Business Cycle Implications of Capacity Constraints under Demand Shocks

Florian Kuhn* Chacko George†‡

March 8, 2018

Abstract

When capacity constraints limit the production of heterogeneous firms, demand shocks can endogenously generate a number of important business cycle regularities: recessions are deeper than booms, economic volatility is countercyclical, the aggregate Solow residual is procyclical and the fiscal multiplier is countercyclical. The model's main mechanism is that the share of firms at their production limit is strongly procyclical. A baseline calibration of a basic New Keynesian DSGE model with capacity constraints delivers more than 25% of the empirically observed asymmetry in output, 18% of the additional cross-sectional dispersion in recessions and around 25% of the additional aggregate volatility, and more than 50% of the fluctuations in the Solow residual. The model implies fluctuations in the fiscal multiplier of around 0.12 between expansions and recessions.

JEL codes: E13, E22, E23, E32
Keywords: Asymmetric business cycles, economic volatility, Solow residual, fiscal multiplier, capacity constraints

*Department of Economics, Binghamton University, PO Box 6000, Binghamton, NY 13902-6000, email: fkuhn@binghamton.edu

†FDIC, Center for Financial Research, 550 17th St NW, Washington, DC 20429, email: cgeorge@fdic.gov

‡We thank Matthias Kehrig, Olivier Coibion, Andy Glover and Saroj Bhattarai for helpful suggestions and comments. Naturally any errors are our own. Opinions expressed in this paper are those of the authors and not necessarily those of the FDIC.

1 Introduction

This paper studies how to reconcile within a simple framework four disparate business cycle facts: the asymmetry of business cycle fluctuations, the countercyclicality of aggregate and cross-sectional volatility, the acyclicality of utilization-adjusted total factor productivity, and counter-cyclical fiscal multipliers. Together, these empirical findings characterize recessions as times when output is especially low, volatility is high, and fiscal policy is particularly effective.

While previous work has considered mechanisms that can account for each fact in isolation, these potential explanations are generally at odds with other facts. For example, one can appeal to asymmetric business cycle shocks to explain the asymmetry in business cycles, but this would not, by itself, account for the observed countercyclicality in the dispersion of cross-sectional firm productivity. Rather than trying to combine all of the mechanisms that could potentially account for each fact individually into an unwieldy model, we instead show that a single mechanism — occasionally binding capacity constraints— can endogenously generate each of these business cycle facts when introduced into an otherwise standard business cycle model.

In the model, firms choose their capital capacity before the realization of idiosyncratic and aggregate demand shocks. After learning about these, they may vary their utilization of capital in a way that is increasingly costly as the utilization rate increases. When the economy experiences positive shocks to the demand for firms' products, they increase their capital utilization and output. With capital predetermined, this endogenous choice of utilization gives rise to procyclical measured total factor productivity even when business cycles are driven by shocks other than TFP. At the same time utilization-adjusted factor productivity may remain acyclical, as documented by Basu et al. (2006).

The combination of predetermined capital and convex utilization costs yields an upper bound on any individual firm's production. Large, positive aggregate shocks, then, increase the number of firms at their capacity constraint. This adds extra concavity to aggregate production as a function of demand and helps explain the three remaining business cycle facts. First, booms are "smaller" than downturns, in the sense that average deviations of output from trend are smaller in absolute value when the economy is far above trend than far below trend. In the calibrated model, capacity constraints generate around one quarter of the observed asymmetry of U.S. business cycles.

Second, capacity constraints provide a channel through which fiscal multipliers can be countercyclical. Higher government spending that increases demand for firms products will have larger effects when the economy is in a downturn than in an expansion. During downturns, few firms are capacity constrained and they can therefore readily expand production. During a boom, on the other hand, firms are already producing at their capacity constraint which reduces the expansionary effects of fiscal policy. Quantitatively, while the extent of countercyclicality of fiscal multipliers remains a point of contention empirically, the model here suggests a difference of about 0.12 between the multiplier in recession and expansion, respectively.

Third, the upper limit to production reduces cross-sectional and aggregate volatility when many firms have high capacity utilization. Idiosyncratic demand shocks generate a non-trivial distribution in the measured productivity of firms. The share of firms at their capacity constraint affects the variance of this distribution: since all constrained firms look very similar in terms of their productivity, a higher share of constrained firms implies a lower variance in the distribution of productivity. Recessions, during which few firms are capacity constrained, are then periods of high cross-sectional productivity dispersion. Occasionally binding capacity constraints therefore provide a previously unexplored channel through which cross-sectional productivity dispersion can endogenously move in a countercyclical manner even in the absence of second-moment shocks. Similarly, when the economy-wide utilization is already high, additional demand shocks do not move aggregate output much. Firms at their constraint are in the flat part of their production function and hence many of them do no respond to changes in demand. This explains why aggregate volatility, as measured by the conditional growth rate of aggregate output, is higher in recessions.

Understanding the properties of recessions matters in the assessment of their welfare costs. For example, while symmetric fluctuations reduce welfare, this loss is more severe if fluctuations exhibit asymmetry and the cost of a downturn is concentrated in a short period of time. Increased volatility in recessions can similarly reduce the welfare of risk-averse agents, and, as recent literature has shown, can have adverse economic effects of its own. The question of how economic fluctuations originate and are transmitted also has important implications for fiscal policy because the efficacy of government spending in general depends heavily on the cause of downturns. For example, the government multiplier is generally acyclical in standard models, whereas in models of uncertainty shocks, government spending can actually be less effective in recessions than in normal times.

The main contribution of this paper is to show that capacity constraints can explain several important features of the behavior of output under few additional assumptions. Second, capacity constraints suggest a novel explanation as to why productivity dispersion among firms is countercyclical. Third, while the traditional Keynesian literature has long emphasized idle capacities as one likely source of high fiscal multipliers when aggregate demand is low, there has been relatively little work on integrating this mechanism into modern DSGE models. This paper provides such a model. Fourth, we document how much this model, in addition to being qualitatively consistent, can contribute quantitatively to the explanation of the four business cycle facts. Finally, we add some empirical evidence to previous work on output asymmetry and find that large recessions on average deviate 30% more from trend output than large booms.

A number of papers study the effects of variable capacity utilization in general equilibrium frameworks. Work by Fagnart et al. (1999), Gilchrist and Williams (2000), Álvarez-Lois (2006) and Hansen and Prescott (2005) investigates capacity constraints with heterogeneous firms. The main difference to the present paper is that they consider shocks to aggregate TFP under putty-clay technology or irreversibilities, whereas we focus on fluctuations in aggregate demand under standard Cobb-Douglas production in which capacity constraints arise endogenously rather than as an assumption on production technology. The closest models are Fagnart et al. (1999) and Álvarez-Lois (2006), who explicitly model the pricing decision of monopolistically competitive firms. Fagnart et al. (1999) focus on the amplification of TFP shocks under putty-clay technology and flexible prices, whereas Álvarez-Lois (2006) looks at the response of firm mark-ups when prices are set one period in advance as well as the internal propagation of the putty-clay mechanism. Gilchrist and Williams (2000) emphasize the asymmetric effects on output following large TFP shocks and the hump-shaped response that is generated through the effects of vintage capital. Hansen and Prescott (2005) generate asymmetries by including a choice along the extensive margin of operating or idling plants.

A strand of papers considers variable capacity utilization in a representative-agent framework (Greenwood et al. (1988), Cooley et al. (1995), Bils and Cho (1994), Christiano et al. (2005)). In contrast, the environment with heterogeneous firms allows us to consider occasionally binding capacity constraints, as well as price setting and demand shocks in the monopolistic competition framework. This firm heterogeneity in turn is driving several of the results in our model, as we show in section 5.

A recent paper that also looks at the interplay of cross-sectional and aggregate asymmetries is Ilut et al. (2016), albeit under a different mechanism. They show that under ambiguity aversion (or more generally any concave reaction of employment growth to expected profitability), news shocks can tightly link countercyclical volatility at the micro and macro level. Their explanation involving firms' decision making offers a complementary alternative to the approach in this paper focusing on firms' production technology.

The key mechanism in our model is that the number of capacity constrained firms is procyclical. While the degree to which capacity constraints bind is hard to measure directly, we show below that three empirical observations are consistent with our mechanism: We see significant procylicality in a) aggregate and sectoral capacity utilization, b) firms' investment in new capacity as documented by Bachmann and Zorn (2016) and c) firms' backlogs of unfilled orders. Additionally, testing a prediction of the mechanism, we also find that sectors with higher variance in capacity utilization exhibit stronger business cycle asymmetry.

The paper is structured as follows: In the next section 2 we review the stylized facts established by recent literature and add evidence on business cycle asymmetry. In section 3 we illustrate in a stylized example how capacity constraints can generate these facts qualitatively. We embed this mechanism in a full DSGE model in section 4, and discuss quantitative results in section 5. Section 6 concludes.

2 Four business cycle regularities and capacity constraints

In the following we review the evidence for the four business cycle facts (asymmetry in output, countercyclical profitability dispersion, strong dependence of the Solow residual's cyclicality on factor utilization, a countercyclical fiscal multiplier) that previous literature has found. Since business cycles can be "asymmetric" in many ways, we discuss the specific type of asymmetry we are interested in and then provide additional evidence from US output series.

Large negative deviations in output from trend are bigger than large positive deviations The question of whether business cycles are asymmetric is fairly old. However, as noted by McKay and Reis (2008), it is also too broad to answer — there are many different ways in which business cycle asymmetry could theoretically manifest itself. As they emphasize, one should therefore be specific in exactly which way one wants to assess asymmetries. Previous literature can be loosely grouped into four ways to research this question: By looking for asymmetry in 1) output growth 2) output levels 3) employment growth 4) employment levels. It is worth recalling that asymmetry in levels and growth rates need not be associated. As discussed for example in Sichel (1993), a time series exhibits asymmetry in levels if, say, troughs are far below trend but peaks are relatively flat. Asymmetry in growth rates would be characterized by, say, sudden drops and slow recoveries. Correspondingly, these two types of asymmetry have been dubbed "deepness" and "steepness", respectively, in the literature.

Our reading of the literature is that there is no strong evidence for asymmetry in output growth rates which most papers have focused on (e.g. DeLong and Summers (1986), Bai and Ng

(2005), McKay and Reis (2008)), although Salgado et al. (2015) find evidence for skewness in growth rates in international data. As documented by Sichel (1993) and Knüppel (2014), there is evidence for skewness in output levels. Employment tends to behave more skewed than output over the cycle: Prior work has found asymmetry in both employment growth and in employment levels (e.g. Ilut et al. (2016), McKay and Reis (2008)).

The focus of this paper is on the claim that large deviations of output from trend are more likely to be negative than positive. This means we are interested in the behavior of output *levels*, for which there is some evidence of asymmetry (Sichel (1993)).

