Income Dispersion and Counter-Cyclical Markups

Chris Edmond and Laura Veldkamp*
New York University
Stern School of Business

July 24, 2007

Abstract

We construct a model of counter-cyclical markups based on cyclical variation in the dispersion of income across agents. The model is neoclassical in most respects, with monopolistically competitive firms facing a distribution of buyers that changes through time. Income dispersion is high during recessions, which reduces the price elasticity of demand and increases markups applied by firms. Using recent estimates of counter-cyclical income dispersion, we calibrate the model and show that it generates realistic markups as well as other salient features of business cycles.

Keywords: business cycles, counter-cyclical markups, income dispersion.

JEL classifications: E32.

*We thank George Alessandria, David Backus, Mark Bils, Mark Gertler, Aubhik Khan, John Leahy, Nick Souleles, Harald Uhlig, Lawrence Uren, Stijn Van Nieuwerburgh, Pierre-Olivier Weill, Michael Woodford, Randy Wright, and seminar participants at Tokyo, Princeton, UCLA, NYU, Philadelphia FRB, Rochester, Iowa, Oslo, Melbourne and the 2007 AEA and 2006 SED meetings for helpful comments and conversations. We especially thank Jeff Campbell whose detailed suggestions greatly improved the paper. Laura Veldkamp also thanks Princeton University for their hospitality and financial support through the Kenen fellowship.

A long line of empirical research suggests that the prices vary less over the business cycle than marginal costs. In other words, markups are counter-cyclical. The question is why. We argue that the cross-section dispersion of earnings might play a role. In recessions, when earnings are more dispersed, buyers' willingness to pay is also more dispersed. If sellers were to reduce prices in recessions, they would attract few additional buyers (the small shaded area in the left panel of figure 1). This low elasticity makes the marginal benefit of lowering prices smaller and induces firms to keep prices high. Therefore when dispersion is high, prices stay high but profits are low. In contrast, in booms when dispersion is low, a seller who lowers her price attracts many additional buyers (the larger shaded area in the right panel of figure 1). Therefore in booms, sellers keep prices low but earn high profits.

While there have been many previous explanations for counter-cyclical markups, the mechanism we propose has two strengths: It is based on observables and is simple enough to embed in a conventional business cycle model.[1] The observable variable is earnings dispersion. Embedding the earnings process estimated by Storesletten, Telmer, and Yaron (2004) in a production economy allows us to compare the model's predictions to salient business cycle aggregates. In particular, we deliver realistic pro-cyclical profits, a feature of the data that many models struggle with. Beyond business cycle facts, the model can also explain long-run trends and cross-state variations in markups.

To illustrate the workings of the model's key mechanisms, section 1 analyzes a static version of the model. There is a competitive sector and a monopolistically competitive sector that produces differentiated products. Both produce with intermediate goods, whose only input is effective labor. Agents choose how much to work, how much of the competitive good to consume and how many of the differentiated products to buy. Income dispersion arises

[1]Seminal papers on counter-cyclical markups are Rotemberg and Saloner (1986) and Bils (1989). For a review, see Rotemberg and Woodford (1999). Recent work on the related phenomenon of real price rigidity includes Nakamura and Steinsson (2006) and Menzio (2007).

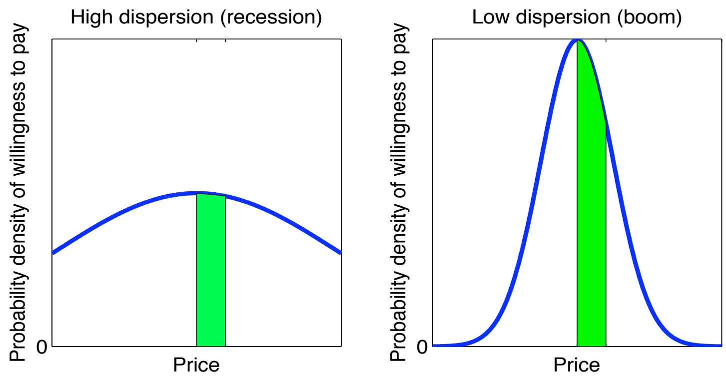

Figure 1: Lowering price is more beneficial when dispersion is low. The shaded area represents the increase in the probability of trade from lowering the price, by an amount equal to the width of the shaded area. This higher probability, times the expected gains from trade, is the marginal benefit to reducing the price. Willingness to pay is based on agents' earnings.

because some agents are more productive. By manipulating the idiosyncratic productivity distribution, we we show that more dispersion results in higher markups and higher prices.

Theory alone cannot tell us if the variation in earnings dispersion is too small or insufficiently counter-cyclical to generate observed markups. Therefore section 2 calibrates and simulates a dynamic version of the model. Section 2.4 documents our main results: The model's optimal markups are determined primarily by earnings dispersion. Since measured dispersion is counter-cyclical, markups are as well. This is consistent with empirical studies in both macroeconomics and industrial organization. The resulting prices look inflexible because they fluctuate less than marginal cost. Yet, there are no price-setting frictions.

The model explains markups without compromising its ability to match macroeconomic aggregates. Section 2.5 compares the model's predictions for employment, real wages, and profits to their empirical counterparts. The model does reasonably well in capturing these aggregates. To keep heterogeneous income processes tractable, our model abstracts from important issues debated in the literature on income heterogeneity and welfare (Krusell and Smith (1998), Rios-Rull (1996) and Krueger and Perri (2005b)), such as capital accumulation or other consumption-savings behavior. Tractability comes at a cost: Omitting capital hurts the performance of the model by making aggregates too correlated with GDP.

A number of other mechanisms can generate counter-cyclical markups. One possibility is that there are simply sticky prices and pro-cyclical marginal costs. Therefore, the difference between price and cost, the markup, is counter-cyclical. The problem with this explanation is that, without additional labor market frictions, it implies counter-cyclical firm profits, strongly at odds with the data. Similarly, while firm entry and exit change the degree of market competition and thus the markup (Jaimovich 2006), free entry implies zero profits. Our model delivers the observed pro-cyclical profits. Booms are times when markups are low but volume is high enough to compensate. In Comin and Gertler (2006), the causality is reversed: They use shocks to markups as the source of business cycle fluctuations. Three models closely related to ours also produce a cyclical elasticity of demand due to changing production technology (Kimball 1995), changing demand composition (Gali 1994), or a change in product variety (Bilbiie, Ghironi, and Melitz 2006).

To argue that earnings dispersion is at least part of the reason for price variation, we look for other evidence that long-run changes and cross-sectional differences in earnings dispersion are correlated with differences in prices, output volatility and profit shares, as predicted by the model. Section 3 shows that the observed increase in earnings dispersion is consistent with the observed decline in business cycle volatility, the slow-down in real wage growth, and the accompanying increase in profit shares. Section 4.1 uses state-level panel data to test the model's predicted relationships between earnings dispersion and prices. Section 4.2 documents additional facts from the empirical pricing literature that when the customer base has a more dispersed earnings, prices tend to be higher.

Our explanation raises an obvious question: Why does earnings dispersion rise in recession? One explanation is that job destruction in recessions is responsible (Caballero and Hammour 1994). Rampini (2004) argues that entrepreneurs' incentives change in recessions, making firm outcomes and owners' earnings more risky. Cooley, Marimon, and Quadrini (2004) and Lustig and Van Nieuwerburgh (2005) argue that low collateral values inhibit risk-sharing in recessions. Any one of these explanations could be merged with this model to produce a model whose only driving process is aggregate technology shocks.

1 An illustrative static model

There is a continuum of agents indexed by $i \in [0, 1]$, with identical preferences over a numeraire consumption good c_i, labor n_i, and a continuum of differentiated products x_{ij}, indexed by $j \in [0, 1]$,

$$U_i = \log(c_i) + \theta(1 - n_i) + \nu \int_0^1 x_{ij} dj, \quad \theta, \nu > 0. \tag{1}$$

Each of the differentiated products is indivisible.[2] An agent either buys good j or not, $x_{ij} \in \{0, 1\}$. But the total quantity of x goods consumed can be adjusted by buying more or fewer goods. Let p_j denote the price of differentiated good j in terms of the numeraire. Then the budget constraint facing agent i is,

$$c_i + \int_0^1 p_j x_{ij} dj \leq w_i n_i + \pi, \tag{2}$$

where π denotes lump-sum profits paid out by firms.

Heterogeneous labor productivity is the source of earnings dispersion. Labor productivity is IID uniform: $w_i \sim \text{unif}[z - \sigma, z + \sigma]$, with $z > \sigma > 0$ so that $w_i > 0$ all i. The distribution of productivity is summarized by its mean z and a measure of dispersion, σ.

[2] Following Kiyotaki and Wright (1989), we assume that the imperfectly competitive x-good is indivisible. We do this both because it makes the model easy to solve and because it reduces sellers' marker power by preventing them from using complicated non-linear pricing strategies. Similar indivisibility assumptions are often made in search models of money.

