at the firm-level. We add the firm subscript for generality.

Price markups differ across firms for two reasons. First, there is a time-invariant component that determines the steady-state level of market power of a firm $\bar{\mu}_{ij}$. For example, firms producing less differentiated products such as orange juice producers are likely to have lower markups than major high-tech firms such as Apple. Second, firms differ in exposure to the aggregate markup process μ_t. For example, a large firm with lots of data might be able to extra more rents from consumers using advances in artificial intelligence (a markup shock) than a smaller firm. In this case, large firms would have a higher fundamental markup exposure λ^{μ}_{ij} than smaller firms. Allowing for heterogeneity in both the level and sensitivity of markups allows us to compare these characteristics' impact on the cross-section of asset prices.

2.3 Representative household

We assume a representative household with Epstein-Zin preferences defined over aggregate consumption, C_t:

$$U_t = \frac{C_t^{1-1/\psi}}{1-1/\psi} + \beta \left(E_t[U_{t+1}^{1-\theta}] \right)^{\frac{1}{1-\theta}}, \tag{15}$$

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 time discount rate. In our calibration, we choose $\psi > \frac{1}{\gamma}$, so that the agent has a preference for early resolution of uncertainty following the long-run risks literature (e.g., Bansal and Yaron (2004)).

The representative household maximizes lifetime utility by choosing consumption and supplying the variable input to firms. We normalize the variable input endowment to one (e.g., labor as in Favilukis and Lin (2016)). The household also participates in financial markets by trading firms' equity. Accordingly, the budget constraint is:

$$C_t + P_t^S S_{t+1} \quad = \quad P_t^V V_t + (D_t + P_t^S) S_t, \tag{16}$$

where P_t^V is the competitive variable input price, V_t is the total variable input supplied by the household, normalized to one, D_t is the aggregate payout from the production sector that equals

$\iint_0^1 d_{ij,t} didj$, P_t^S is the price of the claim to aggregate dividend payout, and S_t is the stock holding in period t.

The household's maximization problem implies the stochastic discount factor (intertemporal marginal rate of substitution), which is used to price all assets in the economy:

$$M_{t+1} = \beta \left(\frac{U_{t+1}}{E_t(U_{t+1}^{1-\theta})^{\frac{1}{1-\theta}}} \right)^{-\theta} \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} \tag{17}$$

In equilibrium, all factor markets clear, that is, the variable input's demand equals the supply, the aggregate capital stock, and aggregate investment are the sum of firm-level capital and investment:

$$\begin{aligned} \iint_0^1 v_{ij,t} didj &= 1 \\ \iint_0^1 k_{ij,t} didj &= K_t \\ \iint_0^1 i_{ij,t} didj &= I_t \end{aligned}$$

where K_t and I_t denote the aggregate capital and the aggregate investment, respectively.

3 Quantitative results

This section evaluates the model quantitatively. We document that the price of markup risk is negative. Apart from successfully matching aggregate quantity and asset pricing dynamics, our heterogeneous firm setup enables us to investigate the relationship between market power, exposures to markup risk, and equilibrium stock returns, and to formulate new predictions that allow us to empirically test them in Section 4.[9]

[9]In the following, we omit the industry subscript j, unless it is necessary to avoid confusion.

3.1 Calibration and quantitative performance

Calibration We start with a description of the calibration used in the benchmark model. All parameters values are summarized in Table 1. The parameters that control household's risk aversion γ and intertemporal elasticity of substitution ψ are calibrated to 10 and 2 respectively, and are in the standard range of values in the long-run risk literature (e.g., Bansal and Yaron (2004)). The assumption that agent prefers a preference for early resolution of uncertainty ($\gamma > \frac{1}{\psi}$) is consistent with asset market evidence as shown in Ai et al. (2020). The time discount rate β is set to be 0.99 to generate a low average risk-free rate.

We set the capital share, α^k, and the variable input share, α^v, to 0.33 and 0.67, respectively. These are standard values used in the literature to match the steady-state input shares when the variable input is labor. The annual depreciation rate for physical capital is 10%, consistent with the steady-state investment rate. The demand elasticity across the industry, τ is chosen based on the estimates in Corhay, Kung, and Schmid (2020a). The investment adjustment cost parameters are set to match the steady-state investment rate and the volatility of the aggregate investment. The values are standard in the production-based asset pricing literature (e.g., Jermann (1998)).

The two aggregate shocks are calibrated to match their empirical counterparts. In particular the aggregate productivity growth process include three parameters: the persistence of the unobservable component ρ_x, the volatility σ_x, and the loading of productivity growth on aggregate investment ϕ_x. We choose the persistence parameter ρ_x to be consistent with the production-based asset pricing literature (e.g., Croce (2014)). The volatility parameter σ_x and the loading on the aggregate investment rate ϕ_x are pinned down by jointly matching the volatility of aggregate output and the unconditional Sharpe Ratio in the data. The parameters that govern the dynamics of the aggregate markup process are set to match our empirical measure described in section 4.1, that is, we set ρ_μ to generate an annual persistence of 0.93 and use a standard deviation for the markup shock of 1%.

We now discuss the parameters driving the cross-section of firms in the model, namely $\bar{\mu}_i$ and λ_i^μ. To keep the model tractable, we discretize the firm distribution in three groups along those two parameter dimensions. In other words, we have three level of average markups $\bar{\mu}_i$: low, medium, and high and for each level of markup, three values for the sensitivity parameter λ_i^μ, for a total of nine firm types. We choose the three values for each of the parameters to replicate the firm

Table 1: Calibrated parameter values

Description	Parameter	Value
Time discount rate	β	0.99
Relative risk aversion	γ	10
IES	ψ	2
Capital share in production	α^k	0.33
Variable input share in production	α^v	0.67
Adjustment cost curvature	Γ_1	15.5
Capital depreciation rate	δ_k	10%/4
Elasticity of substitution	τ	2
Long-run productivity growth	$\Delta\bar{z}$	2%/4
Persistence of productivity shock	ρ_x	$0.9^{1/4}$
Volatility of productivity shock	σ_x	0.24%
Loading of productivity on investment	ϕ_x	2.5%
Persistence of aggregate markup	ρ_μ	0.93
Volatility of markup shock	σ_μ	1%

This table presents the calibrated parameters used in the benchmark model.

distribution in our sample. In particular, the low-, medium-, and high-markup levels are set to 1.10, 1.27, and 1.47, respectively and the fundamental markup exposures λ_i^μ are set to -0.78, 0.86, and 3.86. Finally, the share parameters b_i in the CES aggregator (2) are calibrated so that the share of revenue for all types of firms is identical.

Quantitative performance We solve and simulate our model at a quarterly frequency using a second order perturbation method around the deterministic steady state. We then aggregate quantities and prices into annual level and report a series of key asset pricing and macroeconomic moments in Table 2. Our model is consistent with salient features of macroeconomic quantities and asset prices. In terms of aggregate moments for macro quantities, our model features a low volatility of consumption growth (1.64%) and a relatively high volatility of investment (3.66%). Both consumption and investment growth exhibit strong pro-cyclicality as in the data. For asset pricing moments, our model produces a low risk-free rate (2.34%) and a high equity premium 5.66% on the levered claim. The risk-free rate is smooth with a low volatility of 0.81%. In short, our model provides a good fit to both business cycle quantities and asset prices. We now turn to the impact of markups shocks in the model.

Table 2: Aggregate moments

	Data	Model
A. Macroeconomic Quantities		
$E[\Delta Y]$	2.05%	2.09%
$\sigma(\Delta Y)$	3.05%	1.72%
$\sigma(\Delta C)/\sigma(\Delta Y)$	0.83	0.95
$\sigma(\Delta I)/\sigma(\Delta Y)$	2.61	2.13
$\rho(\Delta C, \Delta Y)$	0.77	0.83
$\rho(\Delta I, \Delta Y)$	0.89	0.63
B. Asset Prices		
$E[R^M - R^f]$	5.71%	5.66%
$\sigma(R^M - R^f)$	17.4%	6.75%
$E[R^f]$	1.10%	2.34%
$\sigma(R^f)$	0.97%	0.81%

This table presents moments from the model simulations and the data, at an annual frequency. Panel A reports several key statistics for macroeconomic quantities, Panel B reports key asset pricing moments.

3.2 The price of markup shocks

In order to shed light on the implications of markup risk for asset prices, we plot the impulse response functions to a positive markup surprise. The results are reported in Figure 1. An increase in markup raises firms' market power and firms optimally cut production to maximize profits. This leads to a drop in intermediate input demand and capital expenditures, consistent with the firm's optimal investment condition (Eq. (11)). The persistent decline in output brings down aggregate consumption. From the investor's perspective, a decline in future consumption lowers her continuation utility, which raises her marginal utility. A positive markup shock is thus associated with negative news for the representative investor, that is, markup shocks carry a negative price of risk. As for the firm, the risk exposure to a markup shock is positive because the firm value and the corresponding excess return increase following the shock.