In Table 1 we report a number of additional observations about the relative magnitude of "strong" booms and recessions. Specifically, we use a detrended output series to construct three measures of differences in large output deviations. For the first measure, we pick an integer N and compare the $N/2$ largest (i.e. positive) deviations with the $N/2$ smallest (i.e. negative) deviations by comparing their means. Here, if business cycles are asymmetric in levels, we would expect the mean deviation in strong recessions to be larger than the mean deviation in strong expansions. Second, in the next column we count how many of the N periods with the largest absolute deviations from trend were positive versus negative. If output is asymmetric as defined above, we would expect the number of periods with negative output deviations to be larger. As a third measure we report the overall skewness of the series (using all periods), defined as the sample estimate of $E\left[(x-\mu)^3/\sigma^3\right]$. This coefficient of skewness is a less direct measure of only large output deviations, but all else equal we would expect the coefficient of skewness to be negative.

We construct these measures for a range of specifications in which we vary the time-series representing "output", the length of the series, the trend filter, as well as the number N of extreme periods considered. The baseline specification uses HP-filtered postwar data. HP filtering often constitutes the weakest case in terms of differences between expansions and recessions since at the edges of the sample this detrending method tends to attribute parts of the cyclical movement into the trend. For almost all specifications in Table 1 we see that large deviations from trend are more likely to be negative.[1] On average across all specifications, recessions appear around 30% deeper than booms are high.

In section 5 we calibrate our model to an HP-1600-filtered quarterly US GDP series, corresponding to the quarterly baseline specification in Table 1. The model will yield trend deviations of 3.27% in an expansion and -3.46% in a recession and thus covers a little more than a quarter of the observed asymmetry under the baseline specification.

Cross sectional and aggregate volatility are countercyclical The second fact is connected to a range of findings that associate recessions with increased microeconomic and macroeconomic volatility. On the microeconomic level, recent literature has found strong evidence for countercyclicality of cross-sectional dispersion among firms in several measures. Eisfeldt and Rampini (2006) show that capital productivity is more dispersed in recessions. Bloom (2009) and Bloom et al. (2012) include empirical evidence associating times of low aggregate production to higher dispersion in sales growth, innovations to plant profitability, and sectoral output. Directly related to levels of firm productivity, Kehrig (2015) finds that the distribution of plant

[1] In fact the only specification in which negative output deviations are not larger than positive deviations is for annual GDP when we start the series in 1929 and use an HP filter which, at the beginning of the sample, picks up the Great Depression as part of the trend.

Table 1: Strong recessions larger than strong expansions

Specification	Mean pos vs neg	# pos vs neg	Skewness
Quarterly GDP			
Baseline	2.73% vs −3.43%	16 vs 24	−0.46
$N = 20$	3.12% vs −4.33%	6 vs 14	−0.46
$N = 80$	2.28% vs −2.87%	40 vs 40	−0.46
Until 2007	2.71% vs −3.36%	18 vs 22	−0.46
Linear filter	7.99% vs −12.70%	6 vs 34	−0.81
Rotemberg filter	4.19% vs −5.68%	6 vs 34	−0.33
Rotemberg filter, $N = 80$	3.74% vs −5.14%	29 vs 51	−0.33
Annual GDP			
Baseline	3.20% vs −4.40%	3 vs 7	−0.35
$N = 6$	3.37% vs −4.83%	0 vs 6	−0.35
$N = 20$	2.99% vs −3.55%	13 vs 7	−0.35
Until 2007	3.20% vs −4.41%	4 vs 6	−0.35
From 1929	16.69% vs −11.61%	6 vs 4	+1.00
Linear filter	7.29% vs −12.51%	2 vs 8	−0.88
Linear filter from 1929	20.50% vs −31.08%	3 vs 7	−0.91
Rotemberg filter	6.23% vs −13.50%	1 vs 9	−0.87
Rotemberg filter from 1929	16.15% vs −36.95%	1 vs 9	−1.22
Monthly industrial production			
Baseline	4.52% vs −5.90%	50 vs 70	−0.65
$N = 40$	5.48% vs −7.57%	7 vs 33	−0.65
$N = 240$	3.71% vs −4.45%	124 vs 116	−0.65
Until 2007	4.39% vs −5.58%	56 vs 64	−0.65
From 1919	11.35% vs −13.59%	54 vs 66	−0.55
Linear filter	17.03% vs −22.69%	33 vs 87	−0.52
Rotemberg filter	7.47% vs −11.23%	46 vs 74	−0.62

Notes: "Mean pos vs neg": Mean of the $N/2$ largest periods vs mean of the $N/2$ smallest periods. "# pos vs neg": Out of the N periods with largest absolute value, how many were positive and how many were negative. "Skewness": Coefficient of skewness defined as $E\left[(x-\mu)^3/\sigma^3\right]$.

For all three series in the baseline, N corresponds to a little less than 1/6 of observations, series were HP filtered and starting date is January 1949. "Quarterly GDP": $N = 40$, end date 2014:4, HP(1600)-filtered. "Annual GDP": $N = 10$, end date 2013, HP(100)-filtered. "Monthly industrial production": $N = 120$, end date 2014/02, HP(10,000)-filtered. Alternative specifications differ from respective baseline only along listed dimensions.

revenue productivity becomes wider in recessions; Bachmann and Bayer (2013) reach a similar result for innovations to the Solow residual in a dataset of German firms. Since we focus on countercyclical dispersion in levels, our closest empirical source is Kehrig (2015) and his result that cross-sectional dispersion is 2.84% higher in recessions than in the long-run average. Quantitatively, in the model dispersion will inrease by 0.48% in recessions, which is 17% of Kehrig (2015)'s estimate.

Broadly, there have been two, not mutually exclusive, approaches to explain the negative correlation of these dispersion measures with output. One fruitful strand of literature starting with Bloom (2009) investigates the effect of exogenous increases in aggregate, cross-sectional, or policy uncertainty on economic conditions. A different set of papers has considered the reverse direction of causality, studying under which conditions a bad aggregate state can cause firm-level dispersion to increase endogenously; examples include Bachmann and Sims (2012), Decker et al. (2015), and Kuhn (2015).

On the macroeconomic level, the fact that many aggregates exhibit increased volatility in recessions has been documented for many real and financial indicators of economic activity (see for example Bloom (2014)'s survey article). In the context of our model, where we focus on the variance of output growth, two recent results are Bloom et al. (2012) who find that recessions are associated with a 23% higher standard deviation of output compared to the long-run average, and Bachmann and Bayer (2013) who find a difference of around 35% between booms and recessions. In our empirical specification, looking at the periods with the largest trend deviations, we find a difference of almost 40% between expansions and recessions as documented in table 4 in section 5.2.3. Our model will generate a difference of 10.7% between booms and busts, and thus explain around a quarter of the empirically observed fluctuations in aggregate volatility quantitatively.

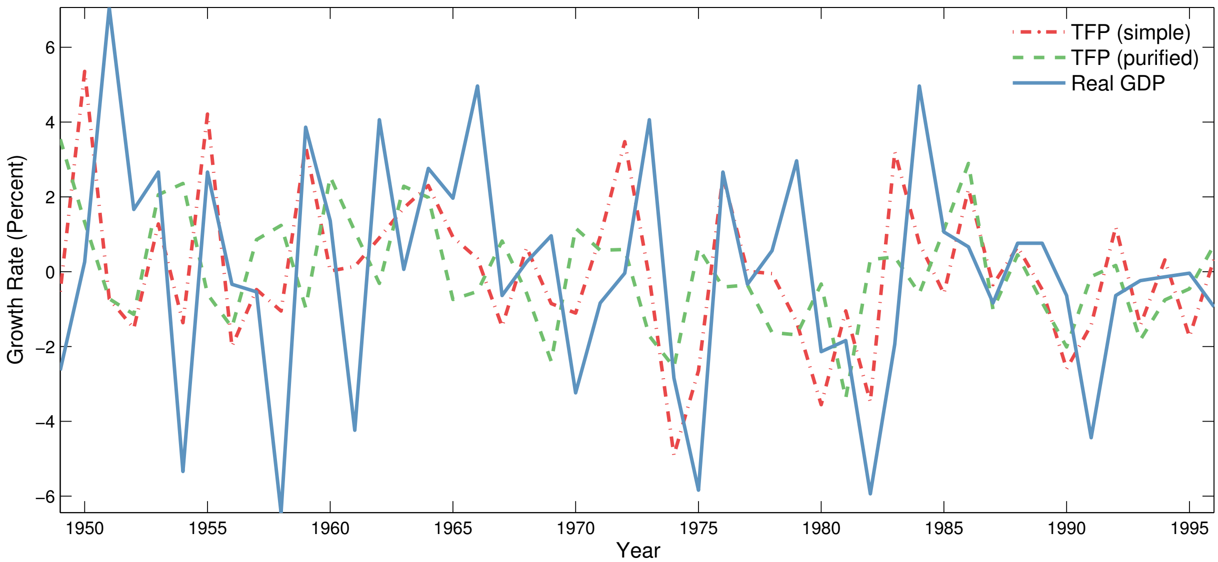
Factor utilization makes TFP look more procyclical The simple Solow residual is strongly procyclical, but much less so if corrected for factor utilization. For this stylized fact we draw on Basu et al. (2006) who discuss ways to improve the measurement of aggregate productivity. In particular, they construct a measure for aggregate technology that accounts for potentially confounding influences of returns to scale, imperfect competition, aggregation across sectors and (especially relevant here), utilization rates of factor inputs. Their uncorrected productivity measure, the Solow residual, is strongly procyclical: Correlation between output growth and simple TFP is 0.74. The corrected measure does not exhibit this strong association with aggregate production, as the correlation of purified TFP with (contemporaneous) output growth is 0.02. Figure 1 visualizes Basu et al. (2006)'s results.

Since the mechanism considered in this paper hinges strongly on the effect of adjustment in factor input utilization, we recalculate the above correlation coefficients using data provided by John Fernald[2] (see Fernald (2014)) which corrects *only* for intensity of capital and labor utilization. This allows us to check if utilization is indeed responsible for the difference in cyclicality between the simple and the purified productivity measure (or if instead the difference stems mainly from the other 'purifying' steps taken by Basu et al. (2006)). Additionally, this dataset spans 15 more years at the end of the sample and is at a quarterly frequency. Again, simple TFP is strongly procyclical, with a correlation of 0.83, whereas utilization-corrected TFP has a coefficient of −0.03.