Profit-maximizing firms transform effective labor 1-for-1 into numeraire goods or differentiated goods. Since aggregate effective labor is $\int_0^1 w_i n_i di$, the labor market clears when

$$\int_0^1 c_i di + \int_0^1 x_j dj = \int_0^1 w_i n_i di, \tag{3}$$

where $x_j := \int_0^1 x_{ij} di$ is the total demand for good j.

Firms choose prices and quantities to maximize profit. Let π_j denote the profits of firm j. Profits are price p_j times aggregate amount sold $x(p_j)$ minus cost,

$$\pi_j = (p_j - 1)x(p_j). \tag{4}$$

Since competitive firms make zero profits, aggregate profits are the integral of all x-good firm profits, $\pi := \int_0^1 \pi_j dj$. Each household gets an equal share of these aggregate profits.

Equilibrium An equilibrium in this economy is: (i) a set of consumption choices c_i and x_{ij} and labor supply choices n_i for each household that maximize utility (1) subject to the budget constraint (2), (ii) a price p_j for each firm that maximizes profit (4) taking as given the demand for the firm's product such that (iii) the markets for c-goods, x-goods, and labor (3) all clear.

Results Optimal consumption of differentiated products x_{ij} follows a cutoff rule, agent i buys good j if the additional utility it provides exceeds the price p_j times the agent's Lagrange multiplier on (2), i.e., if $\nu \geq p_j \lambda_i$. The first order condition for labor supply tells us that $\lambda_i = \theta / w_i$. Combining these two expressions yields the x-good consumption rule,

$$x_{ij} = \begin{cases} 1 & \text{if } w_i \geq \frac{\theta}{\nu} p_j \\ 0 & \text{otherwise} \end{cases} . \tag{5}$$

The fraction of agents who buy a differentiated product is just the probability that each agent has a labor productivity higher than the cutoff value,

$$x(p_j) := \int_0^1 x_{ij} di = \int_{\theta p_j/\nu}^{z+\sigma} \frac{1}{2\sigma} dw_i = \frac{z+\sigma}{2\sigma} - \frac{\theta/\nu}{2\sigma}. \tag{6}$$

The demand facing firm j is linear in its price p_j. Substituting the demand curve (6) into the profit function (4) and differentiating with respect to p_j yields the first order condition characterizing the profit-maximizing price,

$$p_j + \frac{x(p_j)}{x'(p_j)} = 1. \tag{7}$$

The left hand side is the firm's marginal revenue, the right hand side its constant marginal cost. Equivalently, the price set by firm j is a markup over the marginal cost of 1, $p_j = \epsilon(p_j)/(\epsilon(p_j) - 1)$, where $\epsilon(p_j) := -x'(p_j)p_j/x(p_j)$ is the firm's price elasticity of demand. Rearranging (6) delivers the markup,

$$m_j := \frac{p_j}{1} = 1 + \frac{1}{2}\left(\frac{z+\sigma}{\theta/\nu} - 1\right). \tag{8}$$

Firms only produce if they earn non-negative profits, which is when $m_j \geq 1$. To ensure that $m_j \geq 1$, we assume that marginal cost is sufficiently low: $1 \leq (z+\sigma)\nu/\theta$. If this assumption were violated, no firm would produce.

Result. *The markup m_j is strictly increasing in aggregate productivity z and dispersion σ.*

This result follows immediately from differentiating (8) with respect to z and σ. [3] The

[3] A previous version of the paper proved this result holds as long as idiosyncratic productivity is non-negative and its distribution is log-concave. We thank Jeff Campbell for pointing this result out to us.

intuition for the result is that both variables decrease the elasticity of demand,

$$\epsilon(p_j) = \frac{\frac{z+\sigma}{\theta/\nu} + 1}{\frac{z+\sigma}{\theta/\nu} - 1}. \tag{9}$$

When elasticity rises, firms reduce prices because doing so results in many additional sales.

If business cycles involved only changes in productivity, then markups would be procyclical. That is counter-factual. But if dispersion is counter-cyclical, so that σ falls when z rises, then markups may be counter-cyclical. To determine if dispersion is sufficiently counter-cyclical to explain counter-cyclical markups, section 2 builds a dynamic quantitative model.

2 A dynamic quantitative model

Our dynamic model departs from the static model in four ways. First, aggregate productivity z and dispersion σ fluctuate. Second, the distribution of idiosyncratic labor productivity is log-normal, not the simple uniform distribution we used in section 1 for illustrative purposes. Third, marginal cost is variable instead of constant, so that firm profit shares are realistic. Fourth, richer preferences deliver a more realistic wealth effect on the labor supply.

In the model, profits rise in booms. With a strong wealth effect on labor, cyclical profits can make labor counter-cyclical. Although other models encounter this problem, it is particularly acute here because imperfect competition in x goods makes profits larger and more volatile. We use "GHH" preferences (Greenwood, Hercowitz, and Huffman 1988) that eliminate the wealth effect on labor supply to deliver more realistic labor fluctuations.

We omit capital and other assets to keep the model tractable. Since agents have no opportunity to share risk or smooth consumption, such assumptions could distort the results. As a robustness check, we gauge the effect of this distortion by re-calibrating the model to match consumption data, which incorporates the effect of financial income, savings and transfers.

2.1 Model setup

Individuals again have preferences over c_i, n_i and x_{ij} but the utility function is now

$$U_i = \log\left(c_i - \theta\frac{n_i^{1+\gamma}}{1+\gamma}\right) + \nu\int_0^1 x_{ij}dj, \tag{10}$$

which they maximize subject to their budget constraint (2).

The log of aggregate productivity is an AR(1) process

$$\log(z_t) = (1-\rho)\log(\bar{z}) + \rho\log(z_{t-1}) + \varepsilon_{zt}, \qquad \varepsilon_{zt} \sim \mathcal{N}(0,\sigma_z^2). \tag{11}$$

Idiosyncratic labor productivity is log-normal, $\log(w_{it}) = \log(z_t) + \varepsilon_{it}$ where $\varepsilon_{it} \sim \mathcal{N}(0,\sigma_t^2)$. Our model of idiosyncratic productivity follows Storesletten, Telmer, and Yaron (2004) who estimate an earnings process with persistent and transitory shocks. Let

$$\begin{array}{rclcll} \varepsilon_{it} & = & \xi_{it} & + & u_{it}, & u_{it} \sim \mathcal{N}(0,\sigma_u^2) \\ \xi_{it} & = & \rho_\xi \xi_{it-1} & + & \eta_{it}, & \eta_{it} \sim \mathcal{N}(0,\sigma_{\eta t}^2). \end{array} \tag{12}$$

The key feature of the earnings process is that $\sigma_{\eta,t}^2$ increases when GDP is below its long-run mean, specifically $\sigma_{\eta,t}^2 = \sigma_H^2$ if $y_t \geq \overline{y}$ and $\sigma_{\eta,t}^2 = \sigma_L^2$ if $y_t < \overline{y}$, where $\sigma_H^2 < \sigma_L^2$, y_t is GDP, as defined in equation (15) below, and $\overline{y}$ is its long-run mean.

Putting these elements together, our stochastic process for dispersion is

$$\sigma_t^2 = \rho_\xi^2\sigma_{t-1}^2 + (1-\rho_\xi^2)\sigma_u^2 + \begin{cases} \sigma_H^2 & \text{if } y_t \geq \overline{y} \\ \sigma_L^2 & \text{if } y_t < \overline{y} \end{cases} . \tag{13}$$

Finally, we give x-good firms variable marginal costs. They transform effective labor n into x_j goods: $x_j = n^\alpha$, for $0 < \alpha < 1$. Aggregate effective labor is $\int_0^1 w_i n_i di$ and the labor market clears when $\int_0^1 c_i di + \int_0^1 x_j^{1/\alpha} dj = \int_0^1 w_i n_i di$. Profits for firm j are:

$$\pi_j = p_j x_j - x_j^{1/\alpha}, \tag{14}$$

with variable marginal cost $x_j^{(1-\alpha)/\alpha}/\alpha$.

Measuring GDP in the model In order to calibrate and evaluate the model, we need a measure of total value-added to compare to GDP in the data,

$$y := \int_0^1 c_i di + \int_0^1 \int_0^1 p_j x_{ij} dj di. \tag{15}$$

GDP varies both because of changes in the production of each good and because of changes in the relative price of x-goods and c-goods.