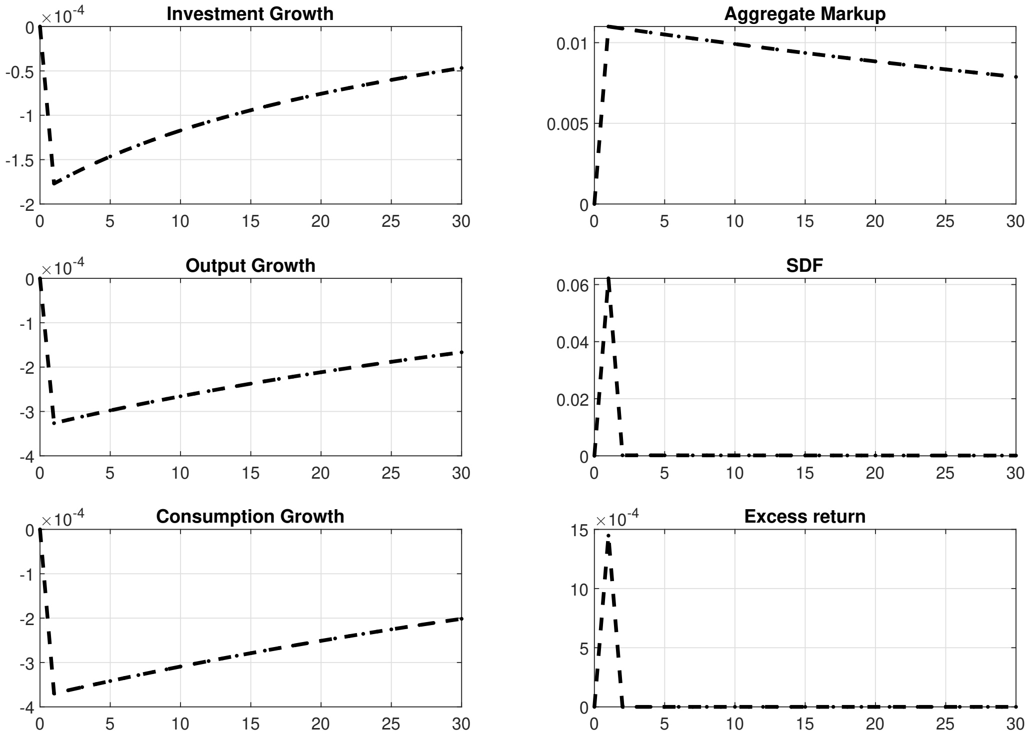


Figure 1: Impulse Response Functions for Markup Shocks

This figure plots the impulse-response functions to a positive markup shock (i.e., $\epsilon^{\mu} > 0$) for investment growth (ΔI), output growth (ΔY), consumption growth ΔC, aggregate markup μ is computed as the log of the sum of firm-level markups weighed by firm sizes, the stochastic discount factor (M), and the excess return on the aggregate stock market R^e. The units of the y-axis are percentage deviations from the steady state.

To better understand how a contractionary shock can cause an increase in asset prices, it is useful to decompose the market value of a firm, $e_{i,t}$ into two components: (i) the present value of rents, and (ii) the value of assets in place. Denoting the market value of firm i by $e_{i,t}$:

$$
\begin{aligned}
e_{i,t} &= d_{i,t} + E_t[M_{t+1}e_{i,t+1}] \\
&= p_{i,t}y_{i,t} - P_t^V v_{i,t} - i_{i,t} + E_t[M_{t+1}e_{i,t+1}] \\
&= d_{i,t}^{\text{rents}} + d_{i,t}^{\text{assets}} + E_t[M_{t+1}e_{i,t+1}] \\
&= e_{i,t}^{\text{rents}} + e_{i,t}^{\text{assets}}
\end{aligned}
$$

where $e_{i,t}^k = d_{i,t}^k + E_t[M_{t+1}e_{i,t+1}^k]$, for $k = \{\text{rents, assets}\}$ is the present value of the rents and assets

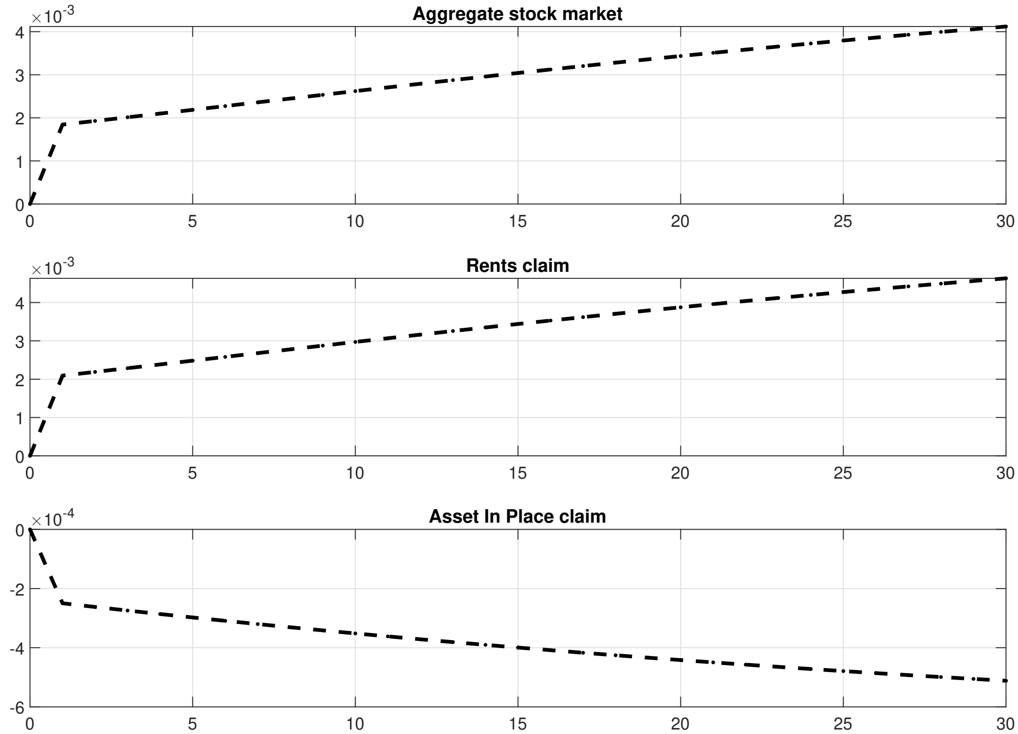


Figure 2: Impulse Response Functions - Firm Value Decomposition

This figure plots the impulse response functions to a positive markup shock (i.e., $\epsilon^{\mu} > 0$) for the present value of the rents claim, the present value of the assets in place claim, and the total stock market value. The present value of rents and asset in place are defined in equation (18) and (19), respectively. The units of the y-axis are quarterly percentage deviations from the steady state.

components of the firm. The dividends of the two components are defined as:

$$d_{i,t}^{\text{rents}} = p_{i,t} y_{i,t} - P_t^V v_{i,t} - R_{i,t}^k k_{i,t} \tag{18}$$

$$d_{i,t}^{\text{assets}} = R_{i,t}^k k_{i,t} - i_{i,t}, \tag{19}$$

where $R_{i,t}^k = \frac{\alpha^k}{\mu_{i,t+1}} \frac{y_{i,t+1}}{k_{i,t+1}}$ is the shadow rental rate of a marginal unit of capital. $d_{i,t}^{\text{rents}}$ denotes the cash-flow from monopolistic rents. When markets are perfectly competitive, this cash-flow is zero in each period causing the present value of rents to be equal to zero. The second dividend component, $d_{i,t}^{\text{assets}}$ represents the income the firms makes on its installed capital. Thus, $e_{i,t}^{\text{assets}}$ represents the value of assets in place.

Figure 2 examines how each of the firm value components changes in response to a positive markup shock. While the total market value increases in response to the shock, the individual components move in opposite direction. An increase in markup raises firms' market power and firms optimally cut production to maximize profits. The reduction in production is associated with a lower demand for capital which reduces the shadow rental rate of capital and the value of assets in place. In other words, when firms have a surprise increase in market power, installed capital is less valuable. In contrast, higher markup power increases the firms' monopolistic rents and thereby raises the present value of the rents component. Overall, the net impact of the markup shock depends on which of these two effects dominate. We find that the change in the rents component dominates, which explains why the aggregate stock return increases following the markup shock.

3.3 Cross-sectional asset pricing implications

We now turn to investigate the impact of heterogeneous exposures to the aggregate markup shock on asset prices. As evident from Equation (10) and (11), the price markup, $\mu_{i,t}$, plays a key role in determining the firms' optimal decisions and asset prices. To study these effects, we compare the impact of a markup shock on firms with different markup exposures. We consider two firms. The first firm, the "low markup exposure" firm, is negatively exposed to the aggregate markup shock. The second firm, the "high markup exposure" firm is positively exposed to the aggregate markup shock.[10] Results are plotted in Figure 3. In response to the positive markup shock, the high markup exposure firm sees its market power increase relatively more, which allows the firm to extract more monopolistic rents. The firm also cuts on investment. The reduction in investment, joint with the increase in monopolistic rents contribute to a rise in the dividend and firm valuation. In contrast, the responses of the low markup exposure firm are opposite.

The firm-level responses in Figure 3 highlight an important result: firm valuations respond very differently to markup shocks. Given that the price of markup risk is negative, a higher exposure to the aggregate process should translate into lower expected return. To assess the quantitative importance of this effect, we sort firms according to their markup sensitivity parameter λ_i^μ into low, medium, and high markup sensitivity portfolios. We then examine the characteristics of these

[10] We pick the markup sensitivity parameters of the low- and the high-markup exposure firm, i.e., λ_L^μ and λ_H^μ, to coincide with the 25% and 75% percentile of the empirical distribution of λ_i^μ's. More details on the sample construction for the λ_i^μ's can be found in section 4.

