Increases in Market Power: Implications for the Real Effects of Nominal Shocks*

Volker Hahn

Department of Economics
University of Konstanz
Box 143
78457 Konstanz, Germany
volker.hahn@uni-konstanz.de

First Version: February 2022
This Version: October 2023

Abstract

In many countries, market power in goods markets has increased over the last decades. We present a menu-cost model with endogenous markups that rationalizes this trend via productivity increases that are concentrated within a small set of firms. We show that these productivity changes entail increases in monetary non-neutrality. Aggregate productivity is procyclical as resources are reallocated across firms over the course of the business cycle. We identify a new amplification mechanism that relies on a direct effect of market concentration on the demand for individual goods. This mechanism strengthens endogenous fluctuations of aggregate productivity and has non-negligible implications for monetary non-neutrality.

Keywords: Price dynamics, increases in market power, menu costs, monetary non-neutrality, endogenous TFP, real rigidities, Kimball aggregator.

JEL: E52, E31, L16.

*I would like to thank Michal Marenčák, Subashini Sundar, seminar participants at AMSE, Marseille, and the National Bank of Slovakia, Bratislava, as well as participants of the 2022 Money-Macro-Finance conference in Warsaw and the 2023 MMF workshop in Portsmouth for many valuable comments and suggestions. I have no conflict of interest to declare.

1 Introduction

In recent decades, there has been a well-documented trend towards higher market power in goods markets as measured by higher markups over marginal costs (Diez et al., 2018; De Loecker et al., 2020).[1] This change is dominated by a comparably small set of firms. Thus average markups in the United States have increased from 21% in 1980 to 61% in 2016, but median markups have hardly changed (De Loecker et al., 2020). Similar patterns can be found for other countries as well (Aquilante et al., 2019).

This paper aims to enhance our understanding of the consequences of increases in market power for monetary non-neutrality, i.e. the degree to which nominal shocks have real consequences. For this purpose, we build a menu-cost model with a new variant of the Kimball aggregator (Kimball, 1995), which allows for endogenous changes in markups (see Baqaee et al., 2023; Edmond et al., 2023).

In our quantitative analysis, we focus on the UK and compare two different periods, a ten-year period before the 2007/2008 financial crisis and a ten-year period after the financial crisis but before the Covid-19 pandemic. According to data provided by the Office of National Statistics (ONS), the median markup in the post-crisis period is only slightly higher than in the pre-crisis period. By contrast, the 90th-percentile of markups over marginal costs increased more strongly from 152% to 192%. In line with the increase of high markups, mean markups over marginal costs increased from 44% to 59%.

In our model, these changes in the distribution of markups can be explained by permanent increases in the productivity of a small group of firms. The rise in the productivity of these firms makes aggregate productivity, which is endogenous in our model, more procyclical and causes a higher degree of monetary non-neutrality. The productivity changes also lead to decreases in the mean frequency of price adjustment and increases in the magnitude of price changes. Both implications for price dynamics are qualitatively in line with the empirical evidence.[2]

As changes in productivity alone cannot explain the empirical magnitudes of the changes in the size and frequency of price adjustment, we also employ an

[1] Autor et al. (2017) find a higher concentration of sales in major sectors of the US economy.

[2] We use the price quote data that is provided by ONS for the UK, which consists of a large number of prices that are used to construct the CPI in the UK. The data are publicly available and have been used by Adam and Weber (2019), Hahn and Marenčák (2020), Blanco (2021), and Carvalho and Kryvtsov (2021).

alternative approach and calibrate our model to the post-crisis period by allowing for different menu costs and sizes of idiosyncratic shocks than in the pre-crisis period. In particular, the observed low frequencies of price changes require larger menu costs. Due to the stronger nominal rigidities, the alternative calibration implies an even larger degree of monetary non-neutrality but smaller endogenous fluctuations of aggregate productivity than the first calibration approach. The higher degree of monetary non-neutrality obtained for the post-crisis period under both calibration approaches is compatible with a flattening of the Phillips curve.[3]

Our paper contributes to the literature on endogenous fluctuations in aggregate productivity in response to demand shocks (Basu, 1995; Meier and Reinelt, 2022; Bai et al., 2019). In particular, it features a misallocation channel, which involves that resources are reallocated from low-productivity firms to high-productivity firms over the business cycle (Baqaee et al., 2023; González et al., 2021; Cooke and Damjanociv, 2021). Baqaee et al. (2023) put differences in the pass-throughs of marginal costs into prices center stage. Because these pass-throughs are larger for low-markup firms than for high-markup firms, expansionary monetary shocks lead to a reallocation of resources from low-markup to high-markup firms, which have comparably high productivity. This shift of resources to highly productive firms enhances the efficiency of the allocation in a boom.

While our paper also features this channel, we identify a new amplification mechanism, which leads to substantially larger endogenous co-movements of TFP with output. Technically, this effect arises due to fluctuations in the Lagrange multiplier of the households' cost minimization problem. We demonstrate that the Lagrange multiplier can be given an interpretation as a measure of output differences across firms or, equivalently, as a measure of market concentration.

While, in most models using a Kimball (1995) aggregator, changes in this measure of market concentration are considered to be too small to be taken into account, we show that the reallocation of resources across firms leads to non-negligible procyclical variations of this measure over the business cycle.[4] These fluctuations in market concentration tend to amplify the effect of changes in the price level on the demand for individual goods for a fixed price of the good. Thus a positive

[3]There is an extensive discussion in the literature about the flattening of the Phillips curve over the last decades (see e.g. Mishkin, 2007). For a recent empirical analysis, see Hazell et al. (2020).

[4]Harding et al. (2022) study a non-linear model of the liquidity trap, which takes the dynamic evolution of the Lagrange-multiplier into account. However, they do not examine how its fluctuations affect business cycle dynamics and the endogenous response of TFP to shocks.

feedback loop arises as fluctuations in market concentration lead to larger real effects of nominal shocks for individual firms and thereby even stronger fluctuations in market concentration. This mechanism strengthens the misallocation channel substantially.

Our paper contributes to other strands of the literature as well. First, our framework involves that changes in market power are caused by changes in productivity. Thus it is related to Ganapati (2021), who shows that rising market power in the US is positively correlated with increases in productivity, as well as De Ridder (2021), who explains increases in market power by the investments of some firms into intangible inputs, which lower marginal costs but increase fixed costs.

Second, to be compatible with sizable endogenous changes in markups, our model uses a variant of the Kimball aggregator (Kimball, 1995). The Kimball aggregator is frequently used to introduce a form of real rigidity that allows for reasonable degrees of nominal rigidities to produce empirically plausible responses of real variables to nominal shocks (see e.g. Ball and Romer, 1990; Smets and Wouters, 2007; Dotsey and King, 2005; Lindé et al., 2016; Harding et al., 2022). Recently, standard parameterizations of the Kimball aggregator have been criticized because they require implausibly large idiosyncratic productivity shocks to generate the empirically observed sizes of price changes (see Klenow and Willis, 2016) and involve values for the super-elasticity of demand, i.e. the price elasticity of the elasticity of demand, that are not in line with microeconomic evidence (Beck and Lein, 2020).

We develop a new variant of the Kimball aggregator, which is characterized by two parameters with clear interpretations. One parameter describes the elasticity of demand for a firm with an intermediate level of output. The other parameter corresponds to the super-elasticity. Thus in our quantitative model, it is straightforward to set the super-elasticity to empirically plausible values. Our model also does not require excessively large idiosyncratic shocks to explain the observed movements in prices and is thus not susceptible to the critique by Klenow and Willis (2016).

Third, the transmission mechanism in standard new Keynesian models is criticized as implausible by Broer et al. (2020) because it involves that monetary stimulus lowers markups and profits, which, due to a negative income effect, leads to an increase in the labor supply.[5] Our framework involves similar or

[5] The cyclicality of markups is discussed by Bils et al. (2018) and Nekarda and Ramey (2020).

even higher profits in response to expansionary nominal shocks and thus is not susceptible to this critique.

Fourth, our paper is related to studies on the connection between market structure and monetary policy (Wang and Werning, 2020; Mongey, 2021). Mongey (2021) compares a menu-cost model with monopolistic competition with an oligopoly economy. The oligopoly model, calibrated to match the same moments of micro price data as the model with monopolistic competition, involves a higher degree of monetary non-neutrality. Bilbiie et al. (2007), Bergin and Corsetti (2008), Lewis and Poilly (2012), Colciago and Silvestrini (2021), and Cooke and Damjanociv (2021) examine how monetary-policy shocks affect firm entry and exit. The present paper abstracts from firm entry and exit.

Our paper is organized as follows. Section 2 presents a new variant of the Kimball aggregator, examines its properties and discusses its implications for price setting. In Section 3, we integrate this aggregator into a general-equilibrium model with menu costs and a roundabout production structure. We explain how we solve and calibrate this model in Section 4. Section 5 presents our main results for aggregate dynamics, the dynamics of prices, and the cyclicality of profits. Several extensions to our basic framework are discussed in Section 6. Section 7 concludes.

2 A New Variant of the Kimball aggregator

Before laying out the general-equilibrium framework, we introduce our variant of the Kimball aggregator and highlight some of its properties in a partial-equilibrium set-up. Moreover, we highlight the relevance of the Lagrange multiplier associated with the constraint on the household's cost-minimization problem for our analysis.

The aggregator proposed in this paper allows for a constant super-elasticity, which is defined as the price elasticity of the elasticity of demand. One advantage of this aggregator is that it is in line with the finding in Beck and Lein (2020) that the distribution of estimates of the super-elasticity is comparably tight. Another advantage is that it is characterized by only two parameters, both of which have a clear economic interpretation.

2.1 Definition

Consider a household that derives utility from consuming C. C is the composite of a continuum of differentiated goods $c(z)$, $z \in [0,1]$, that is implicitly defined by

$$\int_0^1 D\left(\frac{c(z)}{A(z)C}\right) dz = 1, \tag{1}$$

where $D(x)$ is a variant of the Kimball (1995) aggregator that satisfies the standard properties $D(1) = 1$, $D'(x) > 0$, and $D''(x) < 0$ for all $x \geq 0$. $A(z)$ ($A(z) > 0$) is an inverse measure of the quality of good z. Quality shocks are often used in the literature as a more tractable alternative to idiosyncratic productivity shocks (see e.g. Midrigan, 2011; Blanco, 2021).

In this paper, we propose to use the following functional form:

$$D(x) = \int_1^x (\varepsilon - s \ln x')^{\frac{1}{s}} \, dx' + 1, \tag{2}$$

where ε and s are positive parameters, for which we will give an interpretation later. It is straightforward to verify that this aggregator is well-defined for $x < e^{\frac{\varepsilon}{s}}$ and satisfies the properties stated above. The condition $x < e^{\frac{\varepsilon}{s}}$ entails that, for a given of level of C, the quantity of good z that is consumed by the household is lower than the satiation level $A(z)e^{\frac{\varepsilon}{s}}C$.