Our takeaway from this finding is that not correcting for factor input utilization strongly

[2]Data available at www.frbsf.org/economic-research/economists/jfernald/quarterly_tfp.xls

Figure 1: GDP and TFP measures from Basu et al. (2006)



Notes: Annual series for growth rates of GDP (blue solid line), simple TFP as measured by the Solow residual (red dash-dotted line), and purified TFP as constructed by Basu et al. (2006) (green dashed line). Data from Basu et al. (2006). Correlation between output growth and simple TFP growth is 0.74, correlation between output growth and purified TFP growth is 0.02.

increases the relationship between measured aggregate productivity and output. While we do not want to weigh in on the question of which type of shocks drive business cycles, we focus on demand shocks in order to take the extreme stance of constant physical productivity. This allows us to assess how much cyclicality in *measured* TFP can be generated even when the model's correlation of output with *physical* TFP is zero. Quantitatively, in Fernald (2014)'s data the (uncorrected) Solow residual has a standard deviation of 0.87% whereas in our simulation it has a standard deviation of 0.46%.

As suggested by Wen (2004) and Basu et al. (2006), demand shocks under variable capacity utilization are a possible explanation of this fact. Bai et al. (2012) provide an alternative example in which demand shocks can show up as productivity shocks when in a search model search effort is a variable margin.

The government spending multiplier is countercyclical The cause of asymmetries in the business cycle in our model is directly relevant for the effectiveness of policy. Our contribution about capacity constraints and business cycle asymmetries thus complements the literature on cyclical fiscal multipliers. Empirically estimating the level and cyclicality of the government multiplier is difficult because of severe endogeneity issues. Nevertheless, recent empirical work on government multipliers has found significant cyclicality in fiscal multipliers, although the exact size of fluctuations is not identified very precisely. On one end of the spectrum, Auerbach and Gorodnichenko (2012c) estimate the fiscal multiplier in a regime-switching model and find large swings over the cycle ranging from around 0 during a typical boom to around 1.5 during a typical recession, albeit with large confidence intervals. Other papers identifying the multiplier in structural VARs are Mittnik and Semmler (2012) and Bachmann and Sims (2012) who also find significant cyclicality. Auerbach and Gorodnichenko (2012b), Ilzetzki et al. (2013) and Corsetti et al. (2012) all find evidence for state-dependence of the fiscal multiplier in cross-country comparisons. Nakamura and Steinsson (2014) use regional variation in the US to identify a positive relationship between the local spending multiplier and the unemployment rate. Ramey and Zubairy (2018) find that the estimated magnitude of multiplier fluctuations over the cycle is sensitive to the exact specification of the employed empirical model.

Not too much is known about the particular transmission channel through which aggregate conditions affect the multiplier. As Sims and Wolff (2017) point out, several papers model the difference between government spending when interest rates are at the zero lower bound and spending during normal times. Historically however, episodes at the zero lower bound have been relatively rare; and the empirical estimates go beyond these times indicating that the fiscal multiplier also fluctuates with the business cycle when interest rates are positive. Sims and Wolff (2017) explicitly consider multiplier fluctuations over the business cycle in a medium-scale RBC model. Their mechanism is based on households' higher willingness to supply additional labor in recessions. The model by Michaillat (2014) generates a labor multiplier, in which a search friction causes overall employment to respond stronger to government hiring in recessions than in booms.

Here, we focus on the effect of underutilized capacity which complements mechanisms in these papers. Our calibrated model implies average fluctuations of the fiscal multiplier of around 0.12, with the fiscal multiplier increasing with the size of recessions.

Driver in our model: A procyclical share of capacity constrained firms As we will lay out in the following two sections, the elements of our model that explain these four business cycle facts qualitatively are variable capacity utilization in combination with an upper bound to how much firms are willing to produce. The key mechanism in the model is that the fraction of capacity constrained firms is procyclical – this generates business cycle asymmetry, countercyclical volatility, and countercyclical fiscal multipliers. This mechanism is consistent with three empirical facts related to firms' use and building of capacity, which we now turn to. Appendix A describes the following measures in more detail.

First, average capacity utilization across many firms is procyclical. This is true for different levels of aggregation. The quarterly change in the Federal Reserve Board's measure of aggregate capacity utilization in manufacturing has a correlation with GDP growth of 0.74. Similarly, the correlation in first differences between GDP and the indices for durable goods and non-durables are 0.74 and 0.59, respectively.

This procyclicality of aggregate capacity utilization is not primarily driven by reallocation from low-utilization to high-utilization sectors a boom (or vice versa in a bust). When we look at capacity utilization data disaggregated to the sectoral level, we see that of the 21 capacity utilization indices for three-digit NAICS manufacturing industries, only 2 are not significantly procyclical with aggregate output, and the median correlation coefficient in first differences is 0.49.

Our mechanism attributes the procyclical share of constrained firms to shifts in the mean of the distribution of utilization across firms. A natural prediction of this mechanism is that the larger these common shocks to firms, the stronger the effect on the share of constrained firms, and the stronger the observed business cycle effects. To examine this prediction we combine the sectoral utilization data with data on sectoral value added for 18 NAICS manufacturing sectors. We find that a higher variance of sectoral utilization across time (i.e. larger sectoral fluctuations) is associated with stronger asymmetry in the time series of value added for that sector. The correlation of the standard deviation of utilization with asymmetry (as measured by difference in the largest trend deviations) is 0.73, and the correlations with the coefficient of skewness is -0.56. Table 5 in the appendix contains more details and the measures of asymmetry and utilization variance for all sectors.

Second, firms' investment in production capacity is procyclical. In a survey of German firms that asks specifically about the nature of firms' expenditures for capital goods, Bachmann and Zorn (2016) find that capacity investment is strongly procyclical – more so than other types of investment. Capacity investment is also highly concentrated among few firms. We interpret these findings to be consistent with the hypotheses that a) capacity utilization is procyclical at the firm level and that b) firms choose to expand their capacity when their utilization level is very high (since only few firms undertake capacity investment).

Third, we consider data from the Institute for Supply Management on the backlog of orders. In these data firms are asked whether their backlog of unfilled orders has increased, decreased or remained the same. For both manufacturing and non-manufacturing firms these changes in the backlog are strongly correlated with GDP growth. This finding is consistent with a procyclical share of firms that are capacity constrained and unable to meet demand instantaneously. Interestingly, the backlog indices for both manufacturing and non-manufacturing firms additionally are leading indicators for the *level* of GDP with a lead of 3 quarters. This finding is consistent with the possibility that firms adjust their output in response to demand for their products, possibly with a lag to adjust their productive capacity. Appendix A discusses the data sources in more detail.

3 A Simple Example

In order to illustrate the aggregate effects of capacity constraints in a framework of heterogeneous firms in this section we outline a stylized example. Firms choose their capacity before their random demand is realized. A given capacity is associated with an upper bound to production, so that if a firm's demand is greater than this bound, that firm will be constrained and produce just at capacity.

Formally, there is a continuum of ex-ante identical firms indexed by $i \in [0,1]$. Each firm can rent capital (or "capacity") k_i at a real rental price of R at the beginning of the period. A firm's production y_i is a function of *utilized* capital $\tilde{k}_i$, which for simplicity is specified as linear. Capital utilization is free here, however it is subject to the constraint that utilized capital is less than capacity, that is, $y_i = \tilde{k}_i$ s.t. $\tilde{k}_i \leq k_i$. Finally, a firm faces random demand b_i which is distributed according to a cumulative distribution function $F(b)$. The price for each firm's good is constant and normalized to 1.

A firm's sales after realization of b_i will then be $y_i = \min\{b_i, k_i\}$. The firm uses this fact when deciding on the amount of capacity to rent in order to maximize expected profits. The problem can be written as

$$\max_k -Rk + \int_0^k b \, \mathrm{d}f(b) + [1 - F(k)]\, k.$$

The resulting choice for k_i (if interior) requires $1 - F(k_i) = R$, such that for any firm there is a chance of $1-R$ that the capacity constraint binds. Denote the cutoff value for b_i at which the firm just produces at capacity as $\bar{b}_i = k_i$.

Since all firms face the same problem, they choose the same capacity $k_i = k$ and therefore face the same cutoff $\bar{b} = k$. The demand shocks b_i then induce a distribution over y_i with a mass $1 - F(\bar{b})$ concentrated at point $\bar{b}$.

We can introduce aggregate fluctuations into the example by shifting the mean of the distribution $F(b)$, which allows us to show how aggregate shocks qualitatively generate the four stylized facts outlined in the previous section. For concreteness, consider the case that demand b is distributed uniform$(0, 1)$. This implies that the optimally chosen capacity is $k_i = k = 1 - R$.

Output fluctuations Aggregate output under a uniform distribution over b between 0 and 1 is $Y = \int_0^{1-R} b \, \mathrm{d}b + R(1-R) = \frac{1}{2}(1-R^2)$. Now there is an unexpected shift in the mean by ϵ, so that b is drawn uniformly from the interval $(\epsilon, 1+\epsilon)$. Aggregate production then becomes $Y = \frac{1}{2}(1-R^2) + \epsilon(1-R) - \frac{1}{2}\epsilon^2$. By inspection, we can see that output fluctuations are asymmetric. The presence of the second-order term implies that for small values of ϵ, positive and negative output changes are about the same size, while for large values of ϵ, positive output changes are smaller than negative ones.

Fiscal multiplier: An unexpected small increase in demand can be captured by a marginal increase in ϵ. If this increase in demand represented a change in government policy, the resulting increase in output would measure the (marginal) fiscal multiplier. With the second derivative $\mathrm{d}^2 Y / \mathrm{d}\epsilon^2 = -\epsilon$ an additional small increase in aggregate demand affects output less, the higher aggregate demand already is.