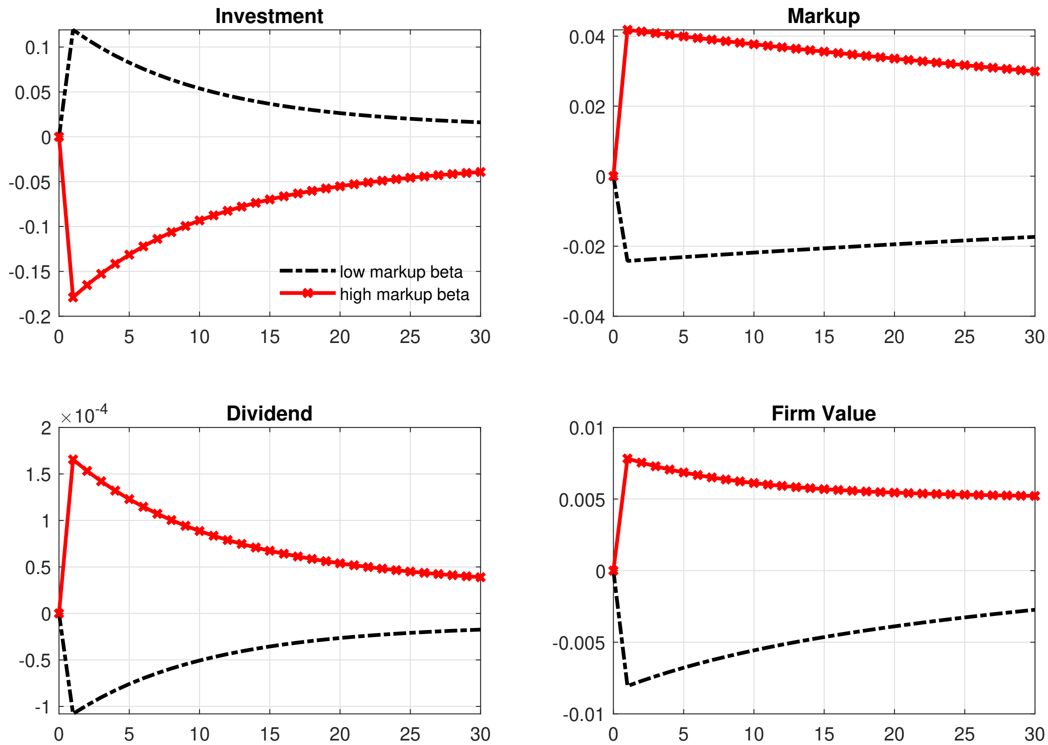


Figure 3: Impulse Response Functions - Markup exposure sorted firms

This figure compares the impulse response functions to a one-standard-deviation increase in markup shock (ϵ^{μ}), conditional on the sensitivity to aggregate markup parameter λ_i^{μ}, for firm investment growth (I_i), markup (μ_i), dividend growth (d_i), and firm value growth v_i. The high (low) case corresponds to a firm with λ_i^{μ} that equals the 75% (25%) percentile in the markup exposure distribution in our simulated data. All values on the y-axis are percentage deviation from the steady state.

three portfolios in Table 3. The upper panel shows that firms that are more exposed to markup risk have lower average excess returns. The return difference is economically significant. The high minus low (H-L) return spread is large, at around -1.2% per annum.

The negative H-L spread is not explained by a differential exposure to the aggregate productivity shock. To see this, we compute the risk exposures (betas) for each portfolio by running the following regression:

$$R_{it}^{e} = \alpha_i + \beta_i^{\mu} \varepsilon_t^{\mu} + \beta_i^{\Delta TFP} \epsilon_t^{x} + \epsilon_{it}, \tag{20}$$

where R_{it}^{e} is the excess return of firm i at time t.

The lower panel of Table 3 reports the risk exposure for each portfolio as well as that of the

H-L portfolio. The H-L portfolio loads positively on the markup shock, but its exposure to the productivity shock is essentially zero. Thus the negative H-L spread is solely explained by a positive exposure to markup risk.

Table 3: Portfolios sorted on exposures to markup shocks in the Model

	L	M	H	H-L
	Portfolio returns			
$E[R^e]$	6.86	6.32	5.68	-1.18
	Risk exposures			
$\beta_{\Delta TFP}$	0.82	0.91	0.84	0.02
β_{μ}	-0.32	0.39	0.80	1.12

This table reports the characteristics of value-weighted portfolios sorted on measures of firms' exposure to the markup shock, using model simulated data. For each measure, we report the lowest (L), medium (M), and highest (H) markup exposure portfolios. We also report the difference between the highest and lowest portfolios (H-L). We consider two markup exposure measures: the fundamental markup exposure λ_i^{μ} (left panel) and the return-based markup beta β_i^{μ} (right panel). Panel A reports the average excess return for each portfolio. Panel B reports the portfolio risk exposures to the TFP shock and markup shock. The risk exposures are estimated using linear regression: $R_{it}^e = \alpha_i + \beta_i^{\mu}\varepsilon_t^{\mu} + \beta_i^{\Delta TFP}\Delta TFP_t + \epsilon_{it}$. The rolling window for estimating exposure to shocks is 12 years as in the data. We generate a long sample with 4000 years of observations.

The asset pricing implications of firms' exposure to markup shocks contrast with those related to the average level of firms' markups. To see this, we run a series of Fama-MacBeth cross-sectional regressions of future returns on both the level of expected markups and the exposure to markup risk. Results are reported in Table 4. Column (1) shows that a higher expected markup is associated with higher future returns, consistent with the idea that the degree of market power increases firm risk (e.g., Corhay et al. (2020a)). Intuitively, the level of price markup endogenously increases the firm's exposure to aggregate productivity risk because monopolistic rents make firms' profits more pro-cyclical. In contrast, column (2) shows that higher markup exposure is associated with lower future returns, which echoes the results in Table 3. Importantly, column (3) shows that both effects remain significant and of similar magnitude when included together. These results suggest that both the level and the dynamics of markups are important to understand the impact of market power on asset prices and offers a very different risk mechanism. While the level of markups affect the exposure to a positively priced source of risk (productivity), the markup exposure mainly loads on a negatively priced source of risk (markup risk).

Table 4: Fama-MacBeth regression: Model-simulated data

	(1)	(4)	(5)
log(expected markup)	2.83***		2.82***
	(23.05)		(36.23)
β^{μ}		-0.65***	-0.57***
		(-8.60)	(-7.66)
Observations	35,100	35,100	35,100
R-squared	0.21	0.79	0.95
Controls	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes

This table reports the results of Fama-MacBeth regressions using the model-simulated data. We simulate the economy such that the sample is comparable to the empirical counterpart. We regress firm-level returns on lagged markup betas and controls. The markup shock exposure is obtained by regressing firm-level returns on the markup shocks, with the market factor as the control variable. It is estimated using a rolling window of 12 years. The markup exposure measures are standardized to have zero mean and a unity standard deviation. The excess returns are in annualized percentage terms. We report Newey-West adjusted t-statistics in parentheses, allowing for 4 lags. The t-statistics are in parenthesis, and *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively. We generate a sample consistent with the data sample in Table 10.

In short, the results in this section show that markup risk is an important source of economic fluctuations that is priced in financial markets. It also highlights an important difference between the level and exposure of markups in terms of asset pricing implications.

3.4 Testable predictions

Our model generates a series of new predictions on the asset pricing implications of markup shocks, which we summarize in this section. First, a surprise increase in markups depresses investment and production input demand and leads to a drop in output. At the same time, the aggregate stock market increases following a markup shock. This happens because higher market power increases the firms' present value of rents. This yields the following testable prediction:

Prediction 1. *A positive markup shock is associated with a contraction in macroeconomic aggregates, but an increase in aggregate stock market valuations.*

Accordingly, positive markups shocks are associated with bad news for the representative investor and thus high marginal utility states. This yields the second prediction:

Prediction 2. *The price of markup risk is negative.*

If the representative investors commend a negative premium for bearing markup risk, then firms with different exposures to markup risk should have different expected returns. This leads to our third prediction:

Prediction 3. *Firms who are more exposed to markup shocks earn lower average stock returns.*

Prediction 4. *Upon an increase in the aggregate markup shock, firms who are more exposed to markup shocks experience higher growth rates in markup, cash flow, and market capitalization. Their investments decrease more than firms with lower exposure to markup shocks.*

In the next section, we formally test those three model predictions using a series of empirical asset pricing tests.

4 Empirical evidence

In this section, we build an empirical measure for the aggregate markup shock. We then test the impact of markup shocks on asset prices and aggregate quantities. Importantly, we show that markup risk is priced in the cross-section of equity returns and carry a significant negative price of risk, consistent with the model predictions.

4.1 Construction of markup shocks

The aggregate markup shock series is built in three steps. First, firm-level markups are estimated following the production approach as in De Loecker, Eeckhout, and Unger (2020). Second, we remove the predictable component of firm-level markups to isolate markup surprises for each individual firm in the sample. Finally, the aggregate markup shock series is obtained by aggregating the firm-level markup surprises. We explain each step in more details in the sections below.

Data sources We first characterize the data used for the empirical analysis. Estimating firm-level markups requires accounting data. We obtain those from the Compustat Annual Database, for the period 1955 to 2020. Following standard practice in the literature, we exclude utility firms (SIC codes between 4900-4999) and financial firms (SIC codes between 6000-6999). We also drop observations with negative or missing sales item, cost of goods sold (COGS), or physical capital (PPEGT). Firm-level variables are deflated using the GDP deflator, obtained from the Bureau of Economic Analysis (BEA). Data on firms' stock return and market capitalization are obtained from CRSP database.

Macroeconomic variables are obtained from the Federal Reserve Economic Data Database maintained by the St. Louis Fed. Inflation is defined as the log change in the core Consumer Price Index. Real output is defined as the aggregate GDP divided by the GDP deflator. Consumption is defined as real personal consumption expenditures. Real investment is defined as the real gross private domestic investment. The relative price of investment goods to consumption goods ratio is calculated as investment deflator divided by consumption deflator. The aggregate series for Total Factor Productivity (TFP) is downloaded from the Federal Reserve Bank of San Francisco. We obtain the market factor and risk-free rate from Ken French's website and the returns of 25 portfolios formed on size and book-to-market, 10 portfolios formed on asset growth rates, and the Fama-French 10 industry portfolios. The 10 portfolios formed on return-on-asset are obtained from Hou, Xue, and Zhang (2015). The sample period is from 1963 to 2020.