With $p(z)$ denoting the nominal price of consumption variety z, the cost-minimization problem of the household is

$$\min_{\{c(z)\}_{z=0}^1} \int_0^1 p(z)c(z)\, dz \qquad \textit{s.t. (1).} \tag{3}$$

This leads to the solution

$$c(z) = A(z) d\left(\frac{p(z)A(z)C}{\varepsilon^{\frac{1}{s}}\widetilde{\theta}}\right) C, \tag{4}$$

where

$$d(x) := e^{(1-x^s)\frac{\varepsilon}{s}} \tag{5}$$

and $\tilde{\theta}$ is the Lagrange multiplier associated with the constraint (1).

Let P be the total nominal cost of one unit of C if the household selects a cost-minimizing bundle of consumption varieties $\{c(z)\}_{z=0}^1$ for given prices $\{p(z)\}_{z=0}^1$

and quality levels $\{A(z)\}_{z=0}^{1}$. It will be useful to introduce the relative quality-adjusted prices of consumption variety z as $q(z) := A(z)p(z)/P$ as well as the transformed multiplier θ as

$$\theta := \left(\varepsilon^{\frac{1}{s}}\widetilde{\theta}\right) / (PC)\,. \tag{6}$$

We will explain later why it is advantageous to consider the transformed multiplier θ rather than the original multiplier $\widetilde{\theta}$.

With the help of $q(z)$ and θ, the demand for variety z can be specified as

$$c(z) = A(z)d\left(\frac{q(z)}{\theta}\right)C. \tag{7}$$

Henceforth we will often refer to quality-adjusted quantities, which are quantities of an individual good divided by its quality level $A(z)$. For example, quality-adjusted consumption of good z amounts to $c(z)/A(z)$.

We note that, for given nominal prices $\{p(z)\}_{z=0}^{1}$ and qualities $\{A(z)\}_{z=0}^{1}$, the multiplier θ and the price level P are implicitly determined by

$$\int_0^1 D\left(d\left(\frac{A(z)p(z)}{P\theta}\right)\right)dz = 1, \tag{8}$$

$$\int_0^1 \frac{A(z)p(z)}{P}d\left(\frac{A(z)p(z)}{P\theta}\right)dz = 1, \tag{9}$$

where the first condition follows from (1) and the second condition from the definition of the price index P.

2.2 Interpretation of θ

We have shown that the demand for a specific differentiated good does not only depend on its quality-adjusted relative price $q(z)$ but also on the value of the transformed Lagrange multiplier θ. As a consequence, we now examine the economic interpretation of θ more closely. For this purpose, it is helpful to consider a scenario where a fraction ν of firms produce identical quantities c_1 and the remaining firms produce identical quantities c_2. Without loss of generality, we assume that $c_2 \geq c_1$ and that all quality levels are equal to one. Now we consider changes in the quantities produced by the two groups of firms, while keeping aggregate consumption fixed.

We obtain the following proposition, which is shown in Appendix B:

Proposition 1. *For $c_2 \geq c_1$, C fixed, and strictly positive parameter s, an increase in c_2 is accompanied by a decrease in c_1 and an increase in θ. The minimum value of θ, which is attained for $c_1 = c_2 = 1$, is 1.*

First, it is obvious that an increase in consumption c_2 implies that c_1 has to be lowered in order to keep aggregate consumption C fixed. Second and more importantly, Proposition 1 suggests that θ can be interpreted as a measure of output differences across firms or a measure of market concentration. This measure of market concentration takes its lowest value of one when firms in the first group produce the same quality-adjusted quantities as firms in the second group. The larger the difference between consumption levels c_1 and c_2, the higher the multiplier θ is.

There are three advantages of formulating the model in terms of the transformed multiplier θ, which was introduced in (6), rather than $\widetilde{\theta}$. First, for a given set of quality-adjusted prices $\{A(z)p(z)\}_{z=0}^{1}$, the solution for θ does not depend on the level of C (see (8) and (9)). Second, multiplying all individual prices by a positive constant yields a corresponding change in the price level but leaves the value of θ unchanged (see again (8) and (9)). Third, the factor $\varepsilon^{\frac{1}{s}}$ in (6) is a normalization that ensures that θ attains a value of one in a completely symmetric situation where all firms produce identical quantities.

2.3 Elasticity and super-elasticity

In this section, we show that parameter s equals the super-elasticity of demand, while parameter ε can be interpreted as the demand elasticity of a firm z that produces an intermediate level of quality-adjusted output $c(z)/A(z) = C$.

It is straightforward to calculate the price elasticity of demand for good z as

$$\epsilon(z) = \varepsilon \left(\frac{q(z)}{\theta} \right)^{s}. \tag{10}$$

For $s > 0$, the elasticity of demand increases with the price, which is a key property in studies employing the Kimball aggregator and leads to demand curves with a smoothed-out kink. Moreover, for $s > 0$, the demand elasticity is a decreasing function of market concentration θ. Intuitively, a more unequal distribution of the quantities of differentiated goods makes it more difficult to substitute different varieties of goods for one another and thereby makes the demand for an individual good less elastic. Later we will show that, in the general equilibrium of our full model, market concentration θ is a procyclical variable. As a consequence, for a given relative quality-adjusted price $q(z)$, the elasticity $\epsilon(z)$ is countercyclical.

To obtain an interpretation of parameter ε, we note that, with the help of (5) and (7), (10) can be rewritten as

$$\epsilon(z) = \varepsilon - s \ln \left(\frac{c(z)}{A(z)C} \right). \tag{11}$$

Equation (11) reveals that, for $s > 0$, the elasticity of demand is a decreasing function of a firm's output. As $c(z)$ approaches the satiation level $A(z)e^{\frac{\varepsilon}{s}}C$, the demand becomes inelastic. In the special case where $c(z)/A(z) = C$, the elasticity of demand is

$$\epsilon(z) = \varepsilon. \tag{12}$$

Therefore ε can be interpreted as the elasticity of demand for a firm that produces an intermediate level of quality-adjusted output.

It is also instructive to compute the super-elasticity of demand, i.e. the price elasticity of the price elasticity. The super-elasticity is given by

$$\frac{\partial \ln \epsilon(z)}{\partial \ln q(z)} = s. \tag{13}$$

Hence our variant of the Kimball aggregator involves a constant super-elasticity of demand, which is given by parameter s.

It may be interesting as well to consider the special case $s \to 0$. According to (10), the elasticity of demand is given by $\epsilon(z) = \varepsilon$ for all firms z, irrespective of the prices that they choose. For $s \to 0$, we thus obtain the special case of a constant-elasticity of demand function and the standard Dixit-Stiglitz aggregator

$$C = \left(\int_0^1 \left(\frac{c(z)}{A(z)} \right)^{\frac{\varepsilon-1}{\varepsilon}} dz \right)^{\frac{\varepsilon}{\varepsilon-1}}. \tag{14}$$

It may be worth noting that $\theta = 1$ in this case, which reflects the fact that the dynamics of the multiplier play no role under a standard CES aggregator.

2.4 Implications for price-setting

To study the implications of our variant of the Kimball aggregator for price-setting, we consider a partial-equilibrium set-up without nominal rigidities where a firm can choose its price subject to its demand (7), taking its marginal cost $mc(z)$ of producing quality-adjusted output as given. In this case, the optimal quality-adjusted relative price $q(z)$ satisfies

$$\varepsilon\left(q(z)\right)^{s-1}\left(q(z) - mc(z)\right) = \theta^s. \tag{15}$$

In Appendix A, we show that, for all admissible values of s, this equation has a unique solution for $q(z)$ that satisfies $q(z) > mc(z)$. Moreover, we prove that the optimal relative price $q(z)$ increases with $mc(z)$. In line with (10), this implies that more productive firms, i.e. those with low marginal costs, face a lower elasticity of demand.

With the help of (10), Equation (15) can be rearranged as

$$q(z) = \frac{\epsilon(z)}{\epsilon(z) - 1} mc(z), \tag{16}$$

where $\frac{\epsilon(z)}{\epsilon(z)-1}$ is the desired markup of a firm under flexible prices.[6] As more productive firms face lower elasticities, markups are higher for these firms. Thus our variant of the Kimball aggregator has the potential to endogenously generate higher markups for subsets of firms that become more productive and accordingly face reduced marginal costs. This mechanism is in line with Ganapati (2021) and De Ridder (2021), who find an empirical association between lower marginal costs and larger market power.

It is also instructive to follow Baqaee et al. (2023) and to examine the desired pass-through, i.e. the elasticity of the firm's price under flexible prices with regard to its marginal cost. In Appendix D, we show

$$\frac{d \ln q(z)}{d \ln mc(z)} = \frac{1}{1 + \frac{s}{\epsilon(z)-1}}. \tag{17}$$

We note that, in the special case where $s = 0$, i.e. for a CES aggregator, desired pass-through is one. In the case of a positive super-elasticity, it is lower. In

[6] Firms always choose a price $q(z)$ that exceeds $\theta/\epsilon^{1/s}$, which is the price that they would choose for zero marginal cost $mc(z)$. Thus (10) yields $\varepsilon(z) > 1$, which ensures $\frac{\epsilon(z)}{\epsilon(z)-1} > 1$.

particular, firms that are very productive and therefore have a low elasticity of demand $\epsilon(z)$ have a low desired pass-through. Our approach thus generates an endogenous negative relationship between markups and pass-throughs, which is also a key element of the analysis in Baqaee et al. (2023). As discussed there, this relationship is supported by empirical evidence (see e.g. Amiti et al., 2019).

We will see that, in our general-equilibrium framework with menu costs, highly productive firms adjust their prices relatively infrequently. This is compatible with our finding of low pass-throughs, which, in a loose sense, implies that highly productive firms are rather reluctant to adjust their prices when economy-wide marginal costs change. Moreover, as price changes tend to be triggered not by changes in marginal costs but by idiosyncratic quality shocks, which have a comparably large variance, the relatively few price changes of highly productive firms tend to be large in our general-equilibrium model.

It is instructive to consider the case where $s = 2$. This value of s is only slightly higher than the one that we will employ for our full model and admits an analytical solution for the optimal price. The optimal price is given by

$$q(z) = \frac{mc(z)}{2} + \sqrt{\frac{\theta^2}{\varepsilon} + \left(\frac{mc(z)}{2}\right)^2}. \tag{18}$$

As a result, in the case where $s = 2$, firm z's markup over marginal cost is

$$\frac{q(z)}{mc(z)} = \frac{1}{2} + \sqrt{\frac{1}{\varepsilon}\left(\frac{\theta}{mc(z)}\right)^2 + \frac{1}{4}} > 1. \tag{19}$$

The desired markup, i.e. the markup that firms would choose in the absence of price rigidity, is thus affected via two channels. It decreases with the marginal cost $mc(z)$, and it increases with market concentration θ. It is straightforward to show with the help of (15) that these two channels are relevant also for values of s different from 2.