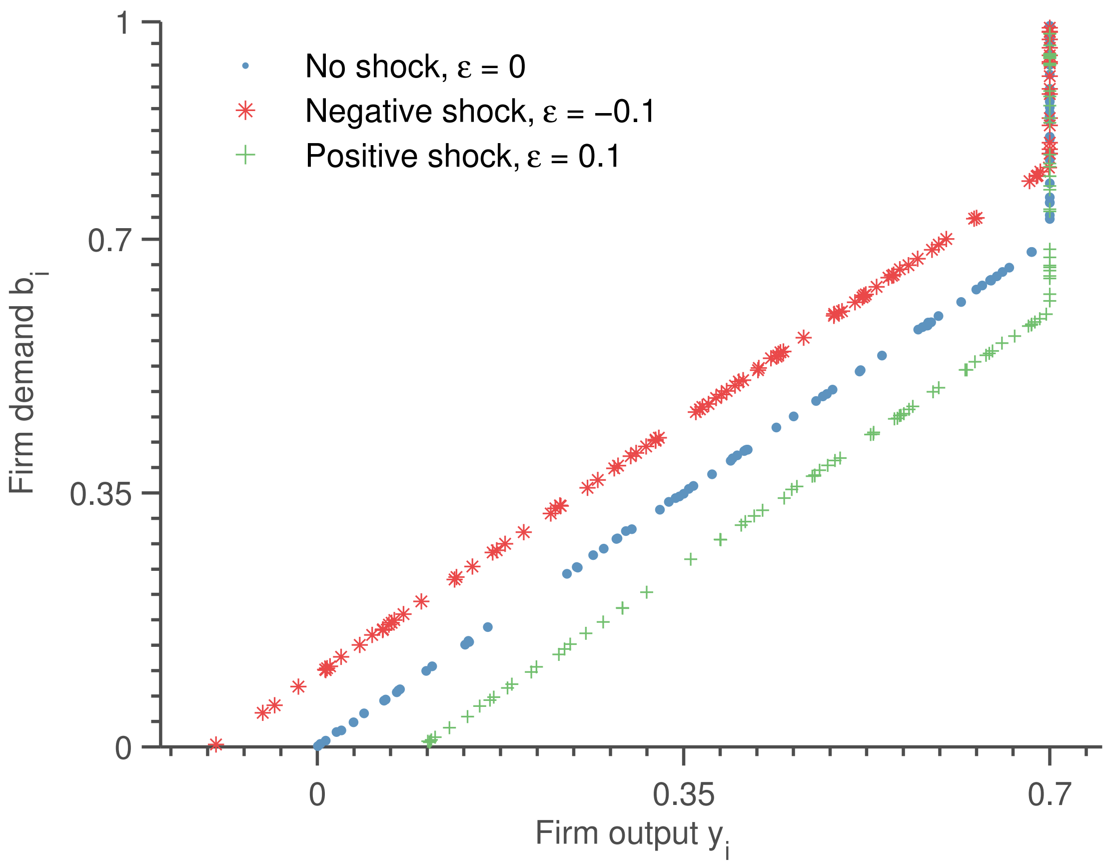
While the government multiplier and asymmetry in output are therefore closely related, they are not quite measuring the same thing. The difference between a large boom and recession is given by the average effect of an increase in demand (that is, the difference in output *between* aggregate states), while the multiplier is determined by the marginal effect (that is, the effect of a small demand shock on output *at* different aggregate states).

Figure 2 displays the mapping from demand shocks b_i into output y_i for an interest rate of 0.3 such that the implied capacity constraint is at 0.7. The three sets of points represent the case without aggregate shock ($\epsilon = 0$) as well as aggregate shocks of $\epsilon = \pm 0.1$.

Cross-sectional and aggregate volatility: The example also illustrates that aggregate fluctuations affect differences between firms. An individual firm's profitability can be measured as $y_i / k_i = y_i/(1-R)$. Since the factor input cost R is the same for all firms this means that the relative cross-sectional variance in profitability at any point is equal to the relative variance in output. While the analytic expression for $\mathrm{Var}(y_i)$ as a function of ϵ is somewhat involved, the intuition is straightforward: the greater the mass of firms at the capacity constraint, the smaller the variance in profitability of the overall distribution. In the extreme case of a very large negative shock (corresponding to $\epsilon < -0.3$ in the example), no firm is capacity constraint and thus dispersion is greatest.

In this stylized example one needs an additional assumption in order to see that the growth rate of output as a measure of aggregate volatility varies with the state of the economy ϵ. In particular, we need to prevent firms from being fully flexible in adjusting their choice of k between periods. Imagine therefore that firms have a fixed capacity level for two periods. If in the first period ϵ is positive, then a further shock in the second period will tend to have relatively small effects, because the relatively large mass of firms at their constraint will not change production – the intuition here is analogous to why fiscal multipliers are smaller in a boom. In the full model of section 4 there will be a price adjustment friction that takes the role of giving persistence to

Figure 2: Distribution of y_i in numerical illustration



Notes: The figure plots simulated output levels y_i (X-axis) for a sample of 100 firms depending on their respective realized demand b_i (Y-axis). Blue •: no aggregate shock, firm output uniformly distributed between 0 and 0.7, and a mass point at 0.7. Green +: For a positive demand shock $\epsilon = 0.1$, additional firms get pushed into their capacity constraint. Output expands less than proportionally, dispersion in output (and profitability) decreases, aggregate capacity utilization and Solow residual increase. Red $*$: The opposite is true for a negative demand shock $\epsilon = -0.1$. The left tail of the distribution becomes wider and the mass of firms at capacity decreases.

aggregate demand shocks.

Measured aggregate productivity: Simple measured aggregate TFP is $Y/K = \frac{1}{2}(1+R) + \epsilon - \frac{1}{2}\epsilon^2/(1-R)$, hence it increases with ϵ due to more intensive use of installed capacity. Measured aggregate TFP is hence endogenously procyclical while TFP corrected for utilization is trivially given by $Y/\tilde{K} = 1$ by definition of the production function.

This example illustrates the basic mechanics with which capacity constraints can qualitatively generate deep recessions along with meek booms, countercyclical fiscal multipliers and a more dispersed productivity distribution in recessions. All of these features arise from a simple shock structure that is perfectly symmetric over time and across firms.

4 Model

We now embed capacity constraints in a New-Keynesian model of aggregate demand shocks to look at the effects in general equilibrium. While the intuition from the previous section about their qualitative implications fully carries through, only a general-equilibrium model will be able to inform us about the size of asymmetries generated by capacity constraints quantitatively.

The main difference relative to the example in the previous section is that the capacity constraint arises endogenously due to convex capital utilization costs. This type of utilization cost can be justified by empirically relevant features such as overtime pay or increased depreciation when operating machinery more intensively. By introducing capacity constraints in this way, a firm's maximal production is given by its willingness to supply goods rather than an assumed technological constraint. To this end, firms not only choose their capacity, but also their goods price at the beginning of the period before any shocks are realized.

The full model also includes the following standard features: Labor constitutes a second flexible factor of production in addition to utilized capital, and an individual firm's demand now comes from a final goods aggregator. Finally, there is a central bank setting nominal interest rates.

There are several reasons why we model firms as setting their price in advance. First, it keeps the model tractable since all firms face the same environment at the time of their decision and hence choose the same price. Second, it will allow us to endogenize capacity constraints as the quantity firms are willing to supply at the set prices. Third, in this context it provides a convenient way of introducing price rigidities which allow preference shocks to affect output through changes in relative prices, as is usual in New Keynesian models.[3]

In order for firm supply to constitute an upper bound to production we will specify that, when supply and demand do not coincide at the set price, quantity traded is given by the minimum of supply and demand, and hence determined by the 'short' market side. This rule differs in particular from an alternative in which the price setter is required to satisfy the other market side's demand or supply at the given price. Fagnart et al. (1999) use a similar setup and discuss the implications for planned and traded quantities in more detail.

[3]Additionally, Kuhn (2015) shows that in general it is important to model firms' pricing behavior explicitly when considering cross-sectional profitability measures: Differences in pricing can prevent firms' profitability from tracking their physical productivity, as highlighted by Foster et al. (2008).

4.1 Timing

The timing within a period is as follows:

1. Households enter a period t with an amount of aggregate capital K_t. At the beginning of the period, before any shocks are realized, a capacity rental market opens where households supply K_t and firms rent their capacity for this period, k_{it}. Simultaneously, firms choose their price p_{it}. (Later in equilibrium, because all firms are the same at the beginning of the period, $k_{it} = K_t$ and $p_{it} = p_t$.)

2. All idiosyncratic and aggregate shocks are realized.

3. The remaining markets open: Firms make their decisions about labor demand and capacity utilization; households decide on their labor supply and desired savings in capital and bonds. Households also receive firm profits and pay taxes. The monetary authority sets the nominal interest rate as a function of inflation. The period ends.

4.2 Final goods aggregator

The final good Y is assembled from a continuum of varieties indexed by $i \in [0, 1]$ according to a standard CES function with parameter σ measuring the elasticity of substitution between intermediate goods

$$Y = \left(\int b_i^{\frac{1}{\sigma}} y_i^{\frac{\sigma-1}{\sigma}} \, \mathrm{d}i \right)^{\frac{\sigma}{\sigma-1}} .$$

The weights $\{b_i\}$ are realizations of *iid* random variables with mean 1.

The perfectly competitive final goods aggregator takes intermediate goods prices as given. It has a nominal budget of $I \geq \int p_i y_i \, \mathrm{d}i$, where p_i is an intermediate variety's nominal price. The aggregator also takes into account the capacity constraint that limits the supply of some varieties. Denoting this upper limit[4] by $\bar{y}$, it therefore has to consider a continuum of inequality constraints $y_i \leq \bar{y} \ \forall i$. The problem can then be expressed as

$$\max_{\{y_i\}, \lambda, \{\mu_i\}} \left(\int b_i^{\frac{1}{\sigma}} y_i^{\frac{\sigma-1}{\sigma}} \, \mathrm{d}i \right)^{\frac{\sigma}{\sigma-1}} + \lambda \left(I - \int p_i y_i \, \mathrm{d}i \right) + \int \mu_i \left(\bar{y} - y_i \right) \mathrm{d}i.$$

After taking first-order conditions (see appendix B), one has

$$y_i^d = b_i \frac{I_U P_U^{\sigma-1}}{p_i^{\sigma}}$$

with $I_U \equiv \int_{y_i < \bar{y}} p_i y_i \, \mathrm{d}i$ the budget spent on unconstrained varieties and $P_U^{1-\sigma} \equiv \int_{y_i < \bar{y}} p_i^{1-\sigma} \, \mathrm{d}i$ a price index over unconstrained varieties. Intuitively, y_i^d is the amount of good i that the aggregator would like to buy if that good were in unlimited supply, given the purchases of all other varieties. As pointed out by Fagnart et al. (1999), of course the aggregator knows that if

[4] In equilibrium the upper bound $\bar{y}$ is equal to the intermediates' maximum supply dictated by costly capacity utilization y^s and will indeed be the same for all firms. One could solve the aggregator's problem more generally using a variety-specific $\bar{y}_i$ at the cost of more notation, but considering a $\bar{y}$ constant across varieties is enough here.

for some particular variety $y_i^d > \bar{y}$ it will not be able to purchase this entire quantity and y_i^d is therefore the demand for *unconstrained* varieties only. For the constrained inputs the aggregator can only purchase the quantity supplied $\bar{y}$. If the constraints bind for some inputs the aggregator will take this into account in the demand for unconstrained varieties y_i^d via the budget and price index over the latter I_U and P_U.[5]

4.3 Firms

We solve the firm's problem backwards: We first determine a firm's optimal utilization and labor input given its realization of b_i and chosen capacity and price, and then the optimal k_i and p_i choices that maximize expected profits.