Firm-level markups The first step consists of obtaining estimates of firm-level markups. Measuring markups at the firm-level is a difficult task as it requires obtaining data on marginal cost of production and output prices, which are not readily available. In this paper, we follow the production approach to estimate firm-level markup, pioneered by Hall (1988). This approach obtains firm-level markups by exploiting the cost-minimization of a variable input of production. To see this more clearly, one can rearrange the first order condition with respect to the variable input of production in the model, i.e., Equation (12), to obtain:[11]

$$\mu_{i,t} \;\; = \;\; \frac{\alpha_i^v}{s_{i,t}^v},$$

where $s_{i,t}^v = \frac{P_t^v v_{i,t}}{p_{i,t} y_{i,t}}$ is the share of variable input expenses as a proportion of total firm's revenues.

The expression above says that the price markup can be estimated as the wedge between a variable input's production elasticity and that specific input's expenditure share in revenue. Thus according to the production approach, the firm-level markup can be identified using *any* variable input, v, observed by the econometrician. In order to accurately measure firm markups in the data, one needs to first obtain data on a flexible production input. De Loecker, Eeckhout, and Unger (2020) use the Compustat item COGS as a measure of v. The idea is that COGS should include all relevant variable costs attributable to the sale of a unit of output, such as the cost of material, electricity, labor, etc. However, Traina (2018) argues that simply using COGS could underestimate the true cost structure of a firm's variable input over time. This happens because firms have increasingly devoted more of their inputs toward marketing and management costs, which are not recorded under COGS, but under the Compustat item SG&A (selling, general and administrative expenses). Therefore in this paper, we define the total variable input of the firm by adding 70% of SG&A expenses to the reported COGS.[12]

An important variable to compute markups is the elasticity parameter α_i^v. We closely follow the procedure in De Loecker et al. (2020) and estimate α_i^v using the standard production function estimation techniques. The firm production function is assumed to be Cobb-Douglas as in the model (see Eq. 6), with a variable input bundle V and a predetermined capital stock K (defined by Compustat item PPEGT). Following the literature, the elasticities are estimated by pooling firms at the two-digit NAICS level. This assumption is reasonable as firms in the same industries are likely to have similar production technologies. In addition, it allows to have more observations to more precisely estimate the elasticity parameters.

A well-known challenge is that the unobserved firm productivity may correlate with the produc-

[11] We ignore the industry subscript for simplicity.

[12] The choice to add 70% of SG&A is consistent with the literature on intangible capital that argues that roughly 30% of SG&A can be capitalized as organizational capital. Gutiérrez and Philippon (2017) also do a similar adjustment to obtain their measure of total variable costs.

tion inputs, leading to biased estimates of the production elasticities. We follow standard practice in the literature (e.g., see Levinsohn and Petrin (2003)) and use a control function approach with a law of motion for unobserved productivity to obtain a consistent estimate of the output elasticities. The detailed estimation procedure can be found in Appendix A. At the end of the first step, we obtain a panel of around 150,000 firm-year observations of price markups.[13] We report the summary statistics for our sample in Table 5.

Table 5: Summary statistics of firm-level markup measures

	mean	median	std	Observations
μ	1.18	1.17	0.28	157,273

This table reports the summary statistics of the estimated firm-level markup measures. The sample is from 1962 to 2020. The detailed estimation procedure can be found in Appendix A.

Firm-level expected markups and markup surprises The second step consists of purging the expected component in firms' markups in order to estimate surprise changes in markups. We compute the expected markup for each firm i, $E_{t-1}[\mu_{i,t}]$ using a predictive regression approach. In particular, we run the following predictive regression at the 4-digit NAICS code industry level:

$$\mu_{i,t} = \eta_i + \rho\mu_{i,t-1} + \beta Z_{i,t-1} + \varepsilon^{\mu}_{i,t}, \tag{21}$$

where η_i denotes firm fixed effect, $Z_{i,t-1}$ is a vector of lagged control variables, and ε_{it} is the markup surprise. The control variables include size, cash flow to asset ratio, book-to-market ratio, and productivity. The productivity is obtained from the estimating firm-level markup, please see details in Appendix A. The choice of control variables is motivated by the empirical findings in De Loecker et al. (2020) that firms' markups are closely related to size and profitability.

Table 6 reports the results of the predictive regression performed on the whole sample and shows that it does a reasonable job at predicting firms' subsequent markups. The high R^2, around

[13] To ensure that we capture well the common components in price markups, we require at least 100 observations for each year. This is a problem for the early sample where data is often missing. Thus, our markup measures starts in 1962.

Table 6: Expected Markup

Predictor	μ	$\log(ME)$	CF/AT	BM	TFP	R^2
Coefficients	0.410	0.002	0.040	0.003	0.011	0.801
	(19.48)	(1.14)	(2.95)	(3.71)	(1.01)	

This table reports the predictive regression as in Equation (21). The lagged control variables include markup μ, the logarithm of market capitalization $\log(ME)$, cash flow to asset ratio CF/AT, and book-to-market ratio BM. We also include firm fixed effect. The sample is from 1962 to 2020. We report t-statistics in parentheses. Standard errors are clustered by firm and year.

80% on average, demonstrates that our choice of predicting variables can explain a large fraction of variations in markups.[14]

We interpret the difference between the firm-level realized and expected markup as the surprise markup shock. More formally, the firm-level markup shock series is obtained as follows:

$$\varepsilon^{\mu}_{i,t} = \mu_{i,t} - \widehat{E_{t-1}[\mu_{i,t}]}. \tag{22}$$

Aggregate markups and markup shocks We are ultimately interested in obtaining a series for the markup shock that affects all firms in the aggregate. Thus, our last step consist of aggregating the markups and markup shocks series.

Following Edmond, Midrigan, and Xu (2018), we define the aggregate markup series by cost-weighting the firm-level observations. In particular, the aggregate and expected markup series are obtained as follows:

$$\mu_t = \sum_i w_{i,t}\mu_{i,t} \tag{23}$$

$$E_{t-1}[\mu_t] = \sum_i w_{i,t}E_{t-1}[\mu_{i,t}], \tag{24}$$

where $w_{i,t}$ is the total variable input of firm i, divided by the total variable input for all firms at time t.

The upper panel in Figure 4 presents the evolution of the aggregate and expected markup

[14]The lagged markup alone can explain more than 60% in an OLS regression without fixed effects.

over time. Consistent with the findings in De Loecker et al. (2020), aggregate markups started to rise in the 1980s. Our average level of markup is lower, on average, which is consistent with the results documented in Traina (2018). The expected markup series follows a very similar pattern as the realized markup, which suggests that our predictive model captures a lot of the variations in markups.

Next, the aggregate markup shock measure is obtained by cost-weighting the markup surprises at the firm level:

$$\varepsilon_t^{\mu} = \sum_i w_{i,t} \varepsilon_{i,t}^{\mu}. \tag{25}$$

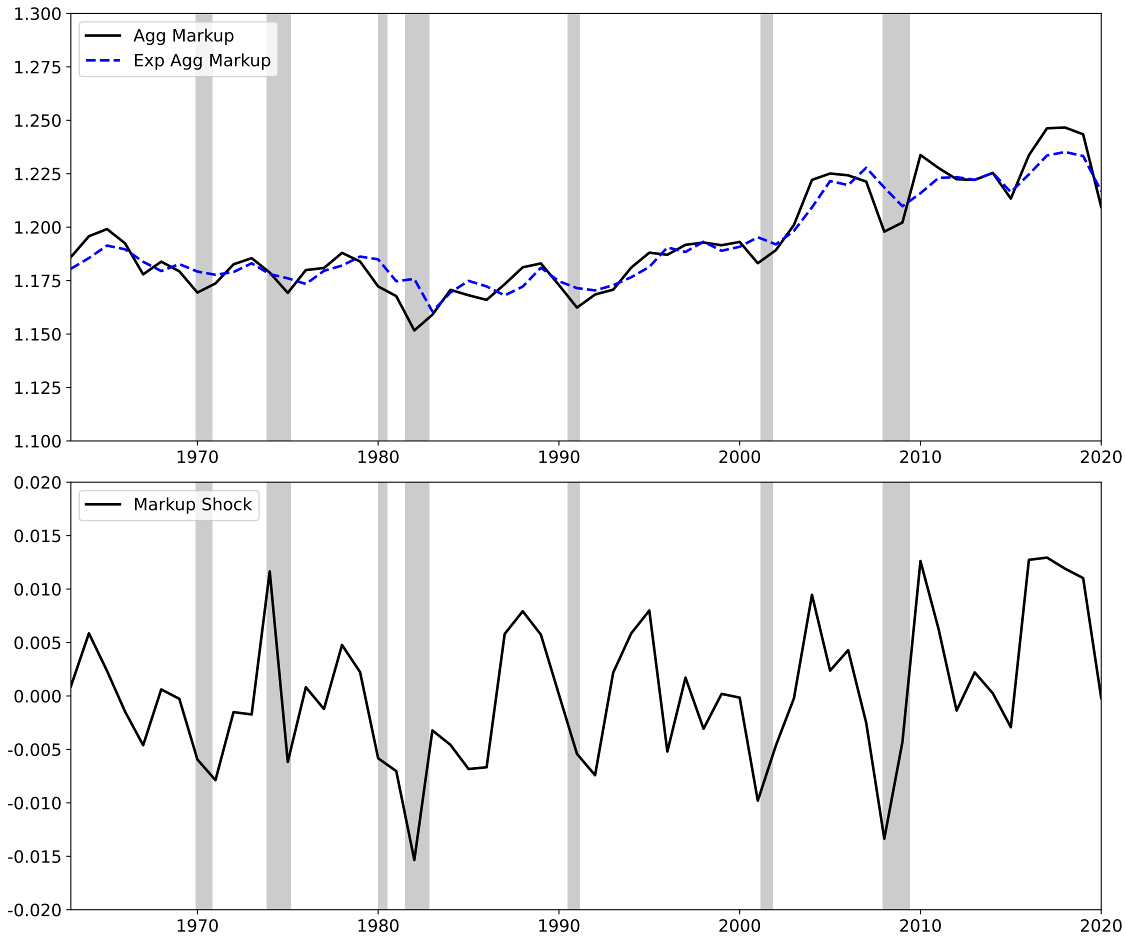
To ensure that the aggregate markup shocks is orthogonal to standard productivity shocks, we project the markup shock series on the contemporaneous changes in TFP and keep the residual as our measured aggregate markup surprises. The volatility of the resulting aggregate markup shocks series is 0.9% per annum, and the persistence is 0.16.