In the general equilibrium of our full model, there are thus two effects regarding the cyclical behavior of desired markups. First, real wages and thus marginal costs are procyclical, which tends to make desired markups countercyclical. Second, we will show that market concentration θ is procyclical because expansionary demand shocks shift resources towards large firms with high productivity. This effect tends to make desired markups procyclical. The second effect is also relevant for the procyclical profits that we will find later.

2.5 Comparison to standard parameterization

Our full model could also employ a standard parameterization of the Kimball aggregator (Klenow and Willis, 2016). One complication with the standard specification would be the tight link between the elasticity and the super-elasticity of demand, as we discuss now.

The standard functional form of the Kimball aggregator is

$$D_{standard}(x) := 1 + (\overline{\varepsilon} - 1)\exp\left(\frac{1}{\overline{\kappa}}\right)\overline{\kappa}^{(\overline{\varepsilon}/\overline{\kappa})-1}\left(\Gamma\left(\frac{\overline{\varepsilon}}{\overline{\kappa}}, \frac{1}{\overline{\kappa}}\right) - \Gamma\left(\frac{\overline{\varepsilon}}{\overline{\kappa}}, \frac{x^{\overline{\kappa}/\overline{\varepsilon}}}{\overline{\kappa}}\right)\right), \quad (20)$$

where $\Gamma(u, v)$ is the incomplete gamma function $\Gamma(u, v) := \int_v^\infty w^{u-1}e^{-w}\, dw$ and parameters $\overline{\theta}$ and $\overline{\kappa}$ satisfy $\overline{\theta} > 1$ and $\overline{\kappa} > 0$. For $\overline{\kappa} \to 0$, we obtain the CES aggregator with elasticity of substitution $\overline{\varepsilon}$.

On p. 452, Klenow and Willis (2016) specify expressions for the elasticity and super-elasticity of demand, which imply that the super-elasticity is always equal to $\overline{\kappa}/\overline{\varepsilon}$ times the elasticity. Later in our quantitative analysis, we use productivity differences to explain differences of 1.16 for median markups versus 2.52 for the 90th percentile. This approximately requires differences in demand elasticities by a factor of $1.16/(1.16-1)/(2.52/(2.52-1)) \approx 4.4$. Due to the tight link between elasticities and super-elasticities, this would also result in differences in super-elasticities by the same factor. As a consequence, a plausible value of the super-elasticity for L-firms would entail a very low value for the super-elasticity of H-firms. The low super-elasticity for H-firms would mean that they approximately face a CES demand function, which would render their markups virtually exogenous. Thus it would be impossible to explain the higher markups for these firms in the late period by increases in their productivities. The specification employed in this paper has the advantage of breaking the tight link between the super-elasticity and elasticity of demand.

3 Model

We integrate the variant of the Kimball aggregator described before into a standard new Keynesian model with idiosyncratic quality shocks, aggregate shocks to nominal spending, a roundabout production structure, and nominal stickiness in the form of menu costs as in Nakamura and Steinsson (2010).

A representative household has instantaneous utility

$$u(C_t) = \frac{C_t^{1-\gamma} - 1}{1-\gamma} - \frac{\omega}{\psi + 1}\mathcal{L}_t^{\psi+1}, \tag{21}$$

where γ is the constant coefficient of relative risk aversion, ψ is the inverse of the Frisch elasticity of labor, and ω is a positive weight. Future utilities are discounted by factor $\beta \in (0, 1)$. The consumption aggregate C_t is defined by (1). The household supplies labor $\mathcal{L}_t$ at real wage w_t. It owns the firms $z \in [0, 1]$ and receives total profits as dividends in every period. The resulting budget constraint is

$$C_t = w_t\mathcal{L}_t + \int_0^1 \pi_t(z)\, dz, \tag{22}$$

where $\pi_t(z)$ is firm z's real profit.

The household's optimal choice of labor is given by

$$w_t = \omega(\mathcal{L}_t)^{\psi}(C_t)^{\gamma}. \tag{23}$$

As has been described before, the household's cost minimization problem involves that its demand for good z is given by (7). For convenience, we repeat equations (8) and (9) as

$$\int_0^1 D\left(d\left(\frac{q_t(z)}{\theta_t}\right)\right) dz \;=\; 1, \tag{24}$$

$$\int_0^1 q_t(z) d\left(\frac{q_t(z)}{\theta_t}\right) dz \;=\; 1. \tag{25}$$

Each firm $z \in [0, 1]$ produces variety z according to the production function

$$y_t(z) = a_t(z)\left[A(z)\big(M_t(z)\big)^{s_m} L_t(z)^{1-s_m} - FC(z)\right]. \tag{26}$$

$L_t(z)$ is the amount of labor employed by firm z in the production process. $M_t(z)$ is a composite of differentiated intermediate goods, which will be discussed in more detail later. s_m with $s_m \in (0, 1)$ is the input share of intermediate inputs.

Firm z's constant productivity is denoted as $A(z)$. $FC(z)$ is a fixed cost of production. We distinguish between two groups of firms. The firms in the interval $[0, \alpha]$, where $\alpha \in (0, 1)$, have a low value of $A(z)$, which we normalize to 1. Their fixed costs are FC_L. The remaining firms, which we label firms of type H, have a high value of $A(z)$, which we label A_H ($A_H \geq 1$), and a different level of fixed costs FC_H. In our quantitative analysis, we will explain increases in market power via increases in productivity A_H.

In line with the literature, the quality shocks $a_t(z)$ show up not only in the consumption aggregator (see (1)) but in the production function as well. $a_t(z)$ follows a random walk

$$\ln a_t(z) = \ln a_{t-1}(z) + \xi_t(z), \tag{27}$$

where the shocks $\xi_t(z)$ are independent and normally distributed with mean zero and variance σ_ξ^2.

We employ a roundabout production structure like Basu (1995), where the different varieties of goods serve as consumption goods and, at the same time, as intermediate inputs. $m_t(z, z')$ is the quantity of intermediate inputs that firm z purchases from firm $z' \in [0, 1]$. The composite of intermediate goods $M_t(z)$ is implicitly given by

$$\int_0^1 D\left(\frac{m_t(z, z')}{a_t(z')M_t(z)}\right) dz' = 1, \tag{28}$$

where $D(x)$ is the same aggregator as the one that was used for the consumption aggregate C (see (2)). Hence the total demand for firm z's good is

$$y_t(z) = c_t(z) + \int_0^1 m_t(z', z)\, dz' = (C_t + M_t)\, a_t(z) d\left(\frac{q_t(z)}{\theta_t}\right), \tag{29}$$

where $M_t = \int_0^1 M_t(z')\, dz'$. Aggregate gross output Y_t is given by

$$Y_t = C_t + M_t. \tag{30}$$

While Y_t is gross output, which includes also the production of intermediate inputs, the variable C_t corresponds to value-added output. For this reason, we should think of C_t rather than Y_t as representing GDP.

Firm z's profits in period t are

$$\pi_t(z) = \frac{q_t(z)y_t(z)}{a_t(z)} - w_t L_t(z) - \int_0^1 \frac{q_t(z')m_t(z, z')}{a_t(z')}\, dz' - \chi(z) w_t I_t(z), \tag{31}$$

where $I_t(z)$ is an indicator variable that is zero if the firm does not change its nominal price, i.e. $p_t(z) = p_{t-1}(z)$, and one otherwise. The $\chi(z)$'s are exogenous parameters that specify how many units of labor a firm must employ in order to change the price of its output. For our calibration, we will choose the $\chi(z)$'s in a way such that the ratio of menu costs to sales is identical for all firms in the steady state.

Firm z's cost-minimization problem leads to

$$M_t(z) = \frac{s_m}{1-s_m} w_t L_t(z), \tag{32}$$

which, together with $M_t = \int_0^1 M_t(z)\, dz$ and $L_t = \int_0^1 L_t(z)\, dz$, implies

$$M_t = \frac{s_m}{1-s_m} w_t L_t, \tag{33}$$

The firm's quality-adjusted marginal cost, i.e. the cost of a marginal increase in quality-adjusted output $y_t(z)/a_t(z)$, is

$$mc_t(z) = \frac{1}{A(z)} \frac{1}{(s_m)^{s_m}(1-s_m)^{1-s_m}} (w_t)^{1-s_m} = \frac{1}{A(z)} mc_t, \tag{34}$$

where we have introduced the definition

$$mc_t := \frac{1}{(s_m)^{s_m}(1-s_m)^{1-s_m}} (w_t)^{1-s_m}. \tag{35}$$

In period t, firms discount profits at future dates t' ($t' \geq t$) with the factor

$$D_{t,t'} = \beta^{t'-t} \left(\frac{C_{t'}}{C_t} \right)^{-\gamma}. \tag{36}$$

The labor-market clearing condition is

$$\mathcal{L}_t = L_t + \int_0^1 \chi(z) I_t(z)\, dz. \tag{37}$$

Thus the total supply of labor, $\mathcal{L}_t$, has to equal the total demand for labor, which consists of the labor used for production ($L_t = \int_0^1 L_t(z)\, dz$) and for adjusting prices ($\int_0^1 \chi(z) I_t(z)\, dz$). As shown in Appendix C, the aggregate production function can be formulated as

$$Y_t = A_t \left[(M_t)^{s_m} (L_t)^{1-s_m} - FC \right], \tag{38}$$

where $FC := \int_0^1 FC(z)/A(z)\, dz$ is an aggregate fixed cost and aggregate productivity is

$$A_t := \left(\int_0^1 \frac{d\left(\frac{q_t(z)}{\theta_t}\right)}{A(z)}\, dz \right)^{-1}. \tag{39}$$

Thus aggregate productivity is endogenous and depends on the joint distribution of relative prices. The inverse of A_t is the measure of price dispersion that is typically used in new Keynesian models (Nakamura et al., 2018; Blanco, 2021). Increases in this measure of price dispersion correspond to a decrease in the efficiency of the allocation of resources across firms and thus to lower aggregate productivity.

We close the model by making the assumption employed by Nakamura and Steinsson (2010), among others, that nominal value-added output $S_t = C_t P_t$ follows an exogenous process

$$\ln S_t = \mu + \ln S_{t-1} + \eta_t, \tag{40}$$

where μ is the long-term growth rate of nominal output and thus determines trend inflation. The innovations η_t are independent and normally distributed with mean 0 and variance σ_η^2.