2.2 Model solution

The first-order condition for labor choice tells us that labor depends only on the wage and on preference parameters

$$n_i = \left(\frac{w_i}{\theta}\right)^{1/\gamma}. \tag{16}$$

This simple relationship, devoid of any wealth effect is what GHH preferences are designed to deliver. It implies that log earnings are proportional to log idiosyncratic productivity, $\log(n_i w_i) = (1+\gamma)\log(w_i)/\gamma - \log(\theta)/\gamma$. But GHH preferences complicate the model's solution because the cutoff rule for x-good demand is no longer linear in the wage. While agent i still buys a unit of x_j if $\nu \geq p_j \lambda_i$, the Lagrange multiplier on their budget constraint is now $\lambda_i = \left(c_i - \theta \frac{n_i^{1+\gamma}}{1+\gamma}\right)^{-1}$. To derive the x-good consumption rule, use (2) and (16) to substitute out c_i and n_i in the λ_i formula. Then, substitute λ_i into the cutoff rule at the indifference point ($p_j = \nu/\lambda_i$). This delivers a critical wage $\hat{w}(p_j)$ such that any agent with wage higher than this threshold buys the good. Thus the aggregate consumption of good j is $x(p_j) = \Pr[w_i \geq \hat{w}(p_j)]$.

Firms' prices are chosen to maximize profit (14), taking agents' demand functions as given. The first order condition for profit maximization equates marginal revenue and marginal cost,

$$p_j + \frac{x(p_j)}{x'(p_j)} = \frac{1}{\alpha} x(p_j)^{(1-\alpha)/\alpha}. \tag{17}$$

The set of equations that determine a solution to the model can no longer be solved in closed form. Appendix A.1 details the fixed point problem solved in the following numerical analysis.

2.3 Calibration

Parameter			Calibration target	
utility weight on leisure	θ	15	steady state hours	0.33
mean of productivity	$\bar{z}$	7.7	steady state profit share	0.3
concavity of production	α	0.24	steady state average markup	11%
utility weight on x-goods	ν	100	steady state x-sector markup	30%
inverse labor supply elasticity	γ	0.6	measured elasticity (GHH)	1.67
productivity innovation std dev	σ_z	0.0032	output std dev	0.017
productivity autocorrelation	ρ	0.80	output autocorrelation	0.78
transitory earnings std dev	σ_u	0.024	STY estimate (annual)	0.065
persistent earnings std dev $y > \bar{y}$	σ_H	0.012	STY estimate (annual)	0.032
persistent earnings std dev $y < \bar{y}$	σ_L	0.020	STY estimate (annual)	0.054
earnings autocorrelation	ρ_ξ	0.988	STY estimate (annual)	0.952

Table 1: PARAMETERS AND THE MOMENT OF THE DATA EACH PARAMETER MATCHES. Appendix A.1 derives the steady state moments of the model. Appendix A.3 details our transformation of STY moments from annual to quarterly.

Table 1 contains a summary of the calibrated parameters. Since average productivity and labor inputs determine firm profits, $\bar{z}$ and θ are chosen to match 33% of time spent working in steady state and a 30% profit share, both standard business cycle calibration targets. The second moments of the productivity process match the persistence and standard deviation of output as reported in Stock and Watson (1999).

Markups in the x-sector are defined as price divided by marginal cost,

$$m_x := \alpha \frac{p}{x^{(1-\alpha)/\alpha}}. \tag{18}$$

Estimates of markups vary widely, depending on the sector of the economy being measured. At the high end, Berry, Levinsohn, and Pakes (1995) and Nevo (2001) document markups of 27-45% for automobiles and branded cereals. For the macroeconomy as a whole, Chari, Kehoe, and McGrattan (2000) argue for a markup size of 11%. Since the competitive c-sector has zero markup, the markup for the economy as a whole is the x-sector markup m_x, times the x-sector expenditure share. Our calibration uses the concavity of production and the utility weight on x goods to match both the x-sector and the aggregate markup facts. Markups of 30% and 11% imply that the x-sector has an expenditure share of 37%. One caveat is that since α depends on a ratio of the markups, which are two small numbers, small changes in the markups do have big effects on the calibrated level of α. Therefore, appendix A.2 explores model results with higher and lower α's.

The relationship between earnings dispersion and output is not something we can manipulate directly because both earnings and GDP are endogenous variables. Therefore, we use the data to craft a process for exogenous variables – aggregate and idiosyncratic labor productivity – that produces endogenous series that fit the data. Because log labor is proportional to log productivity (equation 16), wage dispersion and log output have the same correlation as do productivity dispersion and log output. However, earnings ($w_i n_i$) has a higher dispersion because an individual's labor supply is positively correlated with their productivity. Earnings dispersion is higher by a factor of $(1+\gamma)/\gamma = 1.60/0.60 = 2.67$.

Therefore, our idiosyncratic productivity parameters are the STY estimates, transformed from yearly to quarterly, divided by 2.67. (See appendix A.3 for details.)

To determine whether the economy is in the high or low dispersion state (σ_H or σ_L), we first simulate the model in one state and then check whether GDP is higher or lower than its steady-state level. If realized GDP is inconsistent with the dispersion state, we re-simulate with the correct dispersion parameter. This process does not guarantee that dispersion and GDP have the same correlation as in the data, but the resulting correlations are quite close: -0.29 in the model and -0.30 in the data.

Issues in measuring dispersion A big question is whether earnings dispersion is the right measure of inequality to compare with the model. Either income, including capital income and transfers, or consumption are arguably more appropriate. Labor earnings are only 63% of income for the average household; yet, earnings and income dispersion have remarkably similar levels and cross-sectional variation (Diaz-Gimenez, Quadrini, and Rios-Rull 1997). The primary motivation for calibrating to earnings dispersion is the availability of reliable estimates of its cyclical properties. To measure dispersion in a number of business cycles requires a panel data set with a long time-series dimension. Storesletten, Telmer, and Yaron (2004) overcome this problem by using age characteristics of the PSID respondents to construct synthetic earnings data back to 1930. They do this same estimation with food consumption data. Because people use savings to smooth earnings fluctuations, consumption dispersion is considerably smaller. Appendix A.2 shows that using consumption dispersion instead of income dispersion strengthens our main result.

Storesletten, Telmer, and Yaron (2004)'s estimates have been controversial, because of the difficulty identifying transitory and permanent shocks. Guvenen (2005) and others argue that, because of unmeasured permanent differences in earnings profiles, the persistence of earnings shocks is overestimated. While this distinction is crucial in a consumption-savings

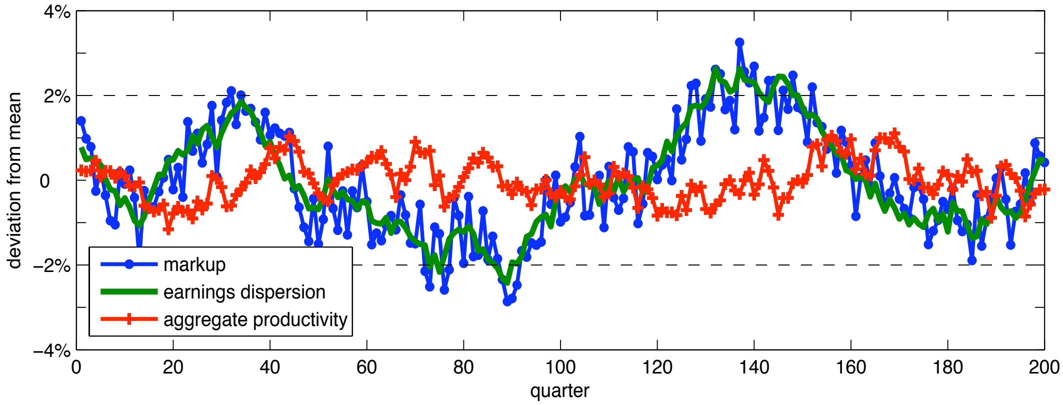

Figure 2: SIMULATED MARKUPS, EARNINGS DISPERSION AND PRODUCTIVITY.

problem, it is not relevant for our model. Whether earnings dispersion is persistent because each person gets persistent shocks or because new workers with more dispersed characteristics enter the sample — this does not matter to our seller who sets the price facing a distribution of willingness to pay. Thus both sides in this debate hold views consistent with our model's predictions.

2.4 Results: counter-cyclical markups

Recessions are times when firms pursue low-volume, high-margin sales strategies. The correlation of markups and log GDP in the simulated model is -0.20. Thus, markups are counter-cyclical. The standard deviation of detrended markups is 1.5% in the model and 2.1% in the data (Gali, Gertler, and Lopez-Salido 2007). In contrast, in a perfectly competitive market, the markup would always be zero. Figure 2 illustrates a simulated time-series of markups.

In the data, counter-cyclical markups have been documented by and Rotemberg and Woodford (1999) using three different methods, by Murphy, Shleifer, and Vishny (1989) using input and output prices, by Chevalier, Kashyap, and Rossi (2003) with supermarket data, by Portier (1995) with French data and by Bils (1987) inferring firms' marginal costs. Besides their negative correlation with output, the other salient cyclical feature of markups is that they lag output. Figure 3 shows that the model's markup is negatively correlated as a lagging variable, but turns to positively correlated when it leads, just as in the data. The difference is that the model's markup must lead by 5 quarters, rather than 2 quarters, to achieve a positive correlation.