The lower panel of Figure 4 plots the times series of markup shocks. Several points are worth noticing. First, markups shocks tend to be high before economic recessions. This is true, even after controlling for the business cycle variations in TFP. This suggests that markup shocks might have important effects on macroeconomic aggregates as predicted by the model. We formally study the business cycle properties of markups shock series in the next section. Second, the markup shock series has experienced several large positive shocks in the later part of the sample. This suggests that some of the forces driving the secular trends in markups are unrelated to changing firm's characteristics or total factor productivity, but driven by another process affecting all firms' markups in the cross-section.

4.2 Markup shocks and the business cycles

Our paper argues that markup risk is an important driver of asset prices. In order to ensure that our constructed series does indeed capture variations in aggregate markups, it is useful to compare the business cycle fluctuations to those inferred from the model (see Figure 1). An increase in markup is associated with a decrease in investment, labor and output. We test these relations in

Figure 4: Aggregate Markup and Markup Shocks



This figure presents the time series of the realized and expected aggregate markup (top panel) and the time series of the aggregate markup shocks (bottom panel). Shaded bars represent NBER recession years. Data are annual from 1963 to 2020.

the data by running the following predictive regressions:

$$\Delta x_{t\rightarrow t+h} = \alpha + \beta_e \varepsilon_t^{\mu} + \gamma \Gamma_t + \varepsilon_{t+h}, \tag{26}$$

where $\Delta x_{t\rightarrow t+h} = \sum_{s=1}^{h} \Delta x_{t+s}$ denotes the h-year ahead growth rate of the variable of interest, Γ_t is a vector of lagged macroeconomic variables used as controls.

We standardize the markup shock to help interpret the results. As shown in Panel A, a one-standard-deviation increase in the markup shock is associated with a 0.5% to 2% standard deviation decrease in consumption growth rates in the next one to two years. The coefficients are large and

Table 7: Markup shocks and aggregate variables

Panel A: Predicting macroeconomic quantities

	$\Delta c_{t\to t+1}$	$\Delta c_{t\to t+2}$	$\Delta c_{t\to t+3}$	$\Delta y_{t\to t+1}$	$\Delta y_{t\to t+2}$	$\Delta y_{t\to t+3}$
β_e	-0.005**	-0.013***	-0.019***	-0.002	-0.007***	-0.014***
(t)	(-2.52)	(-4.18)	(-3.57)	(-0.89)	(-3.15)	(-3.64)
R^2	0.412	0.182	0.165	0.396	0.172	0.164

	$\Delta i_{t\to t+1}$	$\Delta i_{t\to t+2}$	$\Delta i_{t\to t+3}$	$\pi_{t\to t+1}$	$\pi_{t\to t+2}$	$\pi_{t\to t+3}$
β_e	-0.002	-0.017	-0.036***	0.004**	0.009***	0.012**
(t)	(-0.16)	(-1.54)	(-2.68)	(2.55)	(2.70)	(2.42)
R^2	0.320	0.220	0.219	0.864	0.665	0.643

Panel B: Correlation with market valuations

	ΔQ_t	R_t^M
β_e	0.028*	0.41*
(t)	(1.92)	(1.76)

Panel A reports the result of predictive regressions between markup shocks and business cycle variables. Specifically, it reports slope parameter β_e of markup shocks in Equation (26). Panel B presents the contemporaneous correlations between markup shocks and measures of market valuations, using the same set of control variables. We use the growth rate of aggregate Tobin's q (ΔQ) and market excess returns (R^M) to measure market valuations. ΔQ_t is defined as the growth rates of the aggregate market capitalization to private nonresidential fixed capital ratio. The controls include lagged consumption growth Δc_{t-1}, lagged output growth Δy_{t-1}, lagged investment growth Δi_{t-1}, and lagged core inflation π_{t-1}. The markup shock is normalized to have a standard deviation of unity. The Newey-West adjusted t-statistics are reported in parentheses, computed using six lags. We use *, **, and *** to indicate significance at the 10%, 5%, and 1%, respectively. The sample is from 1963 to 2020.

statistically significant. The responses of output and investment growth have similar economic magnitudes. Therefore, a positive shock to markups is associated with a contraction in real quantities, consistent with the model. Importantly, the response of inflation to an increase in markups is positive and statistically significant, which suggests that the forcing process causing the recession also leads to inflationary pressure.[15] Panel B shows the relation between markup shocks and valuation ratios. A positive markup shock is associated with an increase in the aggregate Tobin's

[15]Note that the results are robust to using other measures of inflation, such as total CPI or the GDP deflator.

Q, which corroborates one key prediction from the model: markup shocks are contractionary but lead to an increase in firms' valuations. The reason why an increase in markups lead an opposite effect in aggregate quantities vs. asset markets comes from the fact that higher markups increase the present value of future rents, which boosts firm valuations at the expenses of households (see Figure 1).

Overall, the findings in this section show that the business cycle properties of the markup shocks are in line with those produced by an exogenous markup shock. In the next section, we study the asset pricing properties of markup shocks.

4.3 The price of markup risk

In this section, we formally test the second prediction of the model concerning the impact of markup shocks on asset prices, namely, the price of markup risk is negative. To do so, we perform a standard linear factor asset pricing test (e.g., see Cochrane (2009)). The model features two sources of aggregate risk, namely aggregate productivity shocks and markup shocks. Accordingly, in our empirical tests, we consider a two-factor asset pricing model with the stock market excess return as a proxy for aggregate productivity risk as in Belo, Lin, and Yang (2019) and the markup shock as the second factor. We specify the following stochastic discount factor (SDF):

$$M_t = 1 - b^{MKT} \cdot \mathrm{MKT}_t - b^{\mu} \cdot \varepsilon_t^{\mu}. \tag{27}$$

The equation above states that investors' marginal utility is driven by two aggregate shocks, the market factor MKT_t and the markup shock ε_t^{μ}. The parameters b^{MKT} and b^{μ} denote the loadings on the market factor and markup risk, respectively.

Our goal is to estimate the prices of risk, b^{MKT} and b^{μ}, which is possible using a cross-section of traded assets. In particular, imposing the standard asset pricing moment restrictions, we obtain the following equilibrium condition, which must hold in the absence of arbitrage:

$$E[\big(1 - b^{MKT} \cdot \mathrm{MKT}_t - b^{\mu} \cdot \varepsilon_t^{\mu}\big) R_{it}^{e}] = 0, \tag{28}$$

where $R^e_{i,t}$ is the excess return on a traded asset i in year t. The factor loadings on the two aggregate shocks (b^{MKT} and b_μ) are estimated by the generalized methods of moments (GMM). We also consider the special case of a single-factor model with only the market factor, i.e., the CAPM, to compare it with the full specification in (28). This will help assess the relative importance of markup shocks to capture systematic risk in the economy.

We estimate Eq. (28) using a broad set of testing portfolios, namely, 25 portfolios sorted on size and book-to-market ratio, 10 capital growth (ΔK) portfolios, 10 return-on-asset (ROA) portfolios, and 10 industry portfolios. This large set of portfolios – 55 in total – allows us to improve on the precision of the risk loadings estimates and establish the importance of markup shocks for asset prices. Table 8 reports the GMM estimates of the risk factor loadings and implied mean absolute pricing errors (MAE), obtained from the estimation of the two asset pricing models, using different sets of test assets. The table shows that the estimated factor loadings on the markup shock are negative and significantly different from zero across all the different sets of test assets. In contrast, the factor loading on the market factor is positive. The economic magnitude of the estimated factor loading is large at around -1.5 and quantitatively similar across specifications but somewhat reduces when industry portfolios are included. This suggests that markup shocks affect firms at a more granular level than industries. One can also convert the factor loadings to price of risks estimates.[16] The corresponding market prices of risks for the market factor and the markup shock are around 10 and -1.5, respectively.

Table 8 also reports the MAE for each asset pricing model. The two-factor model does a better job explaining the returns than the CAPM model as the MAE is about 50% lower than those of the CAPM model. For robustness, we consider a specification that includes investment-specific technology (IST) shocks (Papanikolaou (2011)) as a third risk factor in our linear SDF test. The IST shock is defined as the growth rate of the relative price of investment goods to consumption goods. The results in Table 8 show that after controlling for IST shocks, the negative price of markup risk is still statistically significant. Consistent with Papanikolaou (2011), the slope coefficient b_{IST} is negative, which suggests that investment-specific shock carries a negative price of risk.

Taken together, the results in Table 8 show that markup risk is an important source of aggregate

[16]See Cochrane (2009) for details.