In principle, the entire distribution of relative prices $\{q_t(z)\}_{z=0}^1$ represents a state variable. Like Nakamura and Steinsson (2010), we assume that an approach in line with Krusell and Smith (1998) can be pursued and that it is sufficient to use S_t/P_t as an aggregate state variable. In this case, it follows that the firms' inflation expectations can be approximated by a function $\Gamma(x)$ in the following way:

$$\ln(P_t) - \ln(P_{t-1}) = \Gamma(\ln S_t - \ln P_{t-1}). \tag{41}$$

Under the assumption that $C_t = S_t/P_t$ provides a sufficiently accurate description of the aggregate state, firm z's optimization problem can be formulated recursively as

$$\begin{aligned} &V_z(q_t^*(z), C_t) \\ &= \max_{q_t(z)} \Bigg\{ \left[q_t(z) - \frac{mc_t}{A(z)} \right] Y_t d\left(\frac{q_t(z)}{\theta_t}\right) - \frac{FC(z)}{A(z)} mc_t - \chi_t(z) w_t I_t(z) \\ &\qquad + \mathbb{E}_t \left[D_{t,t+1} V_z\left(q_{t+1}^*(z), C_{t+1}\right)\right] \Bigg\}, \\ &\qquad s.t.\ \ q_{t+1}^*(z) = \frac{a_{t+1}(z) P_t}{a_t(z) P_{t+1}} q_t(z). \end{aligned} \tag{42}$$

where $V_z\left(q_t^*(z), C_t\right)$ is firm z's value function and $q_t^*(z)$ is the relative quality-adjusted price that the firm would charge if it did not adjust its nominal price. $I_t(z) = 0$ if $q_t(z) = q_t^*(z)$ and $I_t(z) = 1$ otherwise. An equilibrium of our economy is given by (23)-(25), (27), (30), (33), and (35)-(42).

In our analysis, we compare equilibria for two different periods. The first period is meant to capture the situation in the past with relatively low mean markup ("early period"). The second period describes the more recent situation where a comparably small group of firms, firms H, have very high markups and therefore the mean markup is higher than before ("late period"). We do not model transition dynamics, as the increase in market power is a phenomenon that has emerged over several decades (De Loecker et al., 2020). Together with the observation that prices are typically found to adjust every few quarters in the UK (Hahn and Marenčák, 2020), this makes it unlikely that the gradual transition process itself is important for understanding price dynamics and the aggregate effects of nominal shocks.

4 Numerical Analysis

4.1 Algorithm

The model is solved in a manner similar to Nakamura and Steinsson (2010). First, we compute the steady-state. The aggregate relationships (23), (30), (33), (37), and (38) are log-linearized around this steady state. As a consequence, one can conclude that Γ has to be linear as well and can be written as $\Gamma(x) = ax + b$ with coefficients a and b. Then one starts with a guess for $\Gamma(x)$ as well as a guess for the joint distribution of prices and $\hat{C}$, which is the log deviation of aggregate value-added output from its steady-state value. Based on these guesses, the firms' profits for the different states are calculated and value function iteration is applied to determine the firms' optimal price policies. These policies, in turn, are used to obtain an update for the joint distribution of prices and aggregate value-added output. For each value of the aggregate state $\hat{C}$, one then determines the degree to which (25) is violated, which is then used to adjust the function $\Gamma(x)$. One major difference from existing analyses is that θ has to be computed as a function of the aggregate state. For each value of $\hat{C}$, θ can be computed by evaluating (24) for the current estimate of the respective distribution of relative prices $\{q(z)\}_{z=0}^{1}$. Another point of departure from Nakamura and Steinsson (2010) is that we compute TFP as a function of $\hat{C}$, which can be done by evaluating (39) for the distribution of prices, conditional on different values of $\hat{C}_t$. Then we compute the total amount of labor used to adjust prices. Afterwards we update the aggregate log-linearized relationships. These aggregate relationships are then used to compute updated versions of firms' profits. The updated profits are used to perform value function again. This process is continued until convergence.

4.2 Calibration

For the scenario with low markups ("early period"), we choose the period from February 1996 to January 2006. This period is the earliest ten-year period for which ONS price quote data, which we use to obtain empirical moments of the distribution of price changes, is available.[7] For the second period ("late period"), we select February 2010 to January 2020. This period is the most recent period that spans ten years and does not cover the Covid-19 pandemic, during which a sizable fraction of prices are missing.

parameter	value
s	1.59
β	$0.96^{1/12}$
ψ	1
γ	1
s_m	0.7
$\overline{L}$	1/3

Table 1: Parameterization

Several parameters are chosen in line with other papers in the literature. We set the super-elasticity s to 1.59, which is the value found in Beck and Lein (2020).[8] We select standard values for the discount factor, $\beta = 0.96^{1/12}$, the inverse of the Frisch elasticity of the labor supply, $\psi = 1$, and the inverse of the intertemporal elasticity of substitution, $\gamma = 1$. Parameter ω is chosen such that steady-state labor in a flex-price equilibrium is $L^* = 1/3$. Moreover, we follow Nakamura and Steinsson (2010) and select μ equal to the mean of nominal GDP growth net of

[7] This data is also used in Adam and Weber (2019), Hahn and Marenčák (2020), and Blanco (2021). As is common in the literature, we remove sales. For this purpose, we use ONS' classification of sales. See Hahn and Marenčák (2020) for a more detailed description of the ONS price quote data.

[8] Beck and Lein (2020) use data from Belgium, Germany, and the Netherlands. We thus implicitly assume that the super-elasticity of demand is comparable in the UK.

real GDP growth and σ_η equal to the standard deviation of the growth rate of nominal GDP. The parameter values discussed so far are summarized in Table 1.

We need to determine several additional parameters for the scenario describing the early period, namely the productivity of the more productive H-firms, A_H, the share of firms with a low productivity, α, the fixed costs of both types of firms, FC_L and FC_H, the demand elasticity when $c(z)/A(z) = C$, ε, the ratio of menu costs and sales in the steady state, which we label K, as well as the variance of quality shocks, σ_ξ^2. We use data on the mean, median, and 90th percentile of markups as well as the median and 90th percentile of profit margins published by ONS and compute time averages for the periods under consideration.[9] As parameter ε determines the demand elasticity and markup of a firm with an intermediate output, we determine ε by targeting the median markup. As has been mentioned before, the markups of highly productive firms are high due to their high productivity. Thus we determine A_H by targeting the 90th percentile of markups. Parameter α describes the share of less productive firms. As less productive firms have low markups and the remaining firms have high markups, we use mean markups to pin down α. Parameters FC_L and FC_H stand for the fixed costs of less productive and highly productive firms. They are determined by targeting the 50th and 90th percentiles of profit margins.

The standard deviation of idiosyncratic quality shocks has a sizable impact on the magnitude of price changes. We use the ONS price quote data to calculate the average magnitude of the changes of log prices and use this value as a target for our calibration.[10] Finally, the size of menu costs is a major determinant of the frequency of price adjustment. Thus we try to match the mean frequency of price adjustment, which we calculate from ONS data as well.

For the scenario describing the late period, we follow two different approaches. In approach A, we change only parameters α, A_H, FC_L, and FC_H, which allows us to study the consequences of productivity changes in isolation. For this purpose, we choose A_H to target the 90th percentile of markups, parameter α to target average markups, and FC_L and FC_H to target the median and 90th percentile of profit margins. In approach B, we recalibrate the entire model for the late period.

[9] The data set "Experimental Statistics on markups, market power, productivity growth and business dynamism from the Annual Business Survey (ABS), 1997-2019, Great Britain" contains two different measures of markups, markups on labor costs and markups on intermediate consumption. We use the average of these two measures.

[10] When computing the averages, we use the weights used by ONS for the calculation of the CPI. All time periods are weighted equally.

	early period	late period A	late period B
A_H	5.60	6.91	5.96
α	98.3%	98.1%	96.5%
σ_ξ	0.0477	0.0477	0.0719
ε	7.0	7.0	5.8
K	4.8%	4.8%	10.4%
FC_L	0.006	0.007	0.010
FC_H	1.9	2.9	1.1
μ	0.0014	0.0016	0.0016
σ_η	0.0029	0.0022	0.0022

Table 2: Parameters. The early period captures February 1996 to January 2006. The late period captures February 2010 to January 2020. Approach A for the late period assumes that only parameters values for α, A_H, FC_L, and FC_H differ from those in the first period. Approach B recalibrates all parameters.

	early period	late period A	late period B
size of price changes	12.1% (12.1%)	12.3% (19.0%)	19.0% (19.0%)
freq. of price changes	15.6% (15.6%)	15.3% (14.3%)	14.3% (14.3%)
average markup	1.44 (1.44)	1.59 (1.59)	1.59 (1.59)
markup, 50% quantile	1.16 (1.16)	1.16 (1.20)	1.20 (1.20)
markup, 90% quantile	2.52 (2.52)	2.92 (2.92)	2.92 (2.92)
profit margin, 50% quantile	9% (9%)	8% (8%)	8% (8%)
profit margin, 90% quantile	42% (42%)	47% (47%)	47% (47%)

Table 3: Moments obtained from the model. Empirical values in parentheses. The early period captures February 1996 to January 2006. The late period captures February 2010 to January 2020. Approach A for the late period assumes that only parameters values for α, A_H, FC_L, and FC_H differ from those in the first period. Approach B recalibrates all parameters.

The resulting parameters summarized in Table 2. Table 3 shows the moments from our simulations and the targets in parentheses. One can see that all targets can be met for the early period. Approach A for the late period, where only the productivity-related parameters A_H, α, FC_L, and FC_H are adjusted, implies the correct mean and 90th percentile of markups and generates a reasonable value for the average markup, which was not targeted. As will be discussed more thoroughly later, the increase in A_H also leads to less frequent but larger price changes. Compared to the data, both changes go in the right direction but are not large enough.

Approach B, which recalibrates all internally determined parameters, is able to achieve all targets. For this purpose, it employs larger idiosyncratic shocks and stronger nominal rigidities, i.e. a larger value of K. The implication that price stickiness has gone up is reasonable under the assumption that menu costs are not technological in nature but capture costs of price changes that result from antagonizing customers, for example (see e.g. Blinder et al., 1998). That technological progress does not lead to a reduction in menu costs is also suggested by Nakamura et al. (2018) who find evidence that menu costs have not dropped in the US over four decades. In order to match the level of median markups, which is higher than in the pre-crisis period, the calibration also requires a lower level of ε, which is the elasticity of demand for a firm producing an intermediate level of quality-adjusted output.

As a next step, we examine whether our calibration involves a plausible magnitude for the menu costs. Levy et al. (1997) estimate menu costs to be 0.7% of sales. To compare this with our calibration, we note that K is the size of the menu cost, measured as a fraction of sales, conditional on the firm actually adjusting the price. If we multiply the value of K with the respective mean frequencies of price adjustment for the early period and approach A for the late period, we get values that are very close to 0.7%. Approach B results in menu costs that are approximately twice as large but still in the ballpark of values used in the literature. Finally, we would like to mention that, in all three cases, the error that firms make by using the approximation (41) to form their inflation expectations is negligible.