Technology The intermediate goods firms' production function is $y = \tilde{k}^{\alpha} l^{1-\alpha}$, where l is the hired labor input.[6] There is a quadratic real cost of utilizing capital which depends on the utilization rate $\tilde{k}/k$ and total capacity k given by

$$\mathrm{cu}\left(\frac{\tilde{k}}{k}, k\right) = \frac{\chi}{2}\left(\frac{\tilde{k}}{k}\right)^2 k.$$

This formulation ensures that the utilization costs scale linearly with k and hence the optimal utilization rate is going to be independent of firm size. Note that we do not impose a constraint on production at this stage in the sense that we, theoretically, allow an arbitrarily intensive utilization $\tilde{k}/k$. However, as we will show below, because the cost increases more and more steeply as firms increase utilization, they will choose an upper limit to production that they do not exceed.

There is a Rotemberg-type quadratic real cost of adjusting the nominal price p depending on the relative change p/p_{-1} through

$$\mathcal{C}\left(\frac{p}{p_{-1}}\right) = \frac{\xi}{2}\left(\frac{p}{p_{-1}} - 1\right)^2.$$

We employ this cost because it is the simplest possible way of introducing persistent nominal rigidities — its tractability in the context of this model stems from the fact that all firms choose the same price in equilibrium. Additionally, the price adjustment cost adds an intertemporal dimension to the firm's problem and thus generates some internal propagation of shocks (if $\xi = 0$ the firm's problem is reduced to an infinite sequence of one-shot problems).

Cost function The cost function describes the cheapest way for a firm to produce a fixed output level y given the marginal cost of the input factors which are in turn determined by the

[5]Fagnart et al. (1999) note, using the terminology of Clower (1967), that y_i^d hence constitutes the *effective* demand for variety i since it takes into account the rationing on other markets.

[6]In this section the firm index i is suppressed to save notation. It will reappear in the section on aggregation below.

level of capacity k and the real wage w. It is given by

$$C(y) = \min_{\tilde{k},l} wl + \frac{\chi}{2}\left(\frac{\tilde{k}}{k}\right)^2 k$$
$$\text{s.t. } \tilde{k}^{\alpha} l^{1-\alpha} \geq y.$$

The first-order conditions give optimal input factor quantities as

$$\tilde{k} = \left(\frac{1-\alpha}{\alpha}\frac{\chi}{w}k^{-1}\right)^{-\frac{1-\alpha}{2-\alpha}} y^{\frac{1}{2-\alpha}} \tag{1}$$

$$l = \left(\frac{1-\alpha}{\alpha}\frac{\chi}{w}k^{-1}\right)^{\frac{\alpha}{2-\alpha}} y^{\frac{2}{2-\alpha}}. \tag{2}$$

Given the Cobb-Douglas structure, the optimal input ratio is proportional to the marginal costs of the inputs $\tilde{k}/l = \frac{\alpha}{1-\alpha} w \left(\chi\tilde{k}/k\right)^{-1}$. The cost function follows as

$$C(y) = \frac{2-\alpha}{2\alpha}\left[\chi^{\alpha}\left(\frac{\alpha}{1-\alpha}w\right)^{2(1-\alpha)} y^2 k^{-\alpha}\right]^{\frac{1}{2-\alpha}}.$$

Supply function and cutoff $\bar{b}$ Because the firm incurs convex utilization costs, there will be some cutoff quantity of output more than which the firm will find unprofitable to produce. Since output is increasing in the level demand, the cutoff output quantity will be associated with a cutoff level of the demand shock; here we solve for both output and demand cutoffs, labeled y^s and $\bar{b}$, respectively.

The firm considers the level of output y^s that maximizes profits given its price and cost function, but ignoring its level of demand. In other words, the firm thinks about how much it would produce if demand for its variety was infinite. With $\mathcal{P}$ denoting the nominal price of the final good, it considers its maximal operating profits

$$\max_y \frac{p}{\mathcal{P}} y - C(y)$$

which is solved by[7]

$$y^s(p,k) = \left(\frac{\alpha}{\chi}\right)\left(\frac{1-\alpha}{w}\right)^{\frac{2(1-\alpha)}{\alpha}}\left(\frac{p}{\mathcal{P}}\right)^{\frac{\alpha+2(1-\alpha)}{\alpha}} k. \tag{3}$$

The convexity of the capital utilization cost function ensures that supply given $w, p/\mathcal{P}$ and k is finite.

As mentioned above, there are no contractual arrangements that would require firms to produce more than they desire, so that actual quantity traded y is given by

$$y(p,k,b) = \min\left\{y^d(p,b), y^s(p,k)\right\}. \tag{4}$$

[7]We slightly abuse notation in the following, writing y^s, y^d, $\bar{b}$ and profits π as explicit functions to indicate clearly upon which variables they depend.

This defines a cutoff value $\bar{b}$ for the idiosyncratic demand shock at which $y^s(p,k) = y^d(p,b)$ as

$$\bar{b}(p,k,b)\,\frac{I_U P_U^{\sigma-1}}{p^{\sigma}} \equiv \left(\frac{\alpha}{\chi}\right)\left(\frac{1-\alpha}{w}\right)^{\frac{2(1-\alpha)}{\alpha}}\left(\frac{p}{\mathcal{P}}\right)^{\frac{\alpha+2(1-\alpha)}{\alpha}} k.$$

Any firm with $b > \bar{b}$ will be constrained due to costly utilization, while firms with $b < \bar{b}$ just satisfy demand. An algebraically useful implication is that y^d can be written as

$$y^d(p,b) = (b/\bar{b})y^s. \tag{5}$$

Operating profits, expected profits, and value function Depending on realized demand b, operating profits as a function of p and k are given by

$$\pi(p,k,b) = \begin{cases} \frac{p}{\mathcal{P}} y^d(p,b) - C\left(y^d(p,b);k\right) & \text{if } b \le \bar{b} \\ \frac{p}{\mathcal{P}} y^s(p,k) - C\left(y^s(p,k);k\right) = \frac{p}{\mathcal{P}} y^s(p,k)\frac{\alpha}{2} & \text{if } b > \bar{b} \end{cases}$$

At the beginning of the period the firm can compute expected profits by integrating over b:

$$E\left[\pi(p,k,b)\right] = \int_0^{\bar{b}(p,k,b)} \frac{p}{\mathcal{P}} y^d(p,b) - C\left(y^d(p,b);k\right)\,\mathrm{d}f(b) \; + \int_{\bar{b}(p,k,b)}^{\infty} \frac{p}{\mathcal{P}} y^s(p,k)\frac{\alpha}{2}\,\mathrm{d}f(b).$$

It can now choose its price and capacity in order to maximize expected operating profits minus the rental cost of capacity and the (expected discounted sum of future) costs of price adjustment. In fact, only the price adjustment cost makes the firm problem truly dynamic. The problem is summarized in the firm's value function

$$V(p_{-1}) = \max_{p,k} E\left[\pi(p,k,b)\right] - \left[R - (1-\delta)\right]k - \frac{\xi}{2}\left(\frac{p}{p_{-1}} - 1\right)^2 + \beta E\left[V(p)\right]. \tag{6}$$

The basic tradeoffs for the two choice variables are reflected in the first-order conditions which are provided in appendix C. Intuitively, investing in one more unit of capacity k costs the net real rental rate $R_t - (1-\delta)$ today, but will lower the cost of utilizing capital next period, increasing expected profits. Adjusting the price p has two opposing effects on expected revenue: It will increase marginal revenue per unit sold, whether the firm ends up constrained or unconstrained. The cost is that demand will be reduced, which will be a reduction in quantity sold if the firm produces below the constraint. In addition to these effects on revenue the firm has to take into account the costs of price adjustment, both in the current period as well as the expected costs that will result from future expected adjustments. The first-order condition for the optimal price hence contains terms relating to these four marginal effects on profits.

4.4 Households

There is a price-taking representative household. She maximizes lifetime utility given by

$$E\left[\sum_{t=0}^{\infty}\beta^t\left(\log C_t - \varphi_t\frac{L_t^{1+\varepsilon}}{1+\varepsilon}\right)\right]$$

where C_t is consumption and L_t is hours worked in period t. This formulation of preferences is consistent with existence of a balanced growth path.

As usual in New Keynesian models, changes in the household's period utility are the source of aggregate demand shocks driving the economy. There are two standard ways to model these shocks: When the representative household decreases her demand for consumption today, she must either increase her demand for leisure, or her demand for future consumption. Here, we use a random weight φ_t shifting the relative preference of consumption and leisure to serve as the aggregate demand shock. However, in appendix E we show that the model yields very similar results when using a discount factor shock that affects the demand for current consumption relative to future consumption.

Besides working, the household also earns income from renting capital K_t to firms as well as from holding one-period bonds issued by the central bank. Her real bond demand in t is denoted with S_t, and central bank pays a nominal interest rate of $\mathcal{R}_t$ on these bonds. The household also collects all profits from firms $\widetilde{\pi}_t \equiv \int \pi_{it} - [R - (1-\delta)]\, k_{it} - \frac{\xi}{2}\left(\frac{p_{it}}{p_{i,t-1}} - 1\right)^2 \mathrm{d}i$ and finances any government spending with a lump-sum transfer of G_t. Combining all these payments in units of final goods yields her (real) flow budget constraint

$$C_t + S_t + K_{t+1} = \frac{\mathcal{R}_{t-1}}{\Pi_t} S_{t-1} + R_{t-1}K_t + w_t L_t + \pi_t - G_t.$$

The variable $\Pi_t \equiv \mathcal{P}_t/\mathcal{P}_{t-1}$ denotes inflation.

Her optimality conditions are the labor supply equation

$$w_t = \varphi_t L_t^{\varepsilon} C_t, \tag{7}$$

the Euler equation

$$\frac{1}{C_t} = \beta \mathcal{R}_t E\left[\frac{1}{C_{t+1}\Pi_{t+1}}\right], \tag{8}$$

as well as a no-arbitrage condition between nominal assets and capital

$$\mathcal{R}_t E\left[\frac{1}{C_{t+1}\Pi_{t+1}}\right] = E\left[\frac{R_t}{C_{t+1}}\right]. \tag{9}$$

4.5 Central bank and government

The central bank sets nominal interest rates in accordance with a simple Taylor rule such that inflation fluctuates around its long-run mean of zero:

$$\log(\mathcal{R}_t) = \log(1/\beta) + CB_{rf}\log(\Pi_t). \tag{10}$$

The parameter CB_{rf} determines how strongly the central bank reacts to inflation.

A government undertaking fiscal policy constitutes the second part of the public sector. It can buy goods G_t from the final goods firm which it then consumes. It runs a balanced budget by collecting lump-sum taxes G_t from the household. Since the government's only purpose is to allow us to assess the size of the fiscal multiplier, we fix $G_t = 0$ for all t.