Table 8: Price of risk of markup shocks

	25 SZ-BM		10 ΔK		10 ROA	
b_{MKT}	0.54	0.58	0.43	0.74	0.36	0.64
(t)	(3.34)	(1.98)	(2.76)	(1.87)	(2.30)	(1.03)
b_{μ}		-1.17		-1.72		-1.92
(t)		(-2.07)		(-1.71)		(-1.70)
MAE	2.39	1.98	1.44	1.06	2.23	1.40

	45 SZ-BM, ΔK, ROA		55 SZ-BM, ΔK, ROA, Industry		55 SZ-BM, ΔK, ROA, Industry	
b_{MKT}	0.45	0.60	0.46	0.53	0.46	0.61
(t)	(3.18)	(1.59)	(3.15)	(2.16)	(3.15)	1.54
b_{μ}		-1.41		-0.74		-1.47
(t)		(-2.10)		(-2.59)		(-2.22)
b_{IS}						-0.04
(t)						(-0.10)
MAE	2.30	1.72	2.12	1.94	2.12	1.81

This table shows the estimation results of the linear stochastic discount factor, as in Equation (27). We perform the first-stage estimation using the identity matrix as the weighting matrix. We report from the estimates of the risk factor loadings and the pricing errors. The *t*-statistics are reported in parentheses, computed using the Newey-West estimator allowing for six lags. MAE is the mean absolute pricing errors. The risk factors considered are the market factor and the markup shocks. We also consider a specification with an additional factor, the investment-specific technological shocks (IST). We normalize both of the factors to have a mean zero and a standard deviation of one. The estimation of the asset pricing model is based on 55 portfolios: 25 portfolios sorted on size and book-to-market ratio, 10 capital growth (ΔK), 10 return-on-asset (ROA), and 10 industry portfolios. Data are annual from 1963 to 2020.

risk that is priced in the cross-section of equity returns. Importantly and consistent with the model predictions, the price of markup risk is large and negative, that is, episodes of unexpectedly high market power corresponds to high marginal utility states for investors.

4.4 Markup risk and the cross-section of asset prices

The previous section documented that markup shocks are an important source of systematic risk. As a result, heterogeneous exposures to the aggregate markup shock should result in cross-sectional differences in stock returns. In this section, we provide a set of empirical tests that support the third prediction of the model.

Markup betas and the cross-section of stock returns In the model, the aggregate markup shock affects firms heterogeneously because firm-level markups have different exposure to the aggregate markup process. In this section, we provide further evidence for the key economic channel through which exposures to markup shocks affect the asset prices in the cross-section.

We use stock returns to estimate firms' exposure to aggregate markup shock $\beta^{\mu}_{i,t}$. It is obtained by regressing firm-level excess returns on the markup shock, controlling for the market factor. It is calculated using a 12 year-rolling window. We then sort firms into ten portfolios based on their level of markup exposures and track their subsequent returns.

Table 9 presents the results. The upper panel shows that firms with higher markup exposures earn, on average, 6.18% lower returns than firms with lower markup exposures. The return differential is large and statistically significant at the usual confidence level. Importantly, as shown in the lower panel, while the estimated markup betas exhibit an increasing pattern, the market and IST betas across the different portfolios are mostly flat. Those results are consistent with the model's predictions that markup shocks demand a negative price of risk and that the main force driving cross-sectional differences across markup exposure-sorted portfolios is due to heterogeneous exposure to markup risk (see Table 3 for the model counterpart table). These results corroborates the third prediction of the model, namely, heterogeneous exposures to markup risk is an important driver of the cross-section of asset prices.

We further test the third prediction of the model via Fama-MacBeth cross-sectional regressions of realized stock returns on lagged markup betas. We also use standard firm characteristics that have proven successful at explaining the cross-section of stock returns as control variables, such as the book-to-market, size, ROA, asset growth ΔK, and financial leverage. To ensure that our results are not driven by the average level of markups, we also include the expected markup as a control variable. The regression specification is:

$$R^e_{i,t+1} = a_j + b\log(E_{t-1}[\mu_{i,t}]) + c\beta^{\mu}_{i,t} + \gamma\Gamma_{i,t} + \epsilon_{i,t}, \tag{29}$$

where $\beta^{\mu}_{i,t}$ is the measure of markup exposure, a_j is the industry fixed effect, and $\Gamma_{i,t}$ is a vector of

Table 9: Portfolios sorted on exposures to markup shocks

	L	3	6	8	H	H-L
	Portfolio returns					
$E[R^e]$	13.11	9.98	9.41	8.14	6.94	-6.18
(t)	(5.30)	(4.71)	(4.01)	(6.19)	(3.31)	(-2.75)
	Risk exposures					
α	4.06	1.24	9.88	1.97	-4.82	-8.88
(t)	(0.64)	(0.27)	(5.72)	(1.29)	(-1.22)	(-0.91)
β_μ	-2.53	-2.20	1.11	1.36	2.79	5.32
(t)	(-1.99)	(-1.56)	(0.91)	(1.65)	(1.60)	(1.93)
β_{MKT}	1.07	0.85	0.78	0.72	1.00	-0.07
(t)	(11.73)	(13.08)	(8.50)	(9.79)	(5.72)	(-0.40)
β_{IST}	0.03	-0.30	3.00	0.25	-0.87	-0.89
(t)	(0.01)	(-0.17)	(3.92)	(0.61)	(-0.65)	(-0.29)

This table reports the characteristics of value-weighted portfolios sorted on measures of firms' exposure to the markup shock. We sort firms into 10 portfolios each based on the markup shock exposure measure. We report the characteristics for selected deciles, as well as the difference between the first and the tenth decile portfolios (H-L). Panel A reports the average excess return for each portfolio. Panel B estimates the portfolio risk exposures to the markup shock β_i^μ, the market factor β_i^{MKT}, and the investment-specific technological shock β_i^{IST}. The risk exposures are obtained by the following regression: $R_{it}^e = \alpha_i + \beta_i^\mu \varepsilon_t^\mu + \beta^{MKT}\text{MKT}_t + \beta^{IST}\text{IST}_t + \epsilon_{it}$. The Newey-West adjusted t-statistics with 10 lags are reported in parentheses. The sample period is 1980-2020.

controls. To better interpret our results, $\beta_{i,t}^\mu$ is standardized to have a zero mean and a standard deviation of one. The excess returns $R_{i,t+1}^e$ are annualized.

Table 10 presents the results from the Fama-MacBeth regressions for both the markup beta and the return-based markup beta. The results in column (1) show that the expected markup positively predicts future returns. This firm-level evidence is consistent with our model findings and the existing literature that firms with higher markups have a higher risk. The results in column (2) show that a one-standard-deviation increase in markup exposure is associated with a 1% drop in returns next year. It is economically and statistically significant. This negative relationship between exposure to markup shocks and future returns is robust when we control for the expected markup, as shown in column (3). Additionally, we obtain quantitatively similar results if we use return-based markup beta instead of the markup beta, as presented in column (4) and (5). Overall,

Table 10: Fama MacBeth predictive regressions

	(1)	(2)	(3)
log(expected markup)	2.42		2.80
	(1.30)		(1.25)
β^{μ}		-0.94*	-0.89*
		(-1.88)	(-1.71)
Observations	93,114	85,599	66,509
R-squared	0.05	0.05	0.06
Controls	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes

This table reports the results of Fama-MacBeth regressions based on the specification in Equation (29). The markup beta β^{μ} is obtained by regressing firm-level markup surprises on the markup shocks, with the market factor as the control variable. It is estimated using a rolling window of 12 years. We standardized β^{μ} to have zero mean and a unity standard deviation. The control variables include the book-to-market ratio, size, return on equity, capital growth rate, and debt-to-asset ratio. We also control for industry-fixed effects using the two-digit NAICS code. The excess returns are in annualized percentage terms. We report Newey-West adjusted t-statistics in parentheses, allowing for 10 lags. The sample period is 1980-2020. The t-statistics are in parenthesis, and *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively.

the cross-sectional tests confirm the model predictions on the relation between the level of markup, the exposure to markup risk, and firms' expected return presented in Table 4.

The documented results are consistent with earlier studies that have shown that higher market power is associated with higher expected stock returns. For example, Bustamante and Donangelo (2017) and Corhay et al. (2020a) show that firms operating in more concentrated industries face a higher threat from new entrants, resulting in more risk. Corhay (2017) shows that the embedded put option to default combined with a time-varying competition hedge reduces the cyclicality of competitive firms' cash flow, thereby reducing risk. While these papers rely on productivity risk to explain the relation between markup, industry concentration and returns, our paper identifies a new and complementary force through which markups impact firms' risk. In particular, we find that a larger exposure to markup shocks reduces firms' risk.

To sum up, using the portfolio sorts and Fama-MacBeth regressions, we provide both portfolio and firm-level empirical evidence on the negative relationship between exposures to markup shocks and future returns.

Risk exposures of testing portfolios The previous section showed that heterogeneous exposures to markup shocks is priced in the cross-section of stock returns. In this section, we investigate the relation between markup betas and asset prices for a series of portfolios sorted on various firm-level characteristics that are standard in the empirical finance literature, namely portfolio sorted on book-to-market, size, asset growth, and profitability. For each portfolio sort, we report the average return as well as the risk exposures to the markup shock estimated using

$$R^e_{it} = \alpha_i + \beta_i^{MKT}\mathrm{MKT}_t + \beta_i^{\mu}\varepsilon_t^{\mu} + \varepsilon_{it} \tag{30}$$

where α_i is a constant term, and β_i^{MKT} and β_i^{μ} denote the exposures to the market and markup factors, respectively. The hope is that markup betas explain, at least partially, the well-known "anomalies".

Table 11 reports the results. The estimated markup betas vary significantly within each portfolio sort, usually in a monotonic fashion. Perhaps more interestingly, the markup betas for each portfolio are well-aligned with their respective portfolio returns. For instance, growth firms underperform value firms in terms of equity returns, commonly referred to as the value premium. We find that growth firms are more positively exposed to markup shocks, leading to lower expected returns. One potential reason for this is that growth firms tend to invest more in R&D and organizational capital. De Loecker et al. (2020) show that in the data, firms with high markups tend to also invest more in R&D and advertising expenditures. Therefore, the growth characteristic is likely correlated with the present value of future rents of the firm, which might cause growth firms to be more exposed to shocks that affect current and future markups. Because the aggregate markup shock carries a negative price of risk, growth firms will earn, on average, a lower stock return than value firms.