5 Results

5.1 Price dynamics

Table 4 summarizes key moments from our simulations. In all three scenarios, the frequency of price changes is lower for firms with high productivity than for low-productivity firms. This is compatible with our previous observation that high-productivity firms are more reluctant to adjust prices in response to changes in marginal costs (see our discussion of pass-throughs in Section 2.4).[11] As high productivity firms have high markups, we obtain a negative correlation between markups and the frequency of price adjustment, which is in line with the empirical

[11]For L-firms, the frequency of price adjustment is approximately twice as high as in the case of H-firms. Differences in the frequency of price adjustment can be observed empirically as well (Berger and Vavra, 2019). Our analysis of the ONS price quote data shows that the first quartile of the item-specific distribution of the frequencies of price changes is 3.2% and the third quartile 20.0%. Thus our model-implied differences in the frequencies of price changes are not implausibly large.

	early period	late period A	late period B
size of price changes	12.1%	12.3%	19.0%
freq. of price changes	15.6%	15.3%	14.3%
size of price changes (L)	11.2%	11.1%	17.6%
size of price changes (H)	15.5%	15.7%	23.6%
freq. of price changes (L)	17.3%	17.4%	15.9%
freq. of price changes (H)	9.3%	9.0%	9.1%
size of L sector	78.9%	74.8%	76.6%
mean elasticity	6.3	6.2	5.3
std. dev. $\hat{C}_t$	0.0025	0.0020	0.0021
monetary non-neutrality	0.84	0.89	0.93
TFP change in resp. to $\hat{C}$ change	6.4%	8.2%	6.3%
ch. of log theta in resp. to $\hat{C}$ ch.	4.5%	5.8%	4.3%
steady-state value of theta	1.062	1.082	1.070

Table 4: Simulation outcomes. Mean values weighted by shares of total revenues. The early period captures February 1996 to January 2006. The late period captures February 2010 to January 2020. Approach A for the late period assumes that only parameters values for α, A_H, FC_L, and FC_H differ from those in the first period. Approach B recalibrates all parameters.

evidence presented in Meier and Reinelt (2022) and also a key element in Baqaee et al. (2023).

Compared to less productive firms, highly productive firms adjust prices by larger amounts in all three scenarios. This can be understood by noting that there are two sources of price changes in our model. First, price changes can be caused by changes in markups, which arise in response to aggregate fluctuations. These price changes are typically small. Second, price changes can be triggered by idiosyncratic quality shocks. These changes are typically large. As highly productive firms tend not to respond to changes in markups, the latter source of price changes dominates, which results in comparably large average price changes.

We note that approach A for the late period yields larger but less frequent price changes than the early period, despite identical menu costs and magnitudes of idiosyncratic quality shocks in both scenarios. This is mainly due to the larger market share of H firms in the late period compared to the early period. The higher market share results from a small increase in the number of these firms, $1 - \alpha$, and, more importantly, from the productivity increases, i.e. higher values of A_H, which make highly productive firms produce even larger quantities.

We also note that, according to Table 4, the mean elasticities, computed by weighting firms' elasticities with their respective shares of total sales, are always lower than the respective parameter ε, which is the elasticity that a firm z with an intermediate level of quality-adjusted output $y(z)/A(z) = Y$ would have. This can be explained by noting that the highly productive firms, whose demand is quite inelastic, have comparably high shares of total sales. The values of the mean elasticities are similar to the expenditure-weighted mean elasticities found in Beck and Lein (2020), which are around 5.

5.2 Monetary non-neutrality

In this section, we demonstrate that the higher productivity and thus market power of H-firms in the late period compared to the early period results in stronger real effects of monetary disturbances. We use the ratio of the standard deviation of log value-added output, $\sigma_{\hat{C}}$, and the standard deviation of the shock to the growth rate of nominal spending, σ_η as our measure of monetary non-neutrality. This measure allows us to take the larger size of aggregate nominal shocks in the late period compared to the early period into account. Our choice can be justified further by the fact that $\sigma_{\hat{C}}$ and σ_η are always proportional to one another. In particular, it is straightforward to show that

$$\sigma_{\hat{C}} = \frac{1-a}{\sqrt{1-(1-a)^2}}\sigma_\eta, \tag{43}$$

where a is the slope of $\Gamma(x)$. Thus monetary non-neutrality is

$$mnn := \frac{\sigma_{\hat{C}}}{\sigma_\eta} = \frac{1-a}{\sqrt{1-(1-a)^2}}. \tag{44}$$

As a consequence, the steeper $\Gamma(x)$ is, i.e. the higher a ($0 < a \leq 1$), the lower the degree of monetary non-neutrality mnn. For $a = 1$, we would obtain complete monetary neutrality, i.e. $mnn = 0$. In this case, a shock to the growth rate of nominal spending leads to an increase in the inflation rate by the same amount and leaves aggregate real activity unchanged.

Table 4 shows that, despite a lower value of $\sigma_{\hat{C}}$ for approach A in the late period than in the early period, monetary non-neutrality is slightly higher in the late period. As in Baqaee et al. (2023), TFP is procyclical but, as we will discuss later, in contrast with Baqaee et al. (2023), the cyclicality of TFP relies on the cyclicality of θ to a large extent.

The mechanism underlying the procyclical fluctuations of TFP is illustrated in Figure 1, which shows the ratio of the mean markups for H-firms and for L-firms.

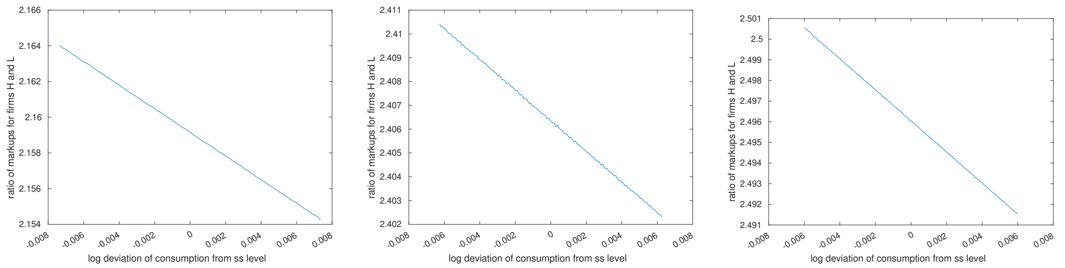

Figure 1: Ratio of the mean markups of H-firms and L-firms as a function of $\hat{C}$. From left to right: early period, late period A, late period B. The early period captures February 1996 to January 2006. The late period captures February 2010 to January 2020. Approach A for the late period assumes that only parameter values for α, A_H, FC_L, and FC_H differ from those in the first period. Approach B recalibrates all parameters.

Because low-productivity firms change their prices more frequently and more strongly in response to aggregate demand shocks, the markups of H-firms decline relative to the markups of L-firms in response to expansionary demand shocks. The relative decline in markups for H-firms leads to a reallocation of resources to H-firms, which have a high level of markups and thus produce inefficiently low quantities. This reallocation enhances the efficiency of the allocation and thus implies a higher aggregate productivity. Baqaee et al. (2023) examine firm-level evidence on markups and firm size and find that this reallocation has empirical support.[12] Moreover, they show that a procyclical behavior of TFP can be found for different measures of aggregate productivity and business-cycle indicators.

Calibration approach B for the late period, which allows for changes in menu costs and the magnitude of idiosyncratic shocks, involves a higher degree of monetary non-neutrality than the first calibration. This is due to higher menu costs and the resulting lower frequency of price adjustment. According to Table 4, the drop in the frequency of price adjustment is driven by the less productive firms. Compared to approach A, this weakens the misallocation channel, which relies on larger average price increases in a boom by less productive firms than by highly productive firms. In line with these arguments, approach B entails smaller changes in TFP concomitant with changes in real activity.

In Baqaee et al. (2023), the impact of demand shocks on endogenous TFP decreases with the size of the Frisch elasticity of the labor supply and vanishes

[12] Using Spanish data, González et al. (2021) identify related effects for investment. Expansionary monetary-policy shocks increase the investment of high-productivity firms and thus increase aggregate productivity. Duval et al. (2021) find that real sales of firms with larger markups respond less to monetary-policy shocks than the sales of low-markup firms.

completely for a very large elasticity. By contrast, we obtain sizable changes in TFP over the business cycle even when we set the inverse of the Frisch elasticity of the labor supply, ψ, to zero. One major difference between Baqaee et al. (2023) and the present paper is that, in our framework, the dynamics of θ_t are important for understanding the magnitude of monetary non-neutrality, as will be explained in the following section.

5.3 Dynamics of θ_t

To understand the new amplification mechanism studied in this paper, it is helpful to examine the dynamics of θ_t in more detail. As has been stressed before, θ_t can be interpreted as a measure of market concentration or output inequality across firms. It is thus plausible that the two scenarios for the late period, which involve a higher productivity and thus a larger market share of the H-firms, imply a higher value of market concentration θ in the steady state (see Table 4).

We have already discussed that expansionary aggregate shocks reallocate resources from low-productivity firms to high-productivity firms. Because high-productivity firms are larger than low-productivity firms, this reallocation leads to increases in market concentration when the economy is hit by an expansionary shock. This is confirmed by Table 4, which shows that θ_t is procyclical in all scenarios considered by us.

	late period B	late period B (θ_t fixed)
size of price changes	19.0%	19.0%
freq. of price changes	14.3%	14.3%
size of price changes (L)	17.6%	17.6%
size of price changes (H)	23.6%	23.6%
freq. of price changes (L)	15.9%	15.9%
freq. of price changes (H)	9.1%	9.1%
size of L sector	76.6%	76.6%
mean elasticity	5.3	5.3
std. dev. $\hat{C}_t$	0.0021	0.0018
monetary non-neutrality	0.93	0.80
TFP change in resp. to $\hat{C}$ change	6.3%	0.4%

Table 5: Relevance of the new amplification mechanism. The column in the middle reproduces results from Tables 2 and 4 for the late period. The column on the right-hand side shows the respective results from a simulation where θ_t is held fixed at its steady-state value.

As a next step, we illustrate that, although the changes in θ_t over the business cycle may appear small, the effects on aggregate dynamics are nevertheless sizable. For this purpose, we have solved our model for scenario B in the late period under the restriction that θ_t is always fixed at its steady-state level. The results are displayed in the last column of Table 5. The unconditional moments of the distribution of price changes are unaffected, which is plausible as the restriction that θ_t is always at its steady-state level does not affect the steady state. However, making θ_t constant dampens the endogenous changes of TFP over the business cycle substantially. Moreover, the assumption of θ_t being fixed also has a non-negligible effect on monetary non-neutrality.