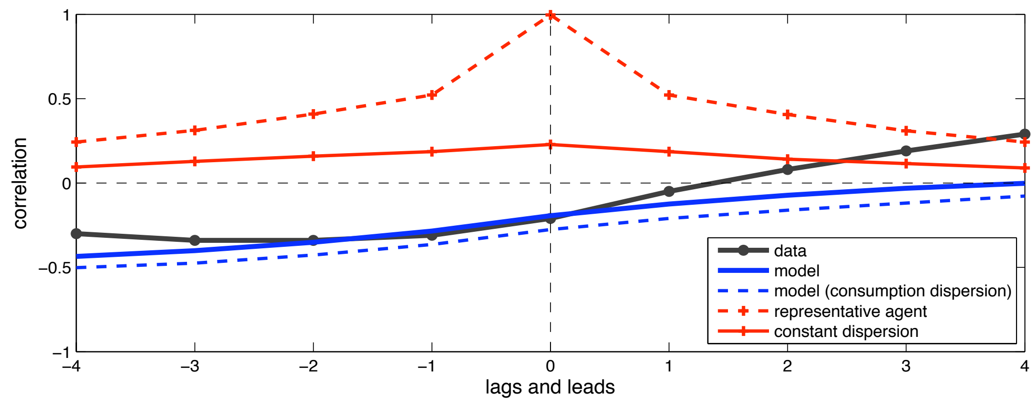

Figure 3: Leads and lags of markup-GDP correlations.

Entries are corr(log(markup$_t$), log(y_{t+k})). Positive numbers indicate leads and negative numbers indicate lags. Data from Rotemberg and Woodford (1999) (table 2, column 2). Markup is estimated using the labor share in the non-financial corporate business sector and an elasticity of non-overhead labor of -0.4.

The reason that the model's markups are lagging is that the earnings dispersion process is highly persistent. In low-productivity periods, it is the shocks to the persistent component of earnings that become more volatile (equation 12). As these high-volatility shocks continue to arrive, the earnings distribution fans out. When productivity picks up and shocks become less volatile, there is enormous dispersion in the persistent component of earnings that takes a long time to revert to its mean. It takes many periods of low-volatility shocks for the earnings distribution to become less dispersed. Since markups are driven by earnings dispersion, which is a lagging variable, markups are lagging as well. This feature of the model is similar to Bilbiie, Ghironi, and Melitz (2006). In their setting, a large fixed cost causes firms to delay entry. Since markups depend on how many firms enter, markups lag the cycle.

2.5 Can the model match standard business cycle moments?

Model variable	relative std dev	corr with GDP
profit share	0.94	0.69
labor	0.49	0.96
real wages	0.28	0.47
Data variable	**relative std dev**	**corr with GDP**
profit share	0.80	0.22 (0.37)
labor (employment)	0.84 (0.82)	0.81 (0.89)
labor (hours)	0.97 (0.98)	0.88 (0.92)
real wages	0.39 (0.36)	0.16 (0.25)*
King and Rebelo (1999)	**relative std dev**	**corr with GDP**
profit share	0.00	0.00
labor	0.48	0.97
real wages	0.54	0.98

Table 2: SECOND MOMENTS OF AGGREGATE VARIABLES IN THE MODEL AND DATA. Standard deviations are divided by the standard deviation of GDP. Most statistics are from Stock and Watson (1999). Labor and wage numbers in parentheses are from Cooley and Prescott (1995). Number with an asterisk is from Rotemberg and Woodford (1999) All capital share statistics come from the labor share statistics reported in Gomme and Greenwood (1995). The second correlation, in parentheses, comes from NIPA data. But the NIPA-based measure counts all proprietors' earnings as profits, although it is part profit and part labor compensation. The first correlation corrects for this by removing proprietor's earnings.

The goal of the paper is to simultaneously explain markups and match business cycle quantities. There is no investment in the model; thus output and consumption are equal. We compare GDP to the other quantities in the model: labor and profits. We also want to investigate properties of prices. But, comparing prices of x and c goods to a measure like the CPI has the problem that the CPI is the rate of exchange between goods and money. Yet there is no money in this model. Therefore, we report a relative price we can interpret, the real wage. In the model, the real wage is the relative price of labor to the expenditure-weighted price index of x and c goods.

Table 2 compares the model aggregates to data. The profit share's cyclical properties do a reasonable job of matching the data. Most importantly, profits are pro-cyclical (although too pro-cyclical and a little too volatile). This is an important piece of evidence that distinguishes this model from sticky price theories, models with free-entry or standard business cycle models such as King and Rebelo (1999). Labor and real wages do slightly less well, but not worse than the standard model. Without a capital stock in the model, wages, labor and output are more driven by changes in productivity. This makes their correlations with output too high.

Understanding counter-cyclical markups is important for business cycle research because the resulting prices are more rigid (less volatile); price rigidity amplifies the effects of productivity shocks on output. In our model, the standard deviation of the log price of the x-good is 0.02 while the standard deviation of log marginal cost is 0.03. Thus, prices are only 2/3rds as volatile as they would be in a standard competitive economy where price equals marginal cost. If our prices were more flexible, they would fall further in recessions so that more x-goods would be sold. But from reading table 2, it appears as though our model does no better than the standard model in explaining macro volatility. But the similarity in the standard deviations of labor and real wages is misleading. Recall that we calibrated our aggregate productivity process to match the volatility of GDP. The calibrated shocks are 1/7th as volatile as those in King and Rebelo (1999).[4] Because our model's recessions are deeper, using the King and Rebelo (1999) productivity process would make our business cycles many times more volatile.

Figure 4 illustrates the behavior of real wages and GDP. It has two features that look familiar. First, real wages look like the mirror image of markups. The intuition for this is that wages are the main component of marginal costs and so wages relative to p behaves

[4]In our calibration, aggregate log productivity has a quarterly persistence of 0.80 and an innovation standard deviation of 0.032 which implies an unconditional quarterly standard deviation of log productivity that is a function of the persistence and volatility of the innovations: $0.0032/\sqrt{1-0.80^2} = 0.005$. In King and Rebelo (1999), the quarterly standard deviation of log productivity is $0.0072/\sqrt{1-0.979^2} = 0.035$.

like the reciprocal of the markup. Furthermore, both real wages and markups are closely correlated with dispersion. Many empirical measures of markups are functions of the inverse real wage (see Rotemberg and Woodford (1999)). The fact that simulated real wages are pro-cyclical means that this alternative measure of the model's markups is counter-cyclical as well. Second, the measure of GDP looks quite similar to the productivity plotted in figure 2. This tells us that, although dispersion has an effect on GDP, it is still primarily driven by productivity shocks.

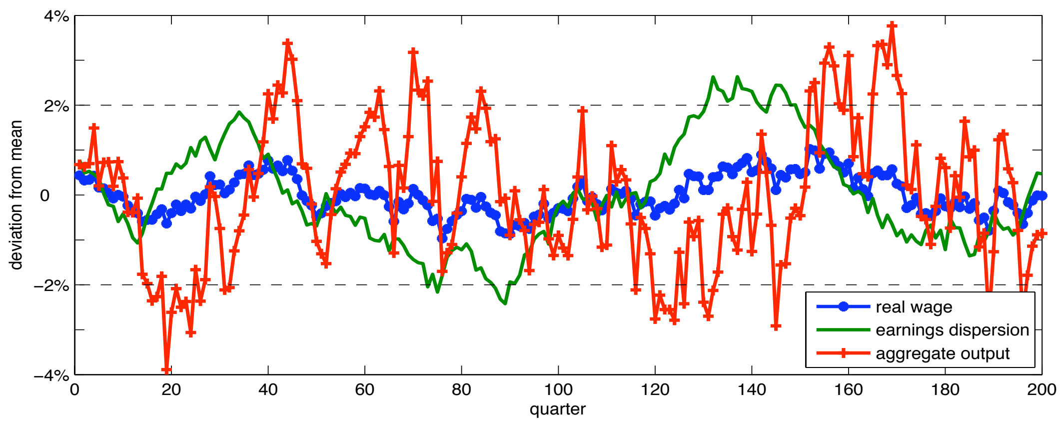

Figure 4: Real wages, earnings dispersion and GDP in the simulated model.

2.6 Benchmark economies

The results so far have been presented in comparison to a standard business cycle model. Because that model has zero markups, the comparison does not tell us which model assumptions drive the markup results. Therefore we compare our model to two more similar benchmarks. The first is an economy where earnings dispersion is constant, always equal to its steady-state value. The second benchmark is an economy where there is no earnings dispersion, only a representative consumer. In both benchmarks, the correlation of markups with GDP is positive, regardless of whether markups are contemporaneous, leading or lagging.