4.6 Aggregation and equilibrium

Firms use their first-order necessary conditions from maximization of their value (6) to determine optimal price and capacity (p_{it}, k_{it}) at the beginning of the period. Since, before realization of period t shocks, all firms share the same state variables, they choose identical prices and capacities such that $p_{it} = p_t$ and $k_{it} = k_t$ $\forall i$. Additionally, firms' decisions about utilization and labor in (1) - (2) and quantity traded in (4) are monomial in $\min \left\{ b_i/\bar{b}, 1 \right\}$. This makes integration over i straightforward and gives aggregate capital utilization costs and labor demand as

$$CU = \frac{\alpha}{2} \frac{p}{\mathcal{P}} y^s \left(\int_0^{\bar{b}} \left(\frac{b}{\bar{b}} \right)^{\frac{2}{2-\alpha}} df(b) + \left[1 - F\left(\bar{b} \right) \right] \right) \tag{11}$$

$$L^d = \frac{1-\alpha}{w} \frac{p}{\mathcal{P}} y^s \left(\int_0^{\bar{b}} \left(\frac{b}{\bar{b}} \right)^{\frac{2}{2-\alpha}} df(b) + \left[1 - F\left(\bar{b} \right) \right] \right) \tag{12}$$

and final goods supply using the aggregator's production function as

$$Y = \bar{b}^{\frac{1}{\sigma-1}} y^s \left\{ \left[\int_0^{\bar{b}} \frac{b}{\bar{b}} df(b) + \int_{\bar{b}}^{\infty} \left(\frac{b}{\bar{b}} \right)^{\frac{1}{\sigma}} df(b) \right] \right\}^{\frac{\sigma}{\sigma-1}}. \tag{13}$$

In equilibrium, the final goods price $\mathcal{P}_t$ as well as the producer price p_t are not determined in levels. These prices, however, only matter relative to each other or their respective values from the previous period. We therefore define the real price of intermediate goods as $\overline{rp}_t = p_t/\mathcal{P}_t$, inflation as $\Pi_t = \mathcal{P}_t/\mathcal{P}_{t-1}$, and producer price inflation as $\Pi_t^{\text{ppi}} = p_t/p_{t-1}$. These relative prices in turn are related according to

$$\Pi_t^{\text{ppi}} = \Pi_t \frac{\overline{rp}_t}{\overline{rp}_{t-1}} \tag{14}$$

as can easily be derived from their definition.

Equilibrium then is defined in the usual way using agents' optimality conditions and clearing of aggregate markets. Notably, the clearing of *aggregate* markets is unaffected by the fact that predetermined prices prevent intermediate goods markets from clearing. Specifically, we define as equilibrium a sequence of prices $\left\{ R_t, \mathcal{R}_t, w_t, \overline{rp}_t, \Pi_t, \Pi_t^{\text{ppi}} \right\}_{t=0}^{\infty}$, and of quantities $\left\{ Y_t, C_t, CU_t, L_t^d, y_t^s, k_t, K_t, L_t \right\}_{t=0}^{\infty}$ and cutoffs $\left\{ \bar{b}_t \right\}_{t=0}^{\infty}$ that satisfy the firms' two optimality conditions derived from (6), their supply (3), aggregate factor demands and final goods supply (11)-(13), the household's optimality conditions (7)-(9), the Taylor rule (10), the definition of producer price inflation (14), as well as market clearing for labor and capital, an aggregate resource constraint, and the aggregator's zero-profit condition. Note that for this definition we have already imposed $y^s = \bar{y}$.

Appendix C collects these equilibrium conditions.

5 Calibration and results

5.1 Calibration

In the following we simulate the model and show that the qualitative results from the example hold up in general equilibrium. Table 2 summarizes the calibration of model parameters in

Table 2: Baseline Calibration

Parameter	Value	Meaning	Calibration
		Standard parameters	
α	$\frac{1}{3}$	Capital share	Standard
β	0.99	Hh discount factor	Standard (quarterly)
δ	0.026	Capital depreciation	Standard (quarterly)
ε	$\frac{1}{2}$	Inv. Frisch elas. labor	Standard
σ	6	E. of S. intermediates	Literature: $\sigma \in [4, 10]$
ρ_φ	0.9	Shock persistence	Standard (quarterly)
σ_φ	0.0044	Shock variance	$\text{sd}(Y) = 1.82\%$
ξ	75	Scale price adj. cost	Ireland (2001)
CB_{rf}	1.75	Taylor rule	Sims and Wolff (2017)
		Model-specific parameters	
χ	0.55	Scale utiliz. cost	sd(capacity util.) $= 3.9\%$
σ_b	0.67	sd idiosync. shocks	$\text{sd}(\Delta\text{TFP}_i) = 0.185$

two groups: The first group contains parameters that have direct empirical interpretations or standard values in the literature, whereas the second group consists of parameters that are specific to the model.

The first group of parameters is set to conventional values found in the literature. Capital's share of income α is set to 1/3, and capital depreciation is $\delta = 2.6\%$ implying an annual rate of 10%. Based on estimates of the average mark-up between around 10% and 30% , the macroeconomic literature uses values for the elasticity of substitution between goods σ between 4 as for example in Bloom et al. (2012) and 10 as for example in Sims and Wolff (2017). We hence choose an interior value of 6. Households have a discount factor of $\beta = 0.99$ such that the annual steady-state interest rate is around 4%. The parameter ε set to 1/2 targets a Frisch elasticity of labor supply of 2 which is also a standard value in macroeconomic models. The aggregate shock follows an AR(1) process in logs such that $\log \varphi_t = \rho_\varphi \log \varphi_{t-1} + u_\varphi$ where u_φ is a mean-zero normal random variable with variance σ_φ^2. We set the persistence parameter $\rho_\varphi = 0.9$. The standard deviation of innovations $\sigma_\varphi = 0.0044$ is chosen to match the empirical standard deviation of quarterly postwar US GDP of 1.82% when detrended with an HP(1600) filter. The price adjustment cost parameter ξ is set to 75, corresponding to the estimate in Ireland (2001). The coefficient measuring how the central bank reacts to inflation is set to 1.75 as in Sims and Wolff (2017).

The second group of parameters describes the utilization cost function and the variance of idiosyncratic shocks, both of which affect (the distribution of) firm-level capacity utilization. We assume the distribution of the *iid* idiosyncratic shock b_i to be log-normal. The variance of this demand shock σ_b will directly affect the variance of firm-level utilization over time, and –importantly for the model's mechanism– the chance that a firm becomes capacity constrained. Ideally, therefore, we would match this idiosyncratic shock variance to an empirical estimate of the share of capacity constrained firms; however we are unaware of such a direct estimate. Since in our model a firm's utilization rate is closely linked to its measured profitability, we hence rely on empirical estimates of the size of innovations to firm profitability to calibrate σ_b. In particular, both Syverson (2011) and Ilut et al. (2016) find a standard deviation of innovations to the log of firm profitability of around 0.185, which in the model corresponds to $\sigma_b = 0.67$. [8]

Finally, we calibrate χ to match the standard deviation[9] in average capacity utilization over time. In particular, the cyclical component of the Federal Reserve Board's index of capacity utilization in industrial production over our baseline period of 1949 to 2014 has a standard deviation of 3.9%. This index constitutes an estimate of current production relative to the maximum possible output given currently established production capacity. In the model this most closely corresponds to Y/y^s, and we hence choose $\chi = 0.55$ so that the standard deviation of Y/y^s equals 3.9%. In section 5.2.6 we discuss the effects of both σ_b and χ on the dynamic behavior of the model.

A central feature of the model is that the fluctuating share of capacity constrained firms generates extra concavity in aggregate production. This causes effect sizes to increase with the magnitude of aggregate fluctuations. For example, if aggregate shocks are small, the response of output to a positive shock is similar to the response to a negative shock. Relative differences between booms and recessions increase as the aggregate shock becomes larger. Model results are therefore somewhat sensitive to the variance σ_φ^2 of innovations to φ. In the baseline calibration we take a conservative stance by detrending the empirical GDP series with an HP-1600 filter, which implies a relatively moderate standard deviation of 1.8% for its cyclical component. If, on the other hand, the underlying growth trend of the empirical series were better described by a linear trend, then the time-series standard deviation of the cyclical component is 4.7%, which significantly amplifies output asymmetry and the fiscal multiplier in our results. We consider this alternative calibration in section 5.2.6.

5.2 Results

5.2.1 Impulse response functions

We simulate business cycles by a shock to the household's preference weight φ_t governing her relative taste for consumption and leisure. While we acknowledge many other possible shocks that can cause aggregate fluctuations, as discussed above we focus on this preference shock as a simple way to generate demand-side effects through distorted relative prices, which in turn allows us to assess how much movement in the measured Solow residual is generated even by a non-technology shock. The results are very similar if we use a discount factor shock to generate fluctuations in demand, as documented in appendix E. (In section 5.2.6 and appendix F we additionally compare the model to a flexible price version under technology shocks to assess the effect of nominal rigidities and demand shocks.)

The model is then solved with a second-order approximation around the non-stochastic steady state using the software package Dynare (see Adjemian et al. (2011)). An approximation of at least second order is necessary since we want to account for the non-linearities generating differences between positive and negative shocks. Under linearization these differences would be lost.

[8]This implies that in the model's non-stochastic steady-state the fraction of capacity-constrained firms is 2.2%, which seems consistent with Bachmann and Zorn (2016)'s finding that capacity *investment* is highly concentrated among a small share of firms.

[9]As in other models with variable capacity utilization the scale parameter χ does not directly affect the steady-state capacity utilization rate.

Figure 3 displays simulated impulse response functions following a 1-standard-deviation increase in the leisure preference of households φ. For approximations of order higher than 1 the effect size of a shock will in general depend on the state of the economy at the time of impact. The standard way of computing impulse responses in such a case is through simulation, which approximates an 'average' effect of the shock across many simulated states.[10]

Most notable is the strong reaction to the shock on impact in period 1. With prices set in advance and hence fully sticky for one period, the standard "New Keynesian" effect of demand shocks via relative prices is fully present in period 1. What remains of the shock in periods 2 and later is driven by the fact that firms' price adjustment costs prevent a full alignment of relative prices, as well as effects through households' labor and capital supply.