The markup beta also lines up well with the portfolios sorted on firm size (ME). Larger firms have lower returns and larger markup betas. This is consistent with the argument that larger firms may benefit more from new data technologies, which increases the market power of such firms at the expenses of the smaller firms (Begenau et al. (2018)). In fact, we find that small firms have a negative exposure to the markup shock as opposed to larger firms, suggesting that markup shocks

Table 11: Risk exposure to markup shocks

	L	2	3	4	5	6	7	8	9	H	H-L
						BM					
$E[R]$	7.14	8.43	7.70	7.38	7.60	8.84	8.07	9.63	11.16	11.63	4.48
(t)	(2.55)	(3.43)	(4.26)	(3.92)	(4.26)	(6.23)	(5.06)	(6.61)	(7.97)	(6.23)	(1.36)
β_μ	0.14	-0.11	-0.31	-0.32	1.13	1.14	-0.31	-0.67	-1.53	-2.41	-2.54
(t)	(0.15)	(-0.14)	(-0.81)	(-0.48)	(1.89)	(1.10)	(-0.24)	(-0.72)	(-2.22)	(-2.62)	(-1.63)
						ME					
$E[R]$	11.34	10.09	10.45	9.71	10.31	9.35	9.60	9.05	8.27	6.86	-4.49
(t)	(4.90)	(5.67)	(6.65)	(6.44)	(6.79)	(5.84)	(6.21)	(5.43)	(4.44)	(3.08)	(-1.36)
β_μ	-3.61	-2.98	-1.46	-2.13	-1.63	-1.20	-0.81	-0.37	0.44	0.29	3.90
(t)	(-1.82)	(-2.36)	(-1.94)	(-2.98)	(-2.30)	(-2.47)	(-1.68)	(-0.75)	(0.99)	(1.14)	(1.87)
						ΔK					
$E[R]$	9.28	7.64	7.89	7.31	6.51	7.14	6.57	7.32	8.16	5.61	-3.67
(t)	(5.48)	(5.63)	(6.02)	(4.45)	(3.28)	(3.87)	(3.21)	(3.40)	(2.98)	(1.86)	(-1.41)
β_μ	-0.50	0.09	0.23	1.21	0.08	-0.01	-0.06	0.23	0.68	1.73	2.23
(t)	(-0.52)	(0.11)	(0.38)	(2.42)	(0.10)	(-0.02)	(-0.08)	(0.35)	(0.65)	(1.82)	(1.86)
						ROA					
$E[R]$	3.55	6.09	6.57	7.55	6.83	8.53	8.39	8.32	8.57	9.54	5.99
(t)	(1.14)	(3.24)	(3.13)	(3.96)	(3.31)	(3.78)	(4.74)	(4.70)	(3.99)	(3.46)	(2.91)
β_μ	2.37	-0.72	0.05	-0.45	-0.19	0.52	-0.67	0.02	0.28	-0.98	-3.35
(t)	(1.48)	(-0.68)	(0.05)	(-0.43)	(-0.31)	(1.27)	(-1.37)	(0.02)	(0.45)	(-1.27)	(-2.56)

This table reports the testing portfolios' risk exposures to markup shocks. We report the estimates of the risk exposure to markup shocks β_i^μ as defined in Equation (30). The Newey-West adjusted t-statistics are reported in parentheses, allowing for 10 lags.

may have important impacts on the distribution of firms.

Similarly, firms that have experienced a high growth of capital (ΔK) are more exposed to the aggregate markup shock, which explains their lower equity returns observed in the data. A potential explanation for this finding is as follows. Firms that grow their assets tend to do so to take advantage of profitable growth opportunities. Those profit opportunities contain a rent component, which is sensitive to surprise changes to the current and future market power of the firm. Thus firms with a higher growth rate of assets have a higher equilibrium markup beta, which explains their lower risk in equilibrium.

Overall, the exposure to markup shocks can explain a wide range of portfolio sorts documented in the literature. Although the model is purposely silent on the economic mechanism linking firms' characteristics to beta exposures, the results in this section provides an interesting avenue for future studies aiming to provide a micro-founded explanations for the documented markup beta spreads.

4.5 Markup exposures and firm dynamics

This section provides empirical evidence that firms with different markup shock exposures will react to aggregate markup differently. These results support the fourth prediction of the model.

To study the role of markup shock exposure, we run the following regression:

$$\Delta y_{it} = a_i + c_t + b \cdot \varepsilon_t^{\mu} \times \beta_{it}^{\mu} + \gamma \beta_{it}^{\mu} + \Gamma Z_{it} + \varepsilon_{it}, \tag{31}$$

where a_i is the firm fixed effect, c_t is the year fixed effect, and ε_t^{μ} is the aggregate markup shock. The markup shock exposure β^{μ} and markup shock ε^{μ} are standardized. Firm-level controls include size $\log(ME)$, book-to-market ratio, book leverage, markup, and return on asset (ROA).[17] The dependent variable $\Delta y_{i,t}$ include the growth rate of markup ($\Delta\mu_{i,t}$), cash flow ($\Delta cf_{i,t}$), market capitalization ($\Delta me_{i,t}$), and physical capital (Δk). The coefficient of the interaction term b captures the impact of aggregate markup shock conditional on the firms' markup exposures.

Table 12 reports the results from estimating the Equation (31). The estimation results show that firms with higher exposure to aggregate markup shocks experience higher growth rates in markup, cash flow, and market capitalization. Their investment will be lower than firms with lower markup shock exposures. In the meanwhile, their investment will fall more. All these results are consistent with our model's prediction four that firms with higher markup shock exposure can extract more monopolistic profits upon a positive markup shock. These results corroborate our model implications, as shown in Figure 3.

[17] To avoid multicollinearity, we control for lagged markup and lagged $\log(ME)$ when the dependent variables are their growth rate.

Table 12: Markup shock exposure and real quantities

	(1) $\Delta\mu$	(2) Δcf	(3) Δme	(4) Δk
$\varepsilon^{\mu} \times \beta^{\mu}$	0.002***	0.053***	0.083***	-0.012***
	(2.894)	(5.507)	(5.124)	(-3.415)
Observations	58,254	52,541	62,064	62,100
R-squared	0.491	0.347	0.544	0.317
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

This table shows the firm-level real quantities in response to aggregate markup shocks. Specifically, we run the regression in Equation (31). β^{μ}_{it} is the firm i's return exposure to the aggregate markup shock, it is normalized to have a standard deviation of 1. Firm-level controls include size $\log(ME)$, book-to-market ratio, book leverage, markup, and return on asset (ROA). We study a broad set of firm-level responses: $\Delta\mu$ is the growth rate of markup, Δcf is the growth rate of EBITDA, Δme is the growth rate of market capitalization, and Δk is the growth rate of PPEGT. We cluster the standard errors by firm and year. Statistical significance level: *** p<0.01, ** p<0.05, * p<0.1.

5 Conclusion

This paper explores the asset pricing implications of markup shocks. We build a new measure of markup shocks using the common component that drives the cross-section of firms' markups. We document two new empirical facts. First, investors require a significant negative risk premium for bearing markup risk. Second, markup shocks generate large differences in risk premia across firms that explains many documented anomalies. Taken together, these findings suggest that markup risk is an important source of aggregate risk that is priced in the cross-section of stock returns.

We rationalize these facts in a real business cycle with heterogeneous firms, imperfect competition, and markup shocks. A positive markup shock allows firms to extract more rents from consumers. Thus, an increase in markup is bad news for the representative household, which explains the negative price of risk. Firms, on the other hand, can benefit from markup shocks because they increase the present value of future rents. In short, the average firm's markup beta is positive, which explains the negative risk premium commanded by investors. We calibrate the model to match the distribution of both the level of markups and the markup exposures to the aggregate markup shock and find that it generates substantial cross-sectional return spreads consistent with the data.

In our model, we adopt a reduced-form approach to model the firm's heterogeneous exposures to markup shocks. We view micro-founding those exposures in a partial equilibrium production-based model (e.g., Zhang (2005)) with markup risk as an important challenge for future research. Our empirical findings also suggest that markup shocks have important impacts on the distribution of firms. Given the importance of markups for the transmission of monetary policy, extending our framework into a Heterogeneous Agent New Keynesian model (e.g., Kaplan et al. (2018)) appears an interesting avenue for future research.

References

Aguerrevere, F. L. (2009): "Real options, product market competition, and asset returns," *The Journal of Finance*, 64, 957–983.

Ai, H., R. Bansal, H. Guo, and A. Yaron (2020): "Identifying preference for early resolution from asset prices," *Working paper.*

Ai, H., J. E. Li, and J. Tong (2022): "Equilibrium Value and Profitability Premiums," *Working paper.*

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

Barkai, S. (2020): "Declining labor and capital shares," *The Journal of Finance*, 75, 2421–2463.

Barrot, J.-N., E. Loualiche, and J. Sauvagnat (2019): "The globalization risk premium," *The Journal of Finance*, 74, 2391–2439.

Begenau, J., M. Farboodi, and L. Veldkamp (2018): "Big data in finance and the growth of large firms," *Journal of Monetary Economics*, 97, 71–87.

Belo, F., X. Lin, and F. Yang (2019): "External equity financing shocks, financial flows, and asset prices," *The Review of Financial Studies*, 32, 3500–3543.

Belo, F. and J. Yu (2013): "Government investment and the stock market," *Journal of Monetary Economics*, 60, 325–339.

Bretscher, L. (2020): "From local to global: Offshoring and asset prices," *Available at SSRN 3268123.*

Bretscher, L., A. Hsu, and A. Tamoni (2020a): "Fiscal policy driven bond risk premia," *Journal of Financial Economics*, 138, 53–73.