When dispersion is constant and equal to its unconditional mean, many of our calibration targets look similar. The x-good markup (33%), the aggregate markup (12%), the average labor supply (0.33), average profit share (0.30), and the standard deviation and autocorrelation of output (0.017, 0.78) are all essentially unchanged. The key difference is the correlation of markups and GDP (0.23). Markups switched from being counter-cyclical to pro-cyclical.

Alternatively, when earnings dispersion is zero, aggregates are either insufficiently volatile or almost perfectly cyclical. Defining markups in this setting is not obvious because sellers either sell 1 or 0 units, making marginal cost sometimes zero, and markups undefined. Defining the markup as price/average cost, we find that the average markup is negative. Furthermore, the markup is perfectly correlated with and more volatile than GDP. This stark contrast makes clear that the effects of our model are being driven by the earnings dispersion mechanism and that both the presence of earnings dispersion and its time-variation are essential for counter-cyclical markups.

3 Evaluating long-run predictions

While our model is built to explain fluctuations at business cycle frequencies, there has been a long-run increase in the level of earnings dispersion that should cause low-frequency changes too. Earnings dispersion increased by 20% from 1967-1996, an average annual rate of 0.66% (Heathcote, Storesletten, and Violante 2006). To determine if our model's predictions are consistent with the long-run facts, we simulate six model economies, with different levels of earnings dispersion. Each economy represents a decade from the 1950's to the 2000's. To make the time-averaged moments of our growing economy like the moments in our benchmark business cycle model, we choose the 1970's to be the same as our benchmark calibration.

In doing this exercise, we run into a well-known problem. GHH preferences are inconsistent with balanced growth. The standard solution to this problem is to scale up the utility weight on leisure as productivity increases so as to keep hours flat. The formula for individual labor supply is $n_i = (w_i/\theta)^{1/\gamma}$. If we proportionately scale up θ with w_i, average hours will not change. In our model with linear preferences over the x-goods, we also need to scale up the utility weight ν. Changing these two parameters at the rate of productivity growth keeps both average hours and expenditure shares constant. Below we refer to results as having 'no correction' when we do not shift preference parameters and to 'balanced growth' results when we do.

3.1 Long-run slowdown in real wage growth

A long-run change that has been of particular concern to policy-makers is the slowdown in the growth of real wages. The left panel of figure 5 illustrates how real wages were keeping pace with productivity growth until the 1970's, when real wage growth slowed down.

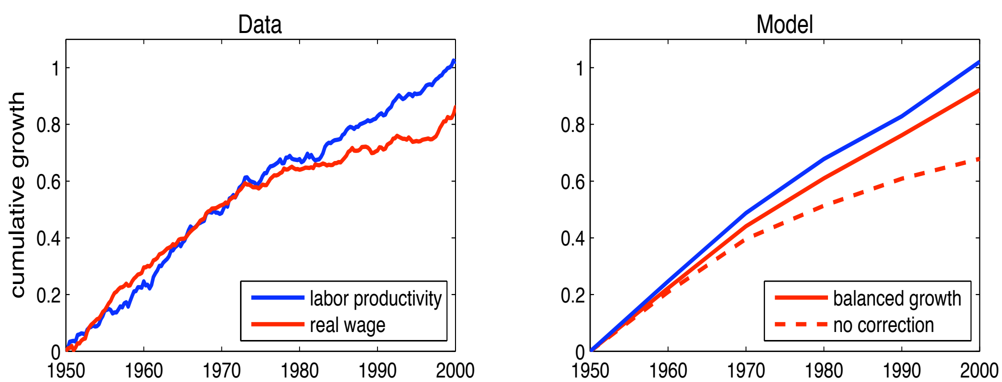

Figure 5: Productivity and wage growth in the data and the simulated model. The trend added to the model is the decade-by-decade increase in labor productivity and in earnings dispersion estimated by Heathcote, Storesletten, and Violante (2006). All other parameter values are listed in table 1. Data: See appendix A.3 for details.

To ask the if the model produces the same effect, we need to calibrate not just the changes in earnings dispersion, but also the changes in aggregate productivity. To do this, we use annual estimates of labor productivity from the BLS, averaged by decade. The right panel of figure 5 shows a pattern in the model similar to that in the data. While our model with balanced growth preferences predicts only half the relative decline in real wages, the baseline model without correction produces an effect twice as strong as that in the data. In contrast, a standard business cycle model would predict that wages and productivity grow in tandem.

The flip side of the relative decline in real wages is an increase in firms' profit shares. The balanced growth model's share of output paid as profits rises steadily from 25% in 1950's to 35% (our calibrated value) in the 1970's to 38% in the 2000's. What happens is that higher dispersion reduces demand elasticity, prompting higher markups and delivering higher profit shares. In the no correction model, rising productivity reinforces this effect, making the rise in profit share more extreme (18% in 1950 to 73% in 2000). As higher productivity increases earnings, the composition of demand changes. Consuming more x-goods means consuming a broader variety of goods and is therefore not subject to the same diminishing marginal returns that set in when c-good consumption increases. Therefore, when earnings increase, x-good consumption rises more than c-good consumption. This effect is big. The expenditure share for x-goods is 22% in 1950 and 75% in 2000. Higher demand for x-goods prompts firms to increase markups, raising profits.

In the data, the evidence on the size of the increase in profit shares is mixed. The share of output not paid out as labor earnings – a very broad definition of profits – rose only by about 5% from 1970-96 (NIPA data). Meanwhile, Lustig and Van Nieuwerburgh (2007) document that net payouts to security holders as a fraction of each firm's value added – a much more narrow definition of profits – rose from 1.4% to 9% (based on flow of funds data) or 2.3% to 7.5% (NIPA data). While the broad measure suggests that our model over-predicts the rise in profits, the 3- to 6-fold rises in profits reflected in the latter measure suggest that the dramatic profit increases predicted by the model may not be so far from the truth.

3.2 The long-run decline in business cycle volatility

One of the most discussed low-frequency changes in the U.S. economy has been the decline of business cycle volatility. One might think that the increase in earnings dispersion over the last 50 years would make the model's cycles more volatile because the individual earnings processes have become more volatile. This concern is not founded. Higher dispersion can lower business cycle volatility because high earnings dispersion makes fewer agents marginal. In other words, it reduces aggregate demand elasticity. Therefore, shocks to labor productivity have less effect on who buys what products. Since producers are producing in anticipation of changes in aggregate demand, when aggregate demand becomes less volatile, GDP volatility falls as well. While our model does not explain the bulk of the fall in business cycle volatility, with the balance growth correction, it can generate a modest decline (see appendix A.4 for details).

4 Evaluating cross-sectional predictions

The model's key mechanism is that higher earnings dispersion increases markups. Because higher markups with a given wage level imply lower real wages, a common measure of markups is the inverse real wage. If this mechanism is operative, then in years and U.S. states with similar productivity, the higher-earnings-dispersion state should have a lower real wage and thus a higher markup. This section documents that pattern and surveys related evidence in the industrial organization literature.

4.1 Testing the model with state-level panel data

The data is a panel of annual observations on 49 U.S. states from 1969-1997. For each state s and year t, our panel contains average labor productivity z_{st}, average real wage, and a measure of cross-county earnings dispersion σ_{st}. As a proxy for state labor productivity, we use real state GDP per employed worker. To measure a state's earnings dispersion, we take the log average salaries in each of the state's counties, weight them by the number of jobs in the county, and take their cross-sectional standard deviation. Of course, the cross-county dispersion is much lower than cross-individual dispersion would be. We measure the state-level markup as the inverse of the real wage. All three variables are transformed into log differences from their national averages. Appendix A.3 gives further details.

To determine if in years and U.S. states with similar productivity, the higher-earnings-dispersion state has a higher markup, we do a double-sort of the data. For each year, sort each state into one of three bins, depending on whether its productivity is in the highest, medium or lowest third for that year. Within each of these three bins, sort states again by earnings dispersion into high-, medium- and low-dispersion bins. That delivers a sort of states into 9 bins, with about 5 states in each bin. Average the markups for the states in each bin. Do this for each year and average the 3-by-3 matrix of markups over all years. Finally, subtract the low-dispersion column from the high-dispersion column. The resulting three numbers (in table 3) reveal what the average difference in markups is between states with similar productivity, relative to the national average, but with higher rather than lower dispersion. The null hypothesis is that the difference in markups is zero. This hypothesis is rejected at the 95% confidence level in each of the three productivity categories.

To compare these numbers to results from our model, we simulate a panel of 50 economies followed for 200 quarters and repeat our double-sort procedure on this artificial data. We calculate the high minus low statistics and then repeat this exercise 100 times and report the average moments in table 3.