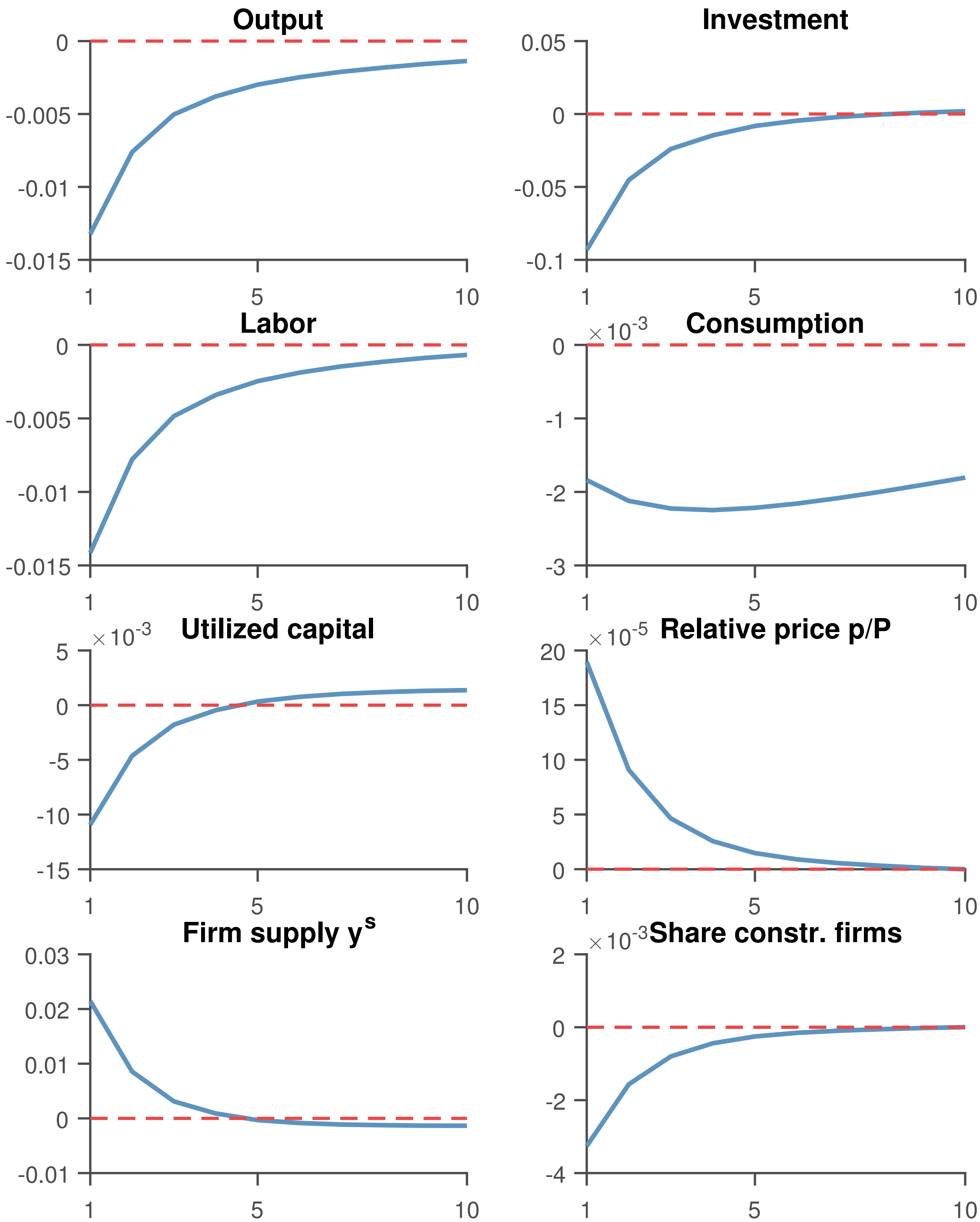
As expected, capacity utilization drops along with aggregate output. The share of firm below their capacity constraint $F\left(\bar{b}\right)$ decreases as well. This is not only due to the reduction in demand for intermediates, but also due to the increase in firms' willingness to supply their respective variety: With nominal intermediate goods prices fixed at p, the decrease in the aggregate price level $\mathcal{P}$ leads to a temporarily high relative price.

Impulse response functions can also visualize the nonlinearities of the model, which matter for output asymmetry and aggregate volatility. Figure 4 shows the difference in responses for expansionary and recessionary shocks, respectively. For the same absolute size, large negative shocks have a stronger impact on real variables than large positive shocks. The reason behind this is that the number of constrained firms responds more strongly following a positive shock, and those firms, once at the constraint, can not adjust their production further. Because the *cumulative* number of firms that become constrained matters, asymmetry depends on the shock size: For small disturbances the response is close to symmetric because the share of constrained firms is small as well.

Conversely, in figure 5 we hold fixed the size and sign of the shock but consider different states in which the impulse hits the economy – specifically, we compare what happens if an expansionary shock arrives when the economy is in a recession versus when it is already in a boom. If the positive shock hits in a recession, real variables respond more strongly to the shock than if it hits in a boom (shown are output and consumption, but the same is true for investment, labor, capacity, and utilization). The reason is again in the number of constrained firms becoming constrained: In a boom, many firms are already at or near full capacity, which despite additional demand does not allow them to increase their output (much) more. In contrast, during a downturn most firms have idle capacities and will be able to ramp up production following an increase in demand. Because quantities can not respond as much in a boom there is a stronger effect on relative prices – the price of the aggregate good increases more strongly, lowering the relative price or the intermediate input. The stronger response of real variables in recessions generates countercyclical aggregate volatility.

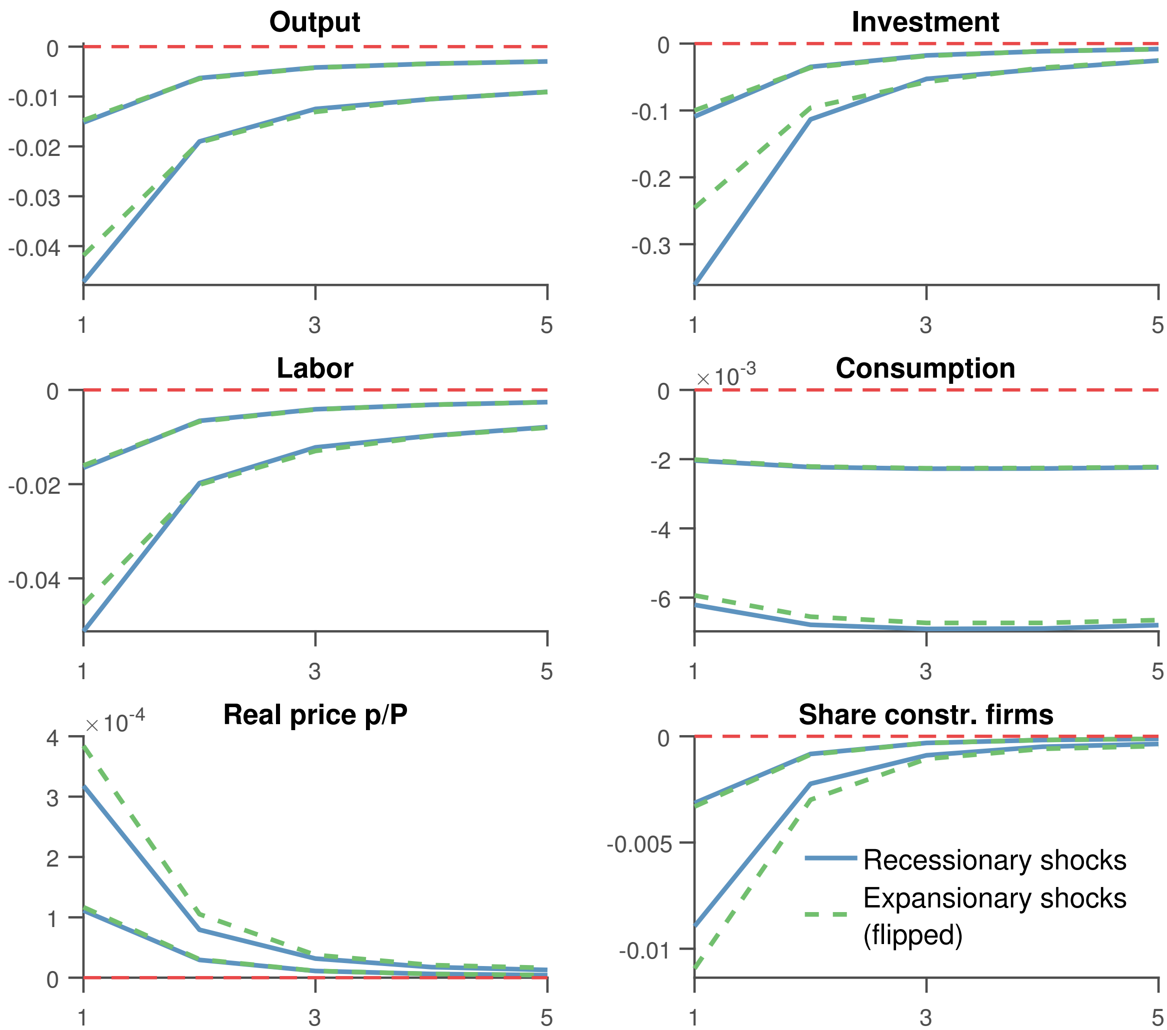
[10] More precisely, one chooses an appropriate 'burn-in' period and a large number I of simulations indexed by i. For each simulation one simulates the model forward such that the model economy is at some random point $S_{i,0}$ of its ergodic state set. Next, one draws a sequence of aggregate shocks $\{Z_{i,t}\}_{t=1}^{T}$ of length T equal to the desired time horizon of the impulse response, and simulates the model forward twice starting from $S_{i,0}$: Once, using only the shocks $\{Z_{i,t}\}$, and once using the same shocks where for $Z_{i,1}$ an additional 1-sd shock the exogenous state variable has been added. The simulated impulse response is then just the difference between the two simulations, averaged over all I repetitions. For more details see, for example, Adjemian et al. (2011).