The intuition for why changes in θ_t have substantial effects on aggregate dynamics can be gleaned from the demand function, (7), which can also be stated as:

$$c_t(z) = a_t(z) d \left(\frac{a_t(z) p_t(z)}{\theta_t P_t} \right) C. \tag{45}$$

Because θ_t is procyclical, it co-varies with the price level P_t. One can think of $\theta_t P_t$ as being the price level that is relevant for the real demand for individual goods. Thus changes in P_t are amplified by corresponding changes in θ_t and are associated with comparably large changes in the demand-relevant price level $\theta_t P_t$. This effect is particularly important for highly-productive firms, which change their nominal prices relatively infrequently. For fixed nominal prices $p_t(z)$, drops in the demand-relevant price level $\theta_t P_t$ then lead to larger drops in relative prices $q_t(z)$ and thus to a stronger reallocation of resources to highly productive firms in the presence of positive demand shocks.

In the following, we discuss the magnitude of changes in the demand-relevant price level θP compared to the changes in P. Suppose that initially consumption is at its steady-state level and that, due to a positive aggregate shock, the log deviation of consumption from its steady-state level increases to $\hat{C} > 0$. It can be shown that this leads to a log price level that is higher by $a/(1-a) \cdot \hat{C}$ than it would be in the steady-state, where a is the slope of the Γ function (see (41)). In our model, typical values of a are around $1/3$, which implies $a/(1-a) \approx 0.5$. According to Table 4, the positive shock leads to increases in the log of θ of approximately $0.05 \cdot \hat{C}$. In the case under consideration, changes in the demand-relevant price level θP are thus amplified by a factor $(0.5 + 0.05)/0.5 = 1.1$ compared to changes in P.

Finally, it may be interesting to relate our findings about the importance of fluctuations in θ_t to Baqaee et al. (2023). They state that changes in the price aggregator that is relevant for demand curves (up to a constant identical to $\theta_t P_t$ in our paper) and changes in the ideal price index (P_t in our paper) are first-order equivalent (see footnote 17 on p. 9 of their paper). In our model, this statement is equivalent to saying that changes in θ_t can be neglected as a first-order approximation. It is instructive to examine this point more closely.

The first-order condition for the household's cost-minimization problem and the definition of θ in (6) imply that we can express θ_t as

$$\theta_t = \varepsilon^{\frac{1}{s}} \left(\int_0^1 D' \left(\frac{c_t(z)}{C_t} \right) \frac{c_t(z)}{C_t} \, dz \right)^{-1}. \tag{46}$$

Clearly, for constant relative prices of differentiated goods, homothetic preferences imply that the ratios $c_t(z)/C_t$ remain fixed for all firms z as C_t changes, which entails that θ_t is unaffected by changes in C_t. However, in our model relative prices vary over the business cycle, which leads to the reallocation of resources towards high-productivity firms in a boom. As a consequence, our model involves non-negligible fluctuations in θ_t in response to demand shocks.

5.4 Cyclical properties of profits

Recently, standard new Keynesian models have been criticized by Broer et al. (2020) because the transmission mechanism of monetary shocks relies on countercyclical profits. More specifically, the low profits caused by a positive shock to nominal spending induce a negative income effect and thereby lead to an increase in the labor supply.

The left panel of Figure 2 confirms that, for the standard case of a CES aggregator, which will be considered in more detail in Section 6.4, aggregate profits are countercyclical. This can be understood by observing that, in standard new Keynesian models, markups are strongly countercyclical because marginal costs are procyclical while price stickiness, coupled with real rigidities, keeps goods prices rather stable in response to aggregate shocks. As an implication of countercyclical markups, profits are countercyclical as well.

The right panel shows that profits are mildly procyclical in the late period for our main model with the Kimball aggregator. In our main model, there are two opposing effects regarding the cyclicality of markups. First, the effect that

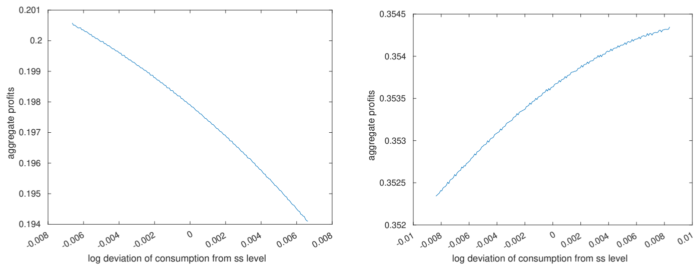

Figure 2: Aggregate profits divided by steady-state value-added output as a function of $\hat{C}$. Left panel: late period with a CES aggregator. Right panel: late period with our variant of the Kimball aggregator.

tends to make markups countercyclical due to price stickiness is also present in our analysis. Second, we have shown in Section 2.4 that procyclical changes in θ_t tend to make desired markups procyclical. This effect makes markups less countercyclical on average and thereby profits procyclical. In principle, there is also a third effect on profits because aggregate menu costs are inversely hump shaped as a function of real activity. As can be shown, this effect is dominant in the first period and thus leads to hump-shaped profits.

5.5 Dynamic response to aggregate shocks

One might be interested in the aggregate dynamics of the economy in response to a shock η_t to the growth rate of nominal demand. As the process for nominal value-added output is fairly simple and as there is only one relevant aggregate state variable (as in Nakamura and Steinsson, 2010), the dynamics of real GDP and inflation are straightforward to describe. It can be shown by combining (40), (41), and $C_t = S_t/P_t$ that the log-deviation of aggregate consumption from its steady-state level follows the law of motion

$$\hat{C}_t = (1-a)\hat{C}_{t-1} + (1-a)\eta_t, \tag{47}$$

where a is the slope of $\Gamma(x)$. Typically, our model implies values of $a \approx 1/3$. Thus the autocorrelation of $\hat{C}_t$ is around 2/3.

A lower value of a, which we obtain for the late period compared to the early period, has two consequences for the dynamics of $\hat{C}_t$. First, a demand shock η_t has a larger effect on impact. Second, the shock has a more persistent effect on $\hat{C}_t$. Both effects of a lower value of a contribute to a higher monetary non-neutrality.

It is immediate to show that inflation $\Pi_t := \ln P_t - \ln P_{t-1}$ satisfies

$$\Pi_t = \frac{a}{1-a}\hat{C}_t + \mu. \tag{48}$$

Thus the deviation of inflation from its long-term trend μ is always proportional to $\hat{C}_t$. It may be worth highlighting that (47) and (48) hold in Nakamura and Steinsson (2010) as well.

6 Robustness

6.1 Overview

In the following, we consider several changes to our framework: stochastic menu costs, idiosyncratic productivity shocks rather than quality shocks, and a CES aggregator as opposed to our variant of the Kimball aggregator. Moreover, we show how our model could be extended easily to match an empirically plausible distribution of firm sizes.

6.2 Stochastic menu costs

To assess the robustness of our findings, we consider a variant of our model with stochastic menu costs where menu costs are zero with probability ϕ and equal to $\chi(z)w_t$ otherwise. We assume that fifty percent of all price changes involve zero cost, which is in line with Blanco (2021). While monetary non-neutrality is somewhat higher than in our main model with non-stochastic menu costs (1.08 in the early period and 1.14 and 1.16 in the two late-period scenarios as opposed to 0.84, 0.89, and 0.93 in the menu-cost model), our results are otherwise qualitatively very similar. More details on this variant of our model can be found in Appendix E.

6.3 Idiosyncratic Productivity Shocks

One might wonder whether our results are robust to the inclusion of idiosyncratic productivity shocks as in Nakamura and Steinsson (2010) rather than quality shocks. A previous version of this paper has examined this scenario and has found very similar results. In particular, the model with productivity shock also involves a misallocation channel in which our amplification mechanism is present. One advantage of the model with quality shocks is that the approximation for inflation expectations in (41) is highly accurate. By contrast, the accuracy of inflation expectations was worse in the previous version of the paper but still comparable to the accuracy reported in Nakamura and Steinsson (2010).

6.4 CES aggregator

It may be instructive to compare our findings with those one would obtain with a standard CES aggregator. In such a model, it is no longer possible to explain differences in markups by differences in productivities across firms. Thus we do not distinguish between L-firms and H-firms. Where possible, all parameters are calibrated as before. For the late period, we recalibrate the model entirely, as in approach B for our main model.

The parameter values are summarized in Table 6. Perhaps surprisingly, some parameter values are similar to those that we obtain for our main model. In particular, the values of menu costs and the magnitudes of idiosyncratic shocks are comparable. As mentioned in Klenow and Willis (2016), parameterizations of the Kimball aggregator used in the literature often require implausibly large idiosyncratic shocks in order to generate plausible magnitudes of price changes. For our main model, we adopt a specification with a plausible value for the super-elasticity. As a consequence, our main model does not require larger shocks than a model based on a CES aggregator.

	early period	late period
σ_ξ	0.0490	0.0739
ε	7.6	6.3
K	5.0%	10.4%
FC	0.004	0.007
μ	0.0014	0.0016
σ_η	0.0029	0.0022

Table 6: CES aggregator: Calibration

Table 7 shows that the targeted moments can be matched. Obviously, average markups and the 90% quantile of markups cannot be matched by a model based on a CES aggregator. Similarly, the model with a CES aggregator fails to produce the high profit margins at the top. This is not surprising as the model does not contain the highly productive H-firms, which have high markups and thus also tend to have high profits.

	early period	late period
size of price changes	12.1% (12.1%)	19.0% (19.0%)
freq. of price changes	15.6% (15.6%)	14.3% (14.3%)
average markup	1.15 (1.44)	1.19 (1.59)
markup, 50% quantile	1.16 (1.16)	1.20 (1.20)
markup, 90% quantile	1.22 (2.52)	1.31 (2.92)
profit margin, 50% quantile	9% (9%)	8% (8%)

Table 7: CES aggregator: Moments

According to Table 8, the CES aggregator leads to smaller real effects of nominal disturbances compared to our main model. This does not come unexpected as the Kimball aggregator is often introduced into new Keynesian models as a means of adding real rigidities, which lead to larger degrees of monetary non-neutrality. The increase in monetary non-neutrality between both periods is due to the higher degree of nominal non-neutrality in the model with the CES aggregator.