As in the data, similar-productivity economies with high dispersion economies have larger markups than those with low dispersion. Moreover for economies in the middle third the

Model	high σ - low σ markup	
low z	3.37	
medium z	3.84	
high z	3.42	
Data	high σ - low σ markup	t-stat
low z	1.69	2.07
medium z	3.67	8.78
high z	2.91	3.45

Table 3: DIFFERENCE IN PERCENTAGE MARKUP, HIGH DISPERSION STATES MINUS LOW DISPERSION STATES, IN THREE PRODUCTIVITY CATEGORIES. High, medium and low states refer to top third, middle third and bottom third of observations. See text and appendix A.3 for further details. The t-statistic for null hypothesis of no significant difference is based on Newey-West standard errors.

difference in markups in the model (3.84%) is close to what it is in the state data (3.67%). For economies in the high and low productivity thirds, the model change in markups is somewhat larger than their empirical counterparts. In short, our main result holds up in the data. In the model higher dispersion raises price because more dispersion reduces the price elasticity of demand. When aggregate demand is less sensitive to price changes, sellers raise prices by raising their markup. In the data, the positive estimates of high σ minus low σ markups confirm this result.

By first sorting states by productivity and then looking for the dispersion effect in each productivity category, we isolate the dispersion effect, without imposing strong parametric assumptions on our estimation. Of course, this exercise does not rule out the possibility that some external factor is causing both markups and dispersion to vary. This would be some force that caused states and years with similar labor productivity but high earnings dispersion to have low real wages (high markups). But since we removed any nationwide source of variation (all variables are deviations from their national averages), any external force generating the covariance cannot be a cyclical effect. It must be independent of aggregate fluctuations.

4.2 Support from empirical pricing studies

Our results are also qualitatively consistent with the findings of Chevalier, Kashyap, and Rossi (2003). Periods of good-specific high demand (e.g., beer on the fourth of July) are times when consumers' values for the goods are more similar. While one might expect that high demand would increase prices, the authors find that prices and markups fall. The same outcome would arise in our calibrated model if productivity dispersion σ falls.

More support for our mechanism comes from a study of the effect willingness-to-pay dispersion has on car sales. Goldberg (1996) estimates that blacks' valuations for new cars are more dispersed than whites'. She then collects data on the initial offer to blacks and whites by a car salesman. The initial offer price is higher, and the probability of sale lower, for the group with more dispersed willingness to pay.

5 Conclusion

Our production economy is set up to capture the intuition that when earnings dispersion is higher, the price elasticity of demand is lower, so sellers optimally raise markups. However, without quantifying the model, the cyclical behavior of prices and markups is ambiguous because the productivity and earnings dispersion effects work in opposite directions. Using estimates of the time-series variation in the earnings distribution, we calibrate the model. Although the model is a simple one, it does a reasonable job of matching the business cycle features of markups and traditional macro aggregates.

Our model provides a theory of real price rigidity, meaning prices that fluctuate less than marginal cost. By themselves, rigid real prices make business cycles more costly. When interacted with a form of nominal rigidity, real rigidities also amplify the real effects of nominal shocks (Ball and Romer (1990) and Kimball (1995)). Future work could merge this theory with a nominal price-setting friction. If such a model delivered enough amplification of small nominal shocks, it could further our understanding of monetary policy's role in the business cycle.

A Appendix

A.1 Solving the GHH model

Agents buy good x_j if and only if $\nu \geq \lambda_i p_j$, where λ_i is the Lagrange multiplier on the budget constraint and satisfies:

$$\lambda_i = \left(c_i - \theta \frac{n_i^{1+\gamma}}{1+\gamma} \right)^{-1}. \tag{19}$$

In equilibrium each seller sets the same price $p = p_j$ all j and each agent buys either none or all of the j goods. Since x goods are sold in discrete $\{0,1\}$ amounts, each agent's expenditure on j goods is either p or 0.

Let $\hat{p}(w_i)$ denote the highest price that an agent with idiosyncratic productivity w_i will pay for the x goods. Using the cutoff rule for x good purchases, this price $\hat{p}(w_i)$ satisfies $\nu = \hat{p}(w_i)\lambda_i$. Substituting for λ_i from (19) gives:

$$\begin{aligned} \nu &= \hat{p}(w_i) \left(c_i - \theta \frac{n_i^{1+\gamma}}{1+\gamma} \right)^{-1} \\ &= \hat{p}(w_i) \left(w_i \hat{n}(w_i) + \pi - \hat{p}(w_i) - \theta \frac{\hat{n}(w_i)^{1+\gamma}}{1+\gamma} \right)^{-1}. \end{aligned}$$

where the second line uses the budget constraint (2) to eliminate c_i and $\hat{n}(w_i)$ is given in (16). Solving for $\hat{p}(w_i)$ yields:

$$\hat{p}(w_i) = \frac{\nu}{1+\nu} \left[w_i \hat{n}(w_i) + \pi - \theta \frac{\hat{n}(w_i)^{1+\gamma}}{1+\gamma} \right]. \tag{20}$$

This is a continuous, strictly increasing and strictly convex function of w_i.

Now let $\hat{w}(p)$ denote the inverse of the reservation price function $\hat{p}(w)$, i.e., $\hat{w}(\hat{p}) = 1$, so that $\hat{w}(p)$ represents the lowest idiosyncratic productivity draw that will lead a buyer to purchase at price p (so $\hat{w}(p)$ is strictly increasing and strictly concave). Then the total demand facing firm j at price p_j is:

$$x_j = \Pr[w \geq \hat{w}(p_j)] = 1 - \Phi\left(\frac{1}{\sigma} \log\left(\frac{\hat{w}(p_j)}{z} \right) \right), \tag{21}$$

where Φ is the standard normal cumulative distribution function.

The resulting profit of firm j is revenue minus costs $\pi_j = p_j x_j - x_j^{1/\alpha}$ Aggregate profits are $\pi = \int_0^1 \pi_j dj$. Since in equilibrium $p_j = p$ for all j, $\pi_j = \pi$ for all j too. In equation (20), aggregate profits π enter the reservation price function $\hat{p}(w)$ and therefore enter the inverse $\hat{w}(p)$ too. To acknowledge this dependence, write $\hat{w}(p, \pi)$. To compute an equilibrium numerically, then, we need to solve a fixed point problem of the form $\pi = F(\pi)$ where:

$$F(\pi) := \max_{p \geq 0} \left\{ p \left[1 - \Phi\left(\frac{1}{\sigma} \log\left(\frac{\hat{w}(p,\pi)}{z} \right) \right) \right] - \left[1 - \Phi\left(\frac{1}{\sigma} \log\left(\frac{\hat{w}(p,\pi)}{z} \right) \right) \right]^{1/\alpha} \right\}. \tag{22}$$

We do this numerically, guessing an initial π_0, iterating on $\pi_{k+1} = F(\pi_k)$ for $k \geq 0$ and then iterating until $|\pi_{k+1} - \pi_k| < 10^{-6}$.

Steady state calibration targets We solve for six parameters $(\alpha, \gamma, \nu, \theta, \sigma, \overline{z})$ such that the steady state of our model delivers the following six properties:

$$\begin{aligned}
\text{elasticity of labor supply} &= \mathbb{E}[d\log(n_i)/d\log(w_i)] &= 1.67\\
\text{earnings dispersion} &= \overline{\sigma}_{\text{STY}} &= 0.29\\
\text{hours worked} &= \mathbb{E}[n_i] &= 0.33\\
\text{labor share} &= \mathbb{E}[w_i n_i]/\overline{y} &= 0.70\\
\text{aggregate markup} &= \overline{m} &= 1.11\\
x\text{-sector markup} &= \overline{m}_x &= 1.30
\end{aligned}$$

We use the following properties of the model repeatedly: individual labor supply is, from (16), $n_i = (w_i/\theta)^{1/\gamma}$. Since log idiosyncratic productivity $\log(w_i)$ is normal with mean $\log(\overline{z})$ and standard deviation σ, log labor supply is:

$$\log(n_i) \sim \mathcal{N}[(1/\gamma)\log(\overline{z}/\theta), (\sigma/\gamma)^2],$$

and so:

$$\mathbb{E}[n_i] = (\overline{z}/\theta)^{1/\gamma}\exp[0.5(\sigma/\gamma)^2].$$

Similarly, log earnings is:

$$\log(w_i n_i) \sim \mathcal{N}[((1+\gamma)/\gamma)\log(\overline{z}) - (1/\gamma)\log(\theta), ((1+\gamma)\sigma/\gamma)^2],$$

and so:

$$\mathbb{E}[w_i n_i] = \overline{z}^{(1+\gamma)/\gamma}\theta^{-1/\gamma}\exp[0.5((1+\gamma)\sigma/\gamma)^2].$$

Given this, the average elasticity of labor supply is $\mathbb{E}[d\log(n_i)/d\log(w_i)] = 1/\gamma$ which equals 1.67 when $\gamma = 0.60$. The standard deviation of log earnings is $[(1+\gamma)/\gamma]\sigma$ which equals the Storesletten, Telmer, and Yaron (2004) estimate of $\overline{\sigma}_{\text{STY}} = 0.29$ when $\sigma = (0.60/1.60)(0.29) = 0.11$.