Figure 3: Impulse Response Functions



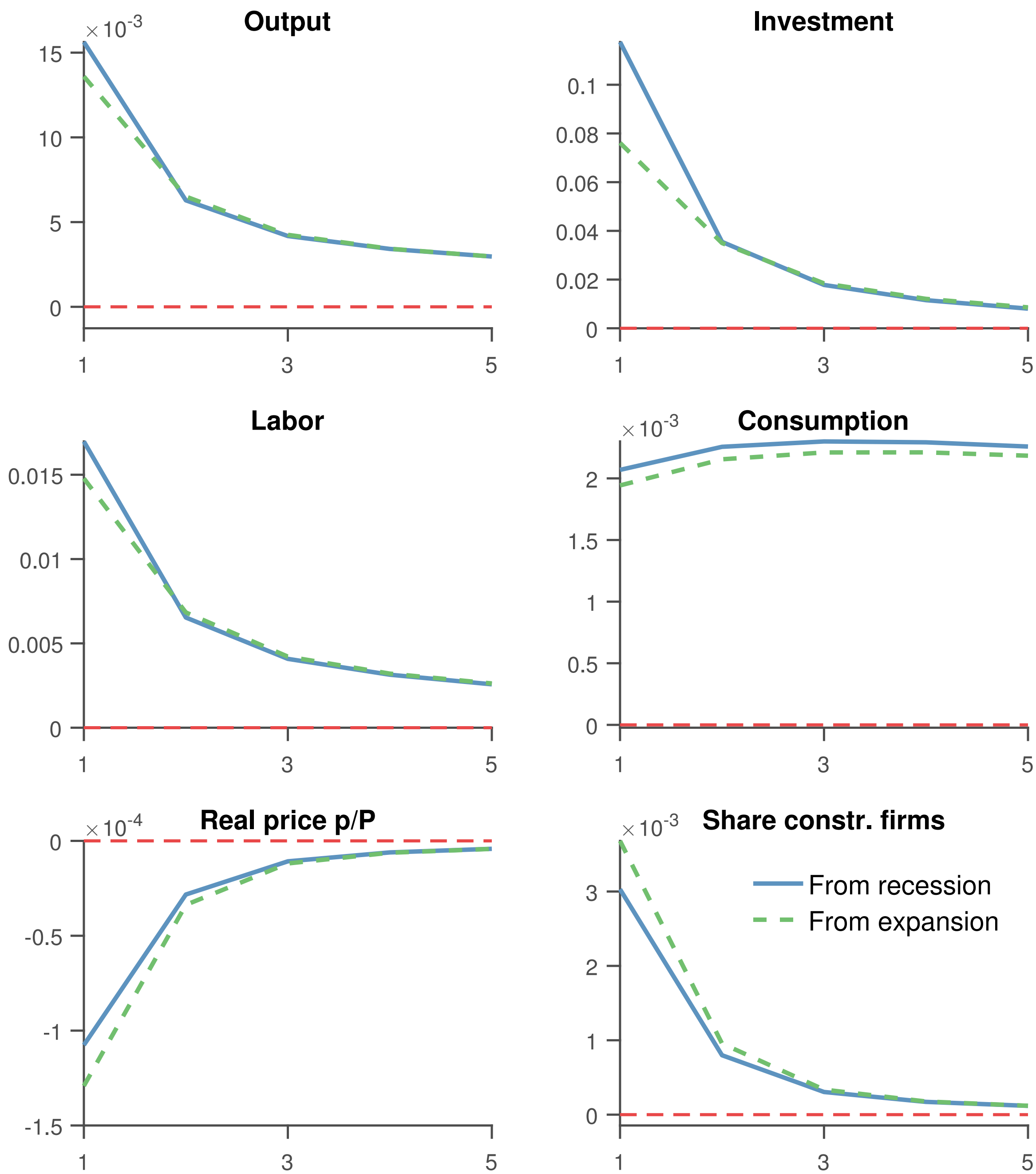
Notes: Simulated impulse response functions for a positive 1-sd shock to the leisure preference φ_t in period 1. Y-axes show log-deviations from the non-stochastic steady state. A description of the simulation procedure is given in footnote 10.

Figure 4: Size and Sign Dependence of Impulse Response Functions



Notes: Simulated impulse response functions for small (1-sd) and large (3-sd) shocks. Expansionary shocks (green dashed line) are mulitplied with −1 to compare quantities.

Figure 5: State Dependence of Impulse Response Functions



Notes: Simulated impulse response functions for an expansionary 1-sd shock in period 1. Solid blue lines: Economy was in a recessionary state before impact. Green dashed lines: Economy was in an expansionary state before impact. States before impact were computed as the average state of the economy in the 1/6th of periods with highest and lowest output, respectively.

Table 3: Asymmetry of Output Levels and Other Variables

	Model		Data	
Variable	Mean pos vs neg	Skewness	Mean pos vs neg	Skewness
Levels				
Output Y	3.27% vs −3.46%	−0.10	2.73% vs −3.43%	−0.46
Labor L	3.38% vs −3.57%	−0.11	3.11% vs −3.99%	−0.40
Investment I	18.84% vs −22.94%	−0.42	13.08% vs −17.12%	−0.53
Consumption C	1.53% vs −1.54%	−0.02	2.23% vs −2.24%	−0.00
Growth rates				
Output ΔY	3.14% vs −3.14%	−0.00	2.30% vs −2.25%	0.19
Labor ΔL	3.45% vs −3.44%	0.00	1.63% vs −2.03%	−0.55
Investment ΔI	23.22% vs−23.00%	0.01	10.00% vs −11.30%	−0.21
Consumption ΔC	0.37% vs −0.38%	−0.07	1.94% vs −1.85%	0.38

Notes: Measures of asymmetry as defined in Table 1 (baseline specification). "Levels" measured in log-deviations from simulation mean (model) or from HP-1600 trend (data), respectively. "Growth rates" measured as log-differences. Data for Output, Investment and Consumption from BEA NIPA tables (Real gross domestic product, personal consumption expenditures, gross private domestic investment", respectively). Data for labor from BLS statistics as hours of all persons in the nonfarm business sector. All data are quarterly.

5.2.2 Output asymmetry

We now turn to an assessment of the implications for the stylized facts in general equilibrium. For the difference between large positive and negative deviations in output, following the approach from the empirical section, we choose an integer N of around 1/6 of the observations ($N = 1666$ out of $10,000$ simulated periods) and compare the mean of the $N/2$ periods with highest output to the $N/2$ periods where output is lowest. As shown in Table 3, the average large recession so defined is −3.46% below trend, whereas the average large expansion is 3.27% above trend. The simulated output series as a whole is negatively skewed with a coefficient of −0.10.

Comparing this to the empirical equivalents in Table 1, the differences between positive and negative output deviations in the model cover around a quarter of those in the data. In the model, recessions are 0.20 percentage points (or 6%) deeper than expansions. As the model was calibrated to match the standard deviation of HP(1600)-filtered output, the closest comparable measure is the first row of Table 1 showing a relative difference of 23%, or 0.70 percentage points.

Regarding the other aggregate time series also listed in Table 3, the model generates asymmetry in levels of investment as well as levels of hours worked, but not for the level of consumption nor the growth rate of output — all these patterns are consistent with empirical findings discussed in section 2 and replicated in the Table. The simulation does not exhibit asymmetry in growth rates of employment, even though there is some empirical evidence for this (e.g McKay and Reis (2008)). The reason behind this is that in the model with its frictionless labor markets the employment and output series move together very closely.

5.2.3 Cross-sectional and aggregate volatility

To assess the relation between profitability dispersion and output, we consider the cross-sectional standard deviation of $\log(\text{profitability}_i)$. Profitability is measured as firm i's priced Solow residual $p_i SR_i = p_i y_i/(k_i^\alpha l_i^{1-\alpha})$ which has the interpretation of "revenue in dollars per input factor basket". As discussed above, this measure uses rented capacity as a measure of capital input — of course firms' true physical productivity $y_i/(\tilde{k}_i^\alpha l^{1-\alpha})$ is constant by construction. Since a firm's profitability is only a function of its price and demand shock, and all firms choose the same price, profitability dispersion in any given period only depends on the variance of realized demand between firms up to capacity $\min\left\{b_i, \bar{b}\right\}$ with

$$\text{Var}\left(\log(p_i SR_i)\right) = \left(\frac{\alpha}{2-\alpha}\right)^2 \text{Var}\left(\log\left(\min\left\{b_i, \bar{b}\right\}\right)\right)$$

(see appendix D).

We then consider the fluctuation of this measure over the business cycle in the simulations to assess the question, how much wider does the firm distribution of profitability become in recessions? Kehrig (2015) finds that for the six recessions in his data ranging from 1972 to 2009, profitability was 2.84% more dispersed in recessions than in the long-run average. If similarly we look at the 1/6 of simulation periods in which output is lowest we find that profitability dispersion increases by 0.48% in recessions, implying that the model captures 17% of the increase in cross-sectional volatility in recessions.

Profitability dispersion in the model is only a function of the cutoff level $\bar{b}$, such that its correlation with output will mirror the correlation of $\bar{b}$ with output. In the simulations the correlation $\text{corr}(\text{sd}(\log(SR_i))_t, Y_t) = -0.91$ is correspondingly strong, and higher than the -0.4 to -0.5 that have been measured in Kehrig (2015) and Bloom et al. (2012). This high correlation in the model results from the close comovement between aggregate output and the level of constrained firms we saw in the impulse response functions.

Turning to aggregate volatility, we construct a measure of aggregate volatility from the simulated output series. For this, we look at the variance in the growth rates of output in recessions and expansions, respectively. Specifically, we compute $\text{sd}(\log(Y_{t+1}/Y_t))$ conditional on Y_t being in its lowest or highest quintile. We expect the variance of output growth to be large in recessions: When a large number of firms are far away from their capacity constraint, output effects of a shock of a given size are stronger. In the model here, the fact that firms can adjust their capacity levels and prices quickly in response to an aggregate shock dampens this effect: It allows firms to lower their capacity after the realization of a bad shock, which in turn increases the number of firms at their constraint. If it took firms longer to react, say with a 'time to build' of two periods instead of one, we would expect to see significantly stronger movements in aggregate conditional volatilities.

Table 4 lists the volatility of several model time-series in the first two columns. Going from boom to recession, the standard deviation of output growth increases from 1.50% to 1.66%. The model's investment and labor series exhibit countercyclical conditional volatilities as well, whereas consumption volatility stays constant over the cycle. In the model, aggregate risk as measured by the volatility of output increases by 10.8%. We also construct the empirical analogues of the volatility measures using US data, which are shown in columns 3 and 4 of Table 4. As in the baseline empirical specification of Table 1, we consider as recessions the 20 quarters since 1949 in which detrended output was lowest. In the data, output volatility

Table 4: Aggregate risk: Conditional Volatilities of Aggregate Variables

	Model		Data	
Variable	Expansion	Recession	Expansion	Recession
Output Y	1.50%	1.66%	0.95%	1.41%
Labor L	1.60%	1.77%	0.68%	1.08%
Investment I	9.64%	11.88%	5.21%	5.91%
Consumption C	0.20%	0.21%	1.08%	0.68%

Notes: Standard deviation of growth rate in expansion/recession in the model. For time series X, conditional volatility in recession is computed as the standard deviation of growth rates following a recessionary quarter; i.e. we compute sd$(\log X_{t+1} - \log X_t | X_t$ in recession). Analogous for expansions. Recessions and expansions as defined in Table 1 (baseline specification) and section 5.2.2; in particular output is among the lowest/highest 20 periods (data) and lowest/highest 833 periods (simulated model series).

in a recession is 39.5% higher in recessions than in booms and thus fluctuates stronger than in the model. Additionally, in the US series both investment and consumption exhibit cyclical volatilities, whereas in our model households are generally able to smooth consumption very well as they do not face any frictions.