Compared to our main model, the model with the CES aggregator involves substantially smaller consequences of demand shocks for aggregate productivity, as the misallocation channel and, in particular, our amplification mechanism are absent. Nevertheless expansionary shocks lead to small increases in aggregate productivity. These are due to the higher frequency of price adjustment, which tends to align the prices of firms and thereby reduces inefficient price dispersion.

	early period	late period
size of price changes	12.1%	19.0%
freq. of price changes	15.6%	14.3%
mean elasticity	7.6	6.3
std. dev. $\hat{C}_t$	0.0020	0.0017
monetary non-neutrality	0.67	0.74
TFP change in resp. to $\hat{C}$ change	0.8%	1.1%

Table 8: CES aggregator: Results

6.5 Firm sizes

While our model matches several moments of the distribution of firms' markups and profit margins, it is not directly able to match an empirically plausible distribution of firm sizes. Nevertheless it is straightforward to specify a model that (i) allows for an arbitrary distribution of firm sizes and thus also for an empirically plausible one and that (ii) delivers results that are identical to those from our model.

For this purpose, one needs to introduce additional positive, constant parameters $S(z)$, $z \in [0,1]$, which affect the size of each firm z in the steady state.[13] With these additional parameters, we change (1), which implicitly specifies the consumption aggregate C_t as a function of the varieties $\{c_t(z)\}_{z=0}^1$, to

$$\int_0^1 S(z) D\left(\frac{c_t(z)}{S(z)a_t(z)C_t}\right) dz = 1, \tag{49}$$

where we use the same aggregator $D(x)$ that we used before. The household's cost-minimization problem leads to the following demand for good z, which is a generalization to (7):

$$c_t(z) = a_t(z)S(z)d\left(\frac{p_t(z)a_t(z)}{P_t\theta_t}\right) C_t, \tag{50}$$

where $d(x)$ is defined as in the main model (see (5)). Thus $S(z)$ shifts the demand for firm z's good but does not affect the elasticity of demand for a given price $p_t(z)$ and quality measure $a_t(z)$.

The equations determining P_t and θ_t, for given $\{p_t(z)\}_{z=0}^1$, $\{a_t(z)\}_{z=0}^1$, and $\{S(z)\}_{z=0}^1$ are straightforward generalizations to (24) and (25):

$$\int_0^1 S(z) D\left(d\left(\frac{a_t(z)p_t(z)}{P_t\theta}\right)\right) dz = 1, \tag{51}$$

$$\int_0^1 \frac{a_t(z)S(z)p_t(z)}{P_t} d\left(\frac{a_t(z)p_t(z)}{P_t\theta_t}\right) dz = 1. \tag{52}$$

It is also useful to scale firm z's fixed cost by parameter $S(z)$ such that, for given price $p_t(z)$ and $a_t(z)$, the ratio of fixed costs over output is unaffected by $S(z)$. The production function (26) is thus modified to

$$y_t(z) = a_t(z)\left[A(z)\big(M_t(z)\big)^{s_m} L_t(z)^{1-s_m} - S(z)FC(z)\right]. \tag{53}$$

It is straightforward to show that, by adjusting the size parameters $\{S(z)\}_{z=0}^1$ and adapting α conformably to keep the relative sizes of the L sector and the H sector constant, any steady-state distribution of firm sizes can be attained. Importantly, the resulting model is equivalent in the sense that it delivers the same results as our main model. As a consequence, our analysis is compatible with any given distribution of firm sizes.

[13] Analogously, Baqaee et al. (2023) introduce "taste shifters" into their model.

It may be worth noting that the model variant with size shifters $S(z)$ requires a minor re-interpretation of θ. θ no longer measures dispersion in quality-adjusted output $c(z)/A(z)$ but differences in $c(z)/(A(z)S(z))$.

7 Conclusions

This paper has proposed a menu-cost model with endogenous markups, in which increases in productivity for a small group of firms lead to increases in average markups while the median markup remains largely constant. It has highlighted an amplification mechanism that has not been studied by the literature so far. Changes in market concentration that occur over the business cycle have a direct impact on the demand for individual goods and the elasticity of demand in particular. These changes in market concentration lead to endogenous movements in aggregate TFP and influence the transmission of monetary shocks.

We have also introduced a new variant of the Kimball aggregator, which is relatively simple as it is characterized by only two parameters yet sufficiently rich to be compatible with empirically plausible changes in markups. Another advantage of our specification is that the two parameters have straightforward interpretations as the super-elasticity of demand as well as the demand elasticity of a firm with an intermediate level of output.

A Optimal Price Implied by (15)

Equation (15) can be rewritten as

$$f\left(q(z), mc(z)\right) = 0, \tag{54}$$

where

$$f\left(q(z), mc(z)\right) := \varepsilon \left(q(z)\right)^{s-1} \left(q(z) - mc(z)\right) - \theta^s. \tag{55}$$

As $f(q(z), mc(z))$ is a continuous function of $q(z)$, $f(mc(z), mc(z)) = -\theta^s$, and $\lim_{q(z)\to\infty} f(q(z), mc(z)) = \infty$, the intermediate value theorem implies that, for a given level of $mc(z)$, (54) has a solution for $q(z)$ that satisfies $q(z) > mc(z)$. The derivative of $f(q(z), mc(z))$ with regard to $q(z)$ is

$$\begin{aligned} \frac{\partial f\left(q(z), mc(z)\right)}{\partial q(z)} &= \varepsilon \left[s \left(q(z)\right)^{s-1} - mc(z)(s-1) \left(q(z)\right)^{s-2}\right] \\ &= \varepsilon \left[sq(z) - mc(z)(s-1)\right] \left(q(z)\right)^{s-2}, \end{aligned} \tag{56}$$

This expression is strictly positive for $q(z) > mc(z)$, which entails that (54) has a unique solution for $q(z)$.

Next we show that the optimal relative price $q(z)$ is an increasing function of $mc(z)$. Computing the total derivative for Equation (54) results in

$$\varepsilon \left[s \left(q(z)\right)^{s-1} \frac{dq(z)}{dmc(z)} - (s-1) \left(q(z)\right)^{s-2} mc(z) \frac{dq(z)}{dmc(z)} - \left(q(z)\right)^{s-1}\right] = 0. \tag{57}$$

Rearranging yields

$$\frac{dq(z)}{dmc(z)} = \frac{q(z)}{sq(z) - (s-1)mc(z)}. \tag{58}$$

Because $q(z) > mc(z)$, this derivative is positive. To sum up, for every value of $mc(z)$, there is a unique optimal price $q(z)$, which satisfies $q(z) > mc(z)$. The optimal price increases with the firm's marginal cost $mc(z)$. □

B Proof of Proposition 1

Without loss of generality, we set $C = 1$. We note that c_1, c_2 and θ have to satisfy:

$$\nu D\left(c_1\right) + (1-\nu) D\left(c_2\right) = 1, \tag{59}$$

$$\left[\nu D'\left(c_1\right) c_1 + (1-\nu) D'\left(c_2\right) c_2\right] \theta = \varepsilon^{\frac{1}{s}}, \tag{60}$$

where Equation (59) follows from (1) and Equation (60) can be obtained from the first-order condition for the household's cost-minimization problem and the definition of θ in (6).

Computing the total differential for (59) and (60) yields

$$\begin{aligned} \nu D'(c_1)\, dc_1 + (1-\nu)D'(c_2)\, dc_2 &= 0, && (61)\\ \nu\left[D'(c_1) + D''(c_1)\, c_1\right] dc_1 + (1-\nu)\left[D'(c_2) + D''(c_2)\, c_2\right] dc_2 &= -\frac{1}{\theta^2}\, d\theta. && (62) \end{aligned}$$

As $D(x)$ is a strictly monotonically increasing function, it is clear from (61) that dc_1 and dc_2 have opposite signs. This is intuitively clear as e.g. a lower level of consumption of goods from the first group, c_1, requires an increase in c_2 in order to keep the consumption aggregate constant.

Combining (61) and (62) yields

$$\nu D''(c_1)\, c_1\, dc_1 + (1-\nu)D''(c_2)\, c_2 dc_2 = -\frac{1}{\theta^2}\, d\theta. \tag{63}$$

With the help of (2), it is straightforward to verify $D''(x) = -\frac{1}{x}\left(D'(x)\right)^{1-s}$. Using this identity and (63), we obtain

$$\nu\left(D'(c_1)\right)^{1-s} dc_1 + (1-\nu)\left(D'(c_2)\right)^{1-s} dc_2 = \frac{1}{\theta^2}\, d\theta. \tag{64}$$

We can use (61) to eliminate dc_1 in (64), which results in

$$(1-\nu)\left(D'(c_2)\right)\left[\left(D'(c_2)\right)^{-s} - \left(D'(c_1)\right)^{-s}\right] dc_2 = \frac{1}{\theta^2}\, d\theta. \tag{65}$$

Because (i) $D(x)$ is strictly concave, (ii) $s > 0$, and (iii) $c_2 > c_1$, the expression in brackets is positive. Thus an increase in c_2 ($dc_2 > 0$) yields an increase in θ ($d\theta > 0$). We have already noted that an increase in c_2 necessitates a decrease in c_1 in order to keep aggregate consumption constant.

It remains to examine the symmetric case where $c_1 = c_2 = 1$. First, we note that this case is compatible with (59). Second, we can use (60) to compute θ as

$$\theta = \frac{\varepsilon^{\frac{1}{s}}}{D'(1)} = 1. \tag{66}$$

□

C Derivation of (38)

According to $L_t = \int_0^1 L_t(z)\,dz$, $M_t = \int_0^1 M_t(z)\,dz$, (26), and (32), which entails $L_t(z)/M_t(z) = L_t/M_t$, we have

$$\begin{aligned} L_t &= \int_0^1 L_t(z)\,dz \\ &= \int_0^1 \frac{y_t(z) + a_t(z)FC(z)}{A(z)a_t(z)} \left(\frac{L_t(z)}{M_t(z)}\right)^{s_m} dz \\ &= \int_0^1 \frac{y_t(z) + a_t(z)FC(z)}{A(z)a_t(z)} \left(\frac{L_t}{M_t}\right)^{s_m} dz. \end{aligned} \tag{67}$$

Together with

$$y_t(z) = a_t(z)Y_t d\left(\frac{q_t(z)}{\theta_t}\right), \tag{68}$$

which follows from (29) and (30), this can be rearranged to yield (38). □

D Derivation of (17)

Equation (16) can be rearranged as

$$\ln q(z) = \ln \epsilon(z) - \ln(\epsilon(z) - 1) + \ln mc(z). \tag{69}$$

Thus we obtain

$$\begin{aligned} \frac{d\ln q(z)}{d\ln mc(z)} &= -\frac{1}{\epsilon(z)(\epsilon(z)-1)} \cdot \frac{d\epsilon(z)}{d\ln mc(z)} + 1 \\ &= -\frac{1}{\epsilon(z)-1} \cdot \frac{d\ln\epsilon(z)}{d\ln mc(z)} + 1. \end{aligned} \tag{70}$$

With the help of (10), we get

$$\frac{d\ln q(z)}{d\ln mc(z)} = -\frac{s}{\epsilon(z)-1} \cdot \frac{d\ln q(z)}{d\ln mc(z)} + 1. \tag{71}$$