——— (2020b): "The real response to uncertainty shocks: The risk premium channel," *Georgia Tech Scheller College of Business Research Paper.*

Bustamante, M. C. and A. Donangelo (2017): "Product market competition and industry returns," *The Review of Financial Studies*, 30, 4216–4266.

Campbell, J. Y. and J. H. Cochrane (1999): "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of political Economy*, 107, 205–251.

Campbell, J. Y. and R. J. Shiller (1988): "The dividend-price ratio and expectations of future dividends and discount factors," *The Review of Financial Studies*, 1, 195–228.

Chen, H., W. Dou, H. Guo, and Y. Ji (2020): "Feedback and contagion through distressed competition," *The Rodney L. White Center Working Papers Series at the Wharton School, Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.*

Cho, T., M. Grotteria, L. Kremens, and H. Kung (2021): "The Present Value of Future Market Power," *Available at SSRN.*

Clara, N. (2018): "Demand elasticities, nominal rigidities and asset prices," in *2018 Meeting Papers*, Society for Economic Dynamics, 790.

Clara, N., A. Corhay, and H. Kung (2021): "Firm Product Concentration and Asset Prices," *Available at SSRN 3488428.*

Cochrane, J. H. (1991): "Production-based asset pricing and the link between stock returns and economic fluctuations," *The Journal of Finance*, 46, 209–237.

——— (2009): *Asset pricing: Revised edition*, Princeton university press.

Corhay, A. (2017): "Industry competition, credit spreads, and levered equity returns," *Rotman School of Management working paper.*

Corhay, A., H. Kung, and L. Schmid (2020a): "Competition, Markups, and Predictable Returns," *Review of Financial Studies*, 33.

——— (2020b): "Q: risk, rents, or growth?" *working paper.*

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

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

Croce, M. M., T. T. Nguyen, S. Raymond, and L. Schmid (2019): "Government debt and the returns to innovation," *Journal of Financial Economics*, 132, 205–225.

Crouzet, N. and J. C. Eberly (2019): "Understanding weak capital investment: The role of market concentration and intangibles," Tech. rep., National Bureau of Economic Research.

——— (2021): "Rents and intangible capital: A Q+ framework," Tech. rep., National Bureau of Economic Research.

De Loecker, J. and J. Eeckhout (2018): "Global market power," Tech. rep., National Bureau of Economic Research.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 135, 561–644.

Doshi, H. and P. Kumar (2021): "Capital Investment and Asset Returns in Dynamic Oligopolies," .

Dou, W. W. and Y. Ji (2021): "External financing and customer capital: A financial theory of markups," *Management Science*, 67, 5569–5585.

Dou, W. W., Y. Ji, and W. Wu (2021a): "Competition, profitability, and discount rates," *Journal of Financial Economics*, 140, 582–620.

——— (2021b): "The Oligopoly Lucas Tree," *The Review of Financial Studies*, hhab120.

Edmond, C., V. Midrigan, and D. Y. Xu (2018): "How costly are markups?" Tech. rep., National Bureau of Economic Research.

Eeckhout, J. and L. Veldkamp (2021): "Data and Market Power," Tech. rep., Columbia University Working Paper Sep.

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

Farhi, E. and F. Gourio (2018): "Accounting for macro-finance trends: Market power, intangibles, and risk premia," Tech. rep., National Bureau of Economic Research.

Favilukis, J. and X. Lin (2013): "Long run productivity risk and aggregate investment," *Journal of Monetary Economics*, 60, 737–751.

——— (2016): "Wage rigidity: A quantitative solution to several asset pricing puzzles," *The Review of Financial Studies*, 29, 148–192.

Galí, J. (2015): *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press.

Greenwald, D. L., M. Lettau, and S. C. Ludvigson (2019): "How the wealth was won: Factors shares as market fundamentals," Tech. rep., National Bureau of Economic Research.

Gutiérrez, G., C. Jones, and T. Philippon (2019): "Entry costs and the macroeconomy," Tech. rep., National Bureau of Economic Research.

——— (2021): "Entry Costs and Aggregate Dynamics," .

Gutiérrez, G. and T. Philippon (2017): "Declining Competition and Investment in the US," Tech. rep., National Bureau of Economic Research.

Hall, R. E. (1988): "The relation between price and marginal cost in US industry," *Journal of political Economy*, 96, 921–947.

Hou, K., C. Xue, and L. Zhang (2015): "Digesting Anomalies: An Investment Approach," *Review of Financial Studies2*, 28, 650–705.

Ireland, P. N. (2004): "Technology shocks in the new Keynesian model," *Review of Economics and Statistics*, 86, 923–936.

Jermann, U. J. (1998): "Asset Pricing in Production Economies," *Journal of Monetary Economics*, 41, 257–275.

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

Kaplan, G., B. Moll, and G. L. Violante (2018): "Monetary policy according to HANK," *American Economic Review*, 108, 697–743.

Knesl, J. (2018): "Automation and the displacement of labor by capital: Asset pricing theory and empirical evidence," *Available at SSRN 3293405.*

Kogan, L. and D. Papanikolaou (2014): "Growth opportunities, technology shocks, and asset prices," *The journal of finance*, 69, 675–718.

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

Levinsohn, J. and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *The review of economic studies*, 70, 317–341.

Loualiche, E. (Forthcoming): "Asset pricing with entry and imperfect competition," *Journal of Finance.*

Manning, A. (2013): *Monopsony in motion*, Princeton University Press.

Palomino, F. (2012): "Bond risk premiums and optimal monetary policy," *Review of Economic Dynamics*, 15, 19–40.

Papanikolaou, D. (2011): "Investment shocks and asset prices," *Journal of Political Economy*, 119, 639–685.

Romer, P. M. (1990): "Endogenous technological change," *Journal of political Economy*, 98, S71–S102.

Rotemberg, J. J. and M. Woodford (1992): "Oligopolistic pricing and the effects of aggregate demand on economic activity," *Journal of political Economy*, 100, 1153–1207.

Smets, F. and R. Wouters (2003): "An estimated dynamic stochastic general equilibrium model of the euro area," *Journal of the European economic association*, 1, 1123–1175.

——— (2007): "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American economic review*, 97, 586–606.

Traina, J. (2018): "Is Aggregate Market Power Increasing?" *Chicago Booth mimeo.*

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

Appendix

A Production function estimation

The estimation procedure closely follows De Loecker et al. (2020). We define firm-level markup as $\mu_{i,t} = \alpha_i^V \frac{P_{it}Y_{it}}{P_{it}^V V_{it}}$, where $P_{it}Y_{it}$ is the firm revenues and $P_{it}^V V_{it}$ is the total variable input expenses. All nominal variables are converted to real terms by dividing GDP deflator. The output elasticity of the variable input α_i^v can be obtained by estimating the following industry-specific Cobb-Douglas production functions:[18]

$$y_{i,t} = \theta_0 + \alpha_i^v v_{i,t} + \alpha_i^k k_{i,t} + \omega_{i,t} + \varepsilon_{it},$$

where $y_{i,t}$ is the log real sales, $v_{i,t}$ is the log real variable input, $k_{i,t}$ is the log real capital stock, $\omega_{i,t}$ is the log productivity, and $\varepsilon_{i,t}$ is the measurement error term.

The estimation follows a two-stage procedure. In the first stage, we compute expected sales by regressing sales on a second-order polynomial approximation in k_{it} and $v_{i,t}$.

$$y_{it} = \theta_0 + \psi(k_{i,t}, v_{i,t}) + \varepsilon_{it}, \tag{A1}$$

where $\psi(k_{i,t}, v_{i,t}) = \sum_{m=0}^{2} \sum_{j=0}^{2-m} \hat{\delta}_{mn} k_{i,t}^m v_{i,t}^n$. We then use the estimated polynomial to predict the firm output as a function of the production inputs, $\widehat{\psi}(k_{i,t}, v_{i,t})$. This allows us to purge the error term $\varepsilon_{i,t}$ out of the firm's output. The underlying assumption is that the firm's output will respond to productivity $\omega_{i,t}$ but not the measurement error $\varepsilon_{i,t}$.

In the second stage, we obtain the estimates for the elasticity parameters, $\hat{\alpha}_i^v$ and $\hat{\alpha}_i^k$. First, using Equation (A1), we compute a prediction for $\omega_{i,t}$ for any candidate value of $\tilde{\alpha}_i^v$ and $\tilde{\alpha}_i^k$ as

[18] Since we do not separately observe the inputs and output price and quantity, we follow De Loecker et al. (2020) and use the sales and expenditures directly when estimating the production elasticities. This approach assumes the equality of marginal and average cost of production. De Loecker et al. (2020) provides a discussion on the conditions under which this estimation generates consistent estimates of the output elasticities.

follows,

$$\widehat{\omega}_{i,t} = \widehat{\psi}(k_{i,t}, v_{i,t}) - \tilde{\alpha}_i^k k_{i,t} - \tilde{\alpha}_i^v v_{i,t}.$$

Next, we assume that the productivity process follows an AR(1) process

$$\hat{\omega}_{i,t} = \gamma_0 + \gamma_1 \hat{\omega}_{i,t-1} + \xi_{it}, \tag{A2}$$

where ξ_{it} is a productivity shock, realized at time t.

The key identification assumption required to estimate the elasticities is that the variable input and next period capital respond to current productivity shocks, but their lagged values do not. Therefore, the following moment conditions can identify the elasticity parameters $\hat{\alpha}_i^v$ and $\hat{\alpha}_i^k$:

$$\mathbb{E}\left(\xi_{it}\left(\alpha^v, \alpha^k\right)\begin{bmatrix} v_{it-1} \\ k_{it} \end{bmatrix}\right) = 0.$$