The remaining four parameters $(\alpha, \nu, \theta, \overline{z})$ have to be solved for simultaneously. From our previous calculations, one condition is immediate:

$$\mathbb{E}[n_i] = (\overline{z}/\theta)^{1/0.60}\exp[0.5(0.11/0.60)^2] = 0.33. \tag{23}$$

We also use labor's share $\mathbb{E}[w_i n_i]/\overline{y} = 0.70$ and since $\overline{y} = \mathbb{E}[w_i n_i] + \overline{\pi}$, we need to have $\mathbb{E}[w_i n_i] = (0.70/0.30)\overline{\pi}$. To calculate $\overline{\pi}$ we need to solve the fixed point problem $\overline{\pi} = F(\overline{\pi})$ outlined above. The solution of the fixed point problem depends on all four parameters and to acknowledge this write $\overline{\pi}(\alpha, \nu, \theta, \overline{z})$. Using the formula for average earnings derived above, we then have a second equation in the four unknowns:

$$\mathbb{E}[w_i n_i] = \overline{z}^{(1.60)/0.60}\theta^{-1/0.60}\exp[0.5((1.60)(0.11)/0.60)^2] = \frac{0.70}{0.30}\overline{\pi}(\alpha, \nu, \theta, \overline{z}). \tag{24}$$

The solution to the fixed point problem also gives us an optimal price $\overline{p}(\alpha, \nu, \theta, \overline{z})$ and associated percentage markup $\overline{m}_x(\alpha, \nu, \theta, \overline{z})$ for the x-sector. Our third equation in the four unknowns is therefore:

$$\overline{m}_x(\alpha, \nu, \theta, \overline{z}) = 1.30. \tag{25}$$

Let $\overline{x}$ denote the amount of the x-good sold by the firm at the price $\overline{p}$ and let $\overline{\varepsilon}_x := (\overline{px})/\overline{y}$ denote the expenditure share on the x-sector. We define the aggregate markup as the expenditure share weighted average of the x-sector and c-sector markups, $\overline{m} := \overline{\varepsilon}_x\overline{m}_x + (1 - \overline{\varepsilon}_x)$, since the c-sector percentage markup is zero by definition. Rearranging gives our fourth equation:

$$\overline{\varepsilon}_x(\alpha, \nu, \theta, \overline{z}) = \frac{\overline{m} - 1}{\overline{m}_x - 1} = \frac{0.11}{0.30}. \tag{26}$$

In short, we solve for the four parameters $(\alpha, \nu, \theta, \overline{z})$ by solving the four equations (23)-(26) simultaneously. This gives us $\alpha = 0.24, \nu = 100, \theta = 15$ and $\overline{z} = 7.68$.

Consumption-dispersion model	level	relative std dev	corr with GDP
x-good markup	24%	0.42	-0.36
profit share	0.32	0.78	0.88
labor	0.33	0.44	0.98
real wages	5.8	0.18	0.32
Low-α model ($\alpha = 0.12$)	level	relative std dev	corr with GDP
x-good markup	28%	0.95	-0.38
profit share	0.46	0.96	0.66
labor	0.33	0.44	0.67
real wages	4.4	0.64	-0.26
High-α model ($\alpha = 0.48$)	level	relative std dev	corr with GDP
x-good markup	35%	0.75	-0.09
profit share	0.21	0.93	0.80
labor	0.33	0.52	0.99
real wages	6.8	0.24	0.89

Table 4: Second moments of aggregate variables in the low-dispersion model. Standard deviations are divided by the standard deviation of GDP.

A.2 Sensitivity analysis

Replace earnings dispersion with consumption dispersion One criticism of the model calibration is that there are many reasons the the distribution of customer willingness to pay that a firm faces may be lower than earnings dispersion. One reason is that agents use savings to smooth out temporary income shocks. Another reason is that stores may not serve the entire population, but cater to a segment of it. For these reasons, it is important that the model's results survive with a lower level of dispersion.

Storesletten, Telmer, and Yaron (2004) also report consumption dispersion estimates, using food consumption data from the PSID. We use the procedure described in appendix A.3 for transforming annual to quarterly estimates. Since the persistence of consumption is slightly lower than earnings, the annual to quarterly conversion factor is different. The quarterly AR1 coefficient in persistent piece of individual earnings is $\rho_\xi = 0.862^{1/4} = 0.964$. This delivers a factor for converting annual to quarterly standard deviations $Q = (1 + \rho_\xi + \rho_\xi^2 + \rho_\xi^3)^{-1} = 0.264$. This conversion yields the parameters of the idiosyncratic earnings process. To convert these to parameters of the idiosyncratic productivity process that we feed into the model, multiply each by $\gamma/(1+\gamma)$. Thus, $\sigma_h = 0.172Q\gamma/(1+\gamma) = 0.017$, $\sigma_l = 0.222Q\gamma/(1+\gamma) = 0.02$, and $\sigma_u Q\gamma/(1+\gamma) = 0.283Q = 0.028$. The resulting steady state dispersion of consumption is 0.21, about 2/3rds of the steady state earnings dispersion (0.29) from the benchmark model. None of the other parameters are changed.

Table 4 shows that lowering the dispersion makes markups more counter-cyclical. Obviously, there is a limit to how low dispersion can fall since in the zero-dispersion case, markups are pro-cyclical. The level of the markup is lower. This is also true for the aggregate markup (9.8%). The moments for the macro variables look very similar to those from the benchmark model. The important take-away from this exercise is that the results are not very sensitive and even get stronger if the dispersion in willingness to pay is substantially lower than the dispersion in earnings.

Vary diminishing returns parameter α Since aggregate markups are used to calibrate α, α is sensitive to the markup level, and markups are not precisely estimated, α is a prime candidate for robustness analysis. Two exercises alleviate these concerns. The bottom two sections of table 4 show that halving or doubling α leaves most of our main results in tact. The most notable exception is that when α is very low, real wages become counter-cyclical. This happens because firms' marginal costs are very volatile. Since those costs are pro-cyclical, it makes prices strongly pro-cyclical and real wages counter-cyclical.

A.3 Data and simulation details

Making annual dispersion quarterly The quarterly persistence and standard deviation of income are derived from the annual estimates of Storesletten, Telmer, and Yaron (2004) as follows: $\rho_\xi = 0.952^{1/4}$, the standard deviation to the persistent component is $0.125Q$ when productivity is above average and is $0.211Q$ when productivity is below average while the standard deviation of the transitory component is $0.255Q$ where the adjustment factor is $Q := 1/(1 + \rho_\xi + \rho_\xi^2 + \rho_\xi^3) = 0.2546$.

Simulations All simulations in this paper begin by sampling the exogenous state variables for a 'burn-in' of 1000 quarters. This eliminates any dependence on arbitrary initial conditions. A cross-section of 2500 individuals is tracked for 200 quarters, corresponding to the dimensions in Storesletten, Telmer, and Yaron (2004). Realizations of endogenous variables are then computed. The moments discussed in the text are averages over the results from 100 runs of the simulation (that is, averages over $200 \times 100 = 20000$ observations).

Aggregate data All data is quarterly 1947:1-2006:4 and seasonally adjusted. Real GDP is from the Bureau of Economic Analysis (BEA). This is nominal GDP deflated deflated by the BEA's chain-type price index with a base year of 2000. We measure aggregate labor productivity by real output per hour and and real wages by real compensation per hour in the non-farm business sector, both from the Bureau of Labor Statistics (BLS) Current Employment Survey (CES) program. Nominal output and compensation are deflated by the BLS's business sector implicit price deflator with a base year of 1992.

State-level data First, we describe the data sources. The state employment, earnings and wage data come from three sources. The first is the County Wage and Salary Summary (CA34), produced by the BEA's Regional Economic Accounts (REA). The data are reported annually from 1963-2004. GDP by state, also from the REA, is aggregated based on weights from the Standard Industrial Classification (SIC) codes from 1963-97, and based on the revised North American Industry Classification System (NAICS) from 1997-2005. Since the underlying methodologies are so different, we do not attempt to splice GDP numbers from the SIC and NAICS accounts and instead, end the sample in 1997. Due to missing data for Alaska, we use 49 states. The District of Columbia is excluded because computing dispersion is impossible with only one county. We use a second source for state consumer price indices, which we use to construct real wages. The BEA reports state price indices, but only from 1990-2005. Del Negro (1998) estimated the indices for 1969-1995. Both measures are annual. The result is a balanced panel of 49 states and 27 time periods for a total of 1323 observations.