Solving for $\frac{d\ln q(z)}{d\ln mc(z)}$ yields (17). □

E Variant with Stochastic Menu Costs

In this Appendix, we study a variant of our model with stochastic menu costs. As mentioned in the main text, we assume that fifty percent of all price changes involve zero cost. Otherwise the model is calibrated in the same way as the model with pure menu costs. The parameter values that are chosen based on external evidence are the same as before (see Table 1). Tables 9 and 10 replicate Tables 4 and 5 for stochastic menu costs. One can readily verify that the findings about the real effects of nominal disturbances as well as price dynamics are very similar (see Table 9). In particular, increases in market power lead to higher degrees of monetary non-neutrality. In addition, the endogenous fluctuations of aggregate productivity are of a similar magnitude as in our main model. Overall, the stochastic-menu-cost model involves somewhat larger levels of monetary non-neutrality, which is plausible as the possibility to adjust prices without cost weakens the selection effect and thereby leads to larger real effects of nominal shocks.

	early period	late period A	late period B
size of price changes	12.1%	12.2%	19.0%
freq. of price changes	15.6%	15.4%	14.3%
size of price changes (L)	11.8%	11.7%	18.5%
size of price changes (H)	13.3%	13.3%	20.7%
freq. of price changes (L)	16.7%	16.7%	15.3%
freq. of price changes (H)	11.8%	11.6%	11.2%
size of L sector	78.3%	74.5%	75.7%
mean elasticity	6.7	6.5	5.9
std. dev. $\hat{C}_t$	0.0032	0.0026	0.0026
monetary non-neutrality	1.08	1.14	1.16
TFP change in resp. to $\hat{C}$ change	6.6%	8.5%	7.2%
ch. of log theta in resp. to $\hat{C}$ ch.	4.6%	5.9%	4.7%
steady-state value of theta	1.068	1.088	1.082

Table 9: Simulation outcomes for stochastic menu costs. Mean values weighted by shares of total revenues.

Table 10 shows that our amplification mechanism, which relies on fluctuations of θ, is relevant for stochastic menu costs as well. Shutting off this mechanism by keeping θ constant leads to a markedly weaker misallocation channel, i.e. smaller changes in endogenous TFP over the business cycle, as well as a lower degree of monetary non-neutrality.

□

	late period	late period (θ_t fixed)
size of price changes	19.0%	19.0%
freq. of price changes	14.3%	14.3%
size of price changes (L)	18.5%	18.5%
size of price changes (H)	20.7%	20.7%
freq. of price changes (L)	15.3%	15.3%
freq. of price changes (H)	11.2%	11.2%
size of L sector	75.7%	75.7%
mean elasticity	5.9	5.9
std. dev. $\hat{C}_t$	0.0026	0.0023
monetary non-neutrality	1.16	1.00
TFP change in resp. to $\hat{C}$ change	7.2%	1.1%

Table 10: Relevance of the new amplification mechanism for the variant with stochastic menu costs. The column in the middle reproduces results from Tables 2 and 4 for the late period. The column on the right-hand side shows the respective results from a simulation where θ_t is held fixed at its steady-state value.

References

Adam, K. and H. Weber (Sept. 2019). "Price Trends over the Product Life Cycle and the Optimal Inflation Target". Bundesbank Discussion Paper 32/2019.

Amiti, M., O. Itskhoki, and J. Konings (2019). "International Shocks, Variable Markups, and Domestic Prices". *Review of Economic Studies* 86(6), pp. 2356–2402.

Aquilante, T., S. Chowla, N. Dacic, A. Haldane, R. Masolo, P. Schneider, M. Seneca, and S. Tatomir (May 2019). "Market Power and Monetary Policy". Bank of England Working Paper #798.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2017). "Concentrating on the Fall of the Labor Share". *American Economic Review (Papers and Proceedings)* 107(5), pp. 180–85.

Bai, Y., J.-V. Ríos-Rull, and K. Storesletten (2019). "Demand Shocks as Productivity Shocks". Manuscript.

Ball, L. and D. Romer (1990). "Real Rigidities and the Non-neutrality of Money". *Review of Economic Studies* 57(2), pp. 183–203.

Baqaee, D., E. Farhi, and K. Sangani (Feb. 2023). "The Supply-side Effects of Monetary Policy". *Journal of Political Economy*. forthcoming.

Basu, S. (1995). "Intermediate Goods and Business Cycles: Implications for Productivity and Welfare". *American Economic Review*, pp. 512–531.

Beck, G. W. and S. M. Lein (2020). “Price Elasticities and Demand-side Real Rigidities in Micro Data and in Macro Models”. *Journal of Monetary Economics* 115, pp. 200–212.

Berger, D. and J. Vavra (2019). “Shocks versus Responsiveness: What Drives Time-varying Dispersion?” *Journal of Political Economy* 127(5), pp. 2104–2142.

Bergin, P. R. and G. Corsetti (2008). “The Extensive Margin and Monetary Policy”. *Journal of Monetary Economics* 55(7), pp. 1222–1237.

Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2007). “Monetary Policy and Business Cycles with Endogenous Entry and Product Variety”. *NBER Macroeconomics Annual* 22, pp. 299–379.

Bils, M., P. J. Klenow, and B. A. Malin (2018). “Resurrecting the Role of the Product Market Wedge in Recessions”. *American Economic Review* 108(4-5), pp. 1118–1146.

Blanco, A. (2021). “Optimal Inflation Target in an Economy with Menu Costs and a Zero Lower Bound”. *American Economic Journal: Macroeconomics* 13(3), pp. 108–141.

Blinder, A., E. R. Canetti, D. E. Lebow, and J. B. Rudd (1998). *Asking About Prices: A New Approach to Understanding Price Stickiness*. Russell Sage Foundation, New York.

Broer, T., N.-J. Harbo Hansen, P. Krusell, and E. Öberg (2020). “The New Keynesian transmission mechanism: A heterogeneous-agent perspective”. *Review of Economic Studies* 87(1), pp. 77–101.

Carvalho, C. and O. Kryvtsov (2021). “Price Selection”. *Journal of Monetary Economics* 122, pp. 56–75.

Colciago, A. and R. Silvestrini (Feb. 2021). “Monetary Policy, Productivity, and Market Concentration”. Manuscript.

Cooke, D. and T. Damjanociv (2021). “Monetary Policy and Welfare with Heterogeneous Firms and Endogenous Entry”. Durham Business Scholl Working Paper No. 2.

De Loecker, J., J. Eeckhout, and G. Unger (2020). “The Rise of Market Power and the Macroeconomic Implications”. *Quarterly Journal of Economics* 135(2), pp. 561–644.

De Ridder, M. (Oct. 2021). “Market Power and Innovation in the Intangible Economy”. Manuscript.

Diez, M. F., M. D. Leigh, and S. Tambunlertchai (2018). “Global Market Power and its Macroeconomic Implications”. IMF Working Paper #18/137.

Dotsey, M. and R. G. King (2005). “Implications of State-dependent Pricing for Dynamic Macroeconomic Models”. *Journal of Monetary Economics* 52(1), pp. 213–242.

Duval, R., D. Furceri, R. Lee, and M. M. Tavares (July 2021). “Market Power and Monetary Policy Transmission”. IMF Working Paper #21/184.

Edmond, C., V. Midrigan, and D. Y. Xu (2023). “How costly are markups?” *Journal of Political Economy* 131(7), pp. 1619–1675.

Ganapati, S. (2021). “Growing Oligopolies, Prices, Output, and Productivity”. *American Economic Journal: Microeconomics* 13(3), pp. 309–327.

González, B., G. Nuño Barrau, D. Thaler, and S. Albrizio (2021). “Firm Heterogeneity, Capital Misallocation and Optimal Monetary Policy”. Bank of Spain Working Paper #2145.

Hahn, V. and M. Marenčák (2020). “Price Points and Price Dynamics”. *Journal of Monetary Economics* 115, pp. 127–144.

Harding, M., J. Lindé, and M. Trabandt (Mar. 2022). “Resolving the Missing Deflation Puzzle”. *Journal of Monetary Economics* 126, pp. 15–34.

Hazell, J., J. Herreno, E. Nakamura, and J. Steinsson (2020). “The Slope of the Phillips Curve: Evidence from US states”. NBER Working paper #28005.

Kimball, M. (1995). “The Quantitative Analytics of the Basic Neomonetarist Model”. *Journal of Money, Credit, and Banking* 27(4), pp. 1241–1277.

Klenow, P. J. and J. L. Willis (2016). “Real Rigidities and Nominal Price Changes”. *Economica* 83(331), pp. 443–472.

Krusell, P. and A. A. Smith Jr (1998). “Income and Wealth Heterogeneity in the Macroeconomy”. *Journal of Political Economy* 106(5), pp. 867–896.

Levy, D., M. Bergen, S. Dutta, and R. Venable (1997). “The Magnitude of Menu Costs: Direct Evidence from Large US Supermarket Chains”. *Quarterly Journal of Economics* 112(3), pp. 791–824.

Lewis, V. and C. Poilly (2012). “Firm Entry, Markups and the Monetary Transmission Mechanism”. *Journal of Monetary Economics* 59(7), pp. 670–685.

Lindé, J., F. Smets, and R. Wouters (2016). “Challenges for Central Banks’ Macro Models”. In: *Handbook of Macroeconomics*. Vol. 2. Elsevier, pp. 2185–2262.

Meier, M. and T. Reinelt (July 2022). “Monetary Policy Markup Dispersion and Aggregate TFP”. *Review of Economics and Statistics*. forthcoming.

Midrigan, V. (2011). “Menu Costs, Multiproduct Firms, and Aggregate Fluctuations”. *Econometrica* 79(4), pp. 1139–1180.

Mishkin, F. S. (2007). “Inflation Dynamics”. *International Finance* 10(3), pp. 317–334.

Mongey, S. (2021). “Market Structure and Monetary Non-neutrality”. NBER working paper #29233.

Nakamura, E. and J. Steinsson (2010). “Monetary Non-neutrality in a Multisector Menu Cost Model”. *Quarterly Journal of Economics* 125(3), pp. 961–1013.

Nakamura, E., J. Steinsson, P. Sun, and D. Villar (Nov. 2018). “The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation”. *Quarterly Journal of Economics* 4(1), pp. 1933–1980.

Nekarda, C. J. and V. A. Ramey (2020). “The Cyclical Behavior of the Price-Cost Markup”. *Journal of Money, Credit, and Banking* 52(S2), pp. 319–353.

Smets, F. and R. Wouters (2007). “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach”. *American Economic Review* 97(3), pp. 586–606.

Wang, O. and I. Werning (2020). “Dynamic Oligopoly and Price Stickiness”. NBER Working Paper #27536.