Next, we construct the three variables productivity z, earnings dispersion σ and markups m. Productivity measures output per worker (labor productivity). It is state GDP, divided by the state CPI to get real state GDP, then divided again by total state employment. Earnings dispersion uses county-level data on the average labor earnings per capita. To get a units-free measure of dispersion, first take the log of earnings in each county. In each period, state earnings dispersion is the standard deviation of log earnings, across all the counties in the state. Finally, markups are the inverse of the real wage. The state real wage the average nominal wage, divided by the state CPI.

To make our data stationary we remove trends from all variables. While we could remove state-specific deterministic trends, we instead remove a single national trend. This helps preserve as much cross-sectional information as possible. For example, according to our measure a state with persistently lower inequality relative to the national average always has below-trend dispersion; if we had instead removed a state-specific trend then this state would sometime have below trend dispersion and sometime above trend dispersion.

We calculate national trends for each variable v as the average across states with each state weighted by its total number of employed workers: $v_{national,t} = \sum_{s=1}^{49} v_{s,t} l_{s,t} / (\sum_{s=1}^{49} l_{s,t})$, where l is the number of employed workers in state s and year t. The de-trending is done by computing the log deviation of the state series from the national average: $\log(v_{s,t}) - \log(v_{national,t})$.

A.4 Computing the decline in business cycle volatility

In each decade, the model's earnings dispersion by decade is chosen so that its log change from the previous decade matches Heathcote, Storesletten, and Violante (2006). When only that change is made to the model, business cycle volatility increases because of an unintended side-effect: When dispersion increases, the average productivity level does as well because idiosyncratic productivity is log-normal. Higher productivity raises hours worked and shifts the expenditure from c-good to consumption to x-good consumption. x-good consumption is more volatile because of its linear utility.

With the balanced growth correction, results improve. Our model predicts essentially flat business cycle volatility. To achieve a decrease in business cycle volatility, we need a smaller and less rapidly growing dispersion process, like that for consumption dispersion. Storesletten, Telmer, and Yaron (2004) report that food consumption has about 2/3rds the dispersion of earnings. We match the level of dispersion in the 70's and its cyclical properties to their estimate. For the long-run increase in consumption dispersion, we use the 5% per-decade increase in non-durable consumption dispersion reported by Krueger and Perri (2005a) for 1970-2000. (See appendix A.2 for further details.) The 10% rise in consumption dispersion from the 1970's-90's results in a 24% drop in the standard deviation of log real GDP, but it does not reproduce the halving of business cycle volatility in the data.

Decade	model (benchmark σ)	model (low σ)	data
50's	1.50		2.1
60's	1.49		1.3
70's	1.55	2.10	2.2
80's	1.55	1.68	1.7
90's	1.53	1.70	0.9
00's	1.50		1.0

Table 5: Business cycle volatility as dispersion increases.
Volatility measured as standard deviation of $\log(y)$ in percent. The trend added to the model is the decade-by-decade increase in earnings dispersion estimated by Heathcote, Storesletten, and Violante (2006). 'Low σ' refers to the model calibrated to match the rise in consumption dispersion as estimated by Krueger and Perri (2005a). In both cases we use the balanced growth correction described in the text. All other parameter values are listed in table 1. Data: standard deviations of quarterly GDP, by decade.

References

BALL, L., AND D. ROMER (1990): "Real Rigidities and the Non-Neutrality of Money," *Review of Economic Studies*, 57(2), 183–203.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

BILBIIE, F., F. GHIRONI, AND M. J. MELITZ (2006): "Endogenous Entry, Product Variety, and Business Cycles," Working Paper.

BILS, M. (1987): "The Cyclical Behavior of Marginal Cost and Price," *American Economic Review*, 77(5), 838–855.

——— (1989): "Pricing in a Customer Market," *Quarterly Journal of Economics*, 104(4), 699–718.

CABALLERO, R., AND M. HAMMOUR (1994): "The Cleansing Effect of Recessions," *American Economic Review*, 84, 1350–68.

CHARI, V., P. KEHOE, AND E. MCGRATTAN (2000): "Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem?," *Econometrica*, 68(5), 1151–1179.

CHEVALIER, J., A. KASHYAP, AND P. ROSSI (2003): "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data," *American Economic Review*, 93, 15–37.

COMIN, D., AND M. GERTLER (2006): "Medium Term Business Cycles," *American Economic Review*, 96(3), 523–551.

COOLEY, T., R. MARIMON, AND V. QUADRINI (2004): "Aggregate Consequences of Limited Contract Enforceability," *Journal of Political Economy*, 112, 817–847.

COOLEY, T., AND E. PRESCOTT (1995): "Economic Growth and Business Cycles," in *Frontiers in Business Cycle Research*, ed. by T. Cooley. Princeton University Press.

DEL NEGRO, M. (1998): "Aggregate Risk Sharing Across US States and Across European Countries," FRB Atlanta Working Paper.

DIAZ-GIMENEZ, J., V. QUADRINI, AND J. RIOS-RULL (1997): "Dimensions of Inequality: Facts on the U.S. Distributions of Earnings, Income and Wealth," *Federal Reserve Bank of Minneapolis Quarterly Review*, 21(2), 3–21.

GALI, J. (1994): "Monopolistic Competition, Business Cycles, and the Composition of Aggregate Demand," *Journal of Economic Theory*, 63, 73–96.

Gali, J., M. Gertler, and D. Lopez-Salido (2007): "Markups, Gaps and the Welfare Costs of Business Cycles," *Review of Economics and Statistics*, 89(1), 44–59.

Goldberg, P. (1996): "Dealer Price Discrimination in New Car Purchases: Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, 104(3), 622–654.

Gomme, P., and J. Greenwood (1995): "On the Cyclical Allocation of Risk," *Journal of Economic Dynamics and Control*, 19, 19–124.

Greenwood, J., Z. Hercowitz, and G. Huffman (1988): "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, 78(3), 402–417.

Guvenen, F. (2005): "An Empirical Investigation of Labor Income Processes," U. Texas Working Paper.

Heathcote, J., K. Storesletten, and G. Violante (2006): "The Macroeconomic Implications of Rising Wage Inequality in the United States," NYU Working Paper.

Jaimovich, N. (2006): "Firm Dynamics, Markup Variations, and the Business Cycle," Stanford University Working Paper.

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

King, R., and S. T. Rebelo (1999): "Resuscitating Real Business Cycles," in *Handbook of Marcoeconomics*, ed. by J. Taylor, and M. Woodford. Elsevier.

Kiyotaki, N., and R. Wright (1989): "On Money as a Medium of Exchange," *Journal of Political Economy*, 97(4), 927–954.

Krueger, D., and F. Perri (2005a): "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies*, 73, 163–193.

——— (2005b): "Public versus Private Risk Sharing," Working Paper.

Krusell, P., and A. Smith (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106(5).

Lustig, H., and S. Van Nieuwerburgh (2005): "Housing Collateral, Consumption Insurance and Risk Premia: An Empirical Perspective," *Journal of Finance*, 60(3), 1167–1219.

——— (2007): "Why Have Payouts to U.S. Corporations Increased?," Working paper.

Menzio, G. (2007): "A Search Theory of Rigid Prices," U. Penn working paper.

Murphy, K., A. Shleifer, and R. Vishny (1989): "Building Blocks of Market-Clearing Models," *NBER Macroeconomics Annual*, pp. 247–286.

Nakamura, E., and J. Steinsson (2006): “Price Setting in Forward-Looking Customer Markets,” Harvard University working paper.

Nevo, A. (2001): “Measuring Market Power in the Ready-to-eat Cereal Industry,” *Econometrica*, 69(2), 307–342.

Portier, F. (1995): “Business Formation and Cyclical Markups in the French Business Cycle,” *Annales D'Economie et de Statistique*, 37-38, 411–440.

Rampini, A. (2004): “Entrepreneurial Activity, Risk, and the Business Cycle,” *Journal of Monetary Economics*, 51, 555–573.

Rios-Rull, J. (1996): “Life-Cycle Economies and Aggregate Fluctuations,” *Review of Economic Studies*, 63, 465–490.

Rotemberg, J., and G. Saloner (1986): “A Supergame-Theoretic Model of Price Wars during Booms,” *American Economic Review*, 76(3), 390–407.

Rotemberg, J., and M. Woodford (1999): “The Cyclical Behavior of Prices and Costs,” in *Handbook of Marcoeconomics*, ed. by J. Taylor, and M. Woodford. Elsevier.

Stock, J., and M. Watson (1999): “Business Cycle Fluctuations in US Macroeconomic Time Series,” in *Handbook of Marcoeconomics*, ed. by J. Taylor, and M. Woodford. Elsevier.

Storesletten, K., C. Telmer, and A. Yaron (2004): “Cyclical Dynamics in Idiosyncratic Labor Market Risk,” *Journal of Political Economy*, 112(3), 695–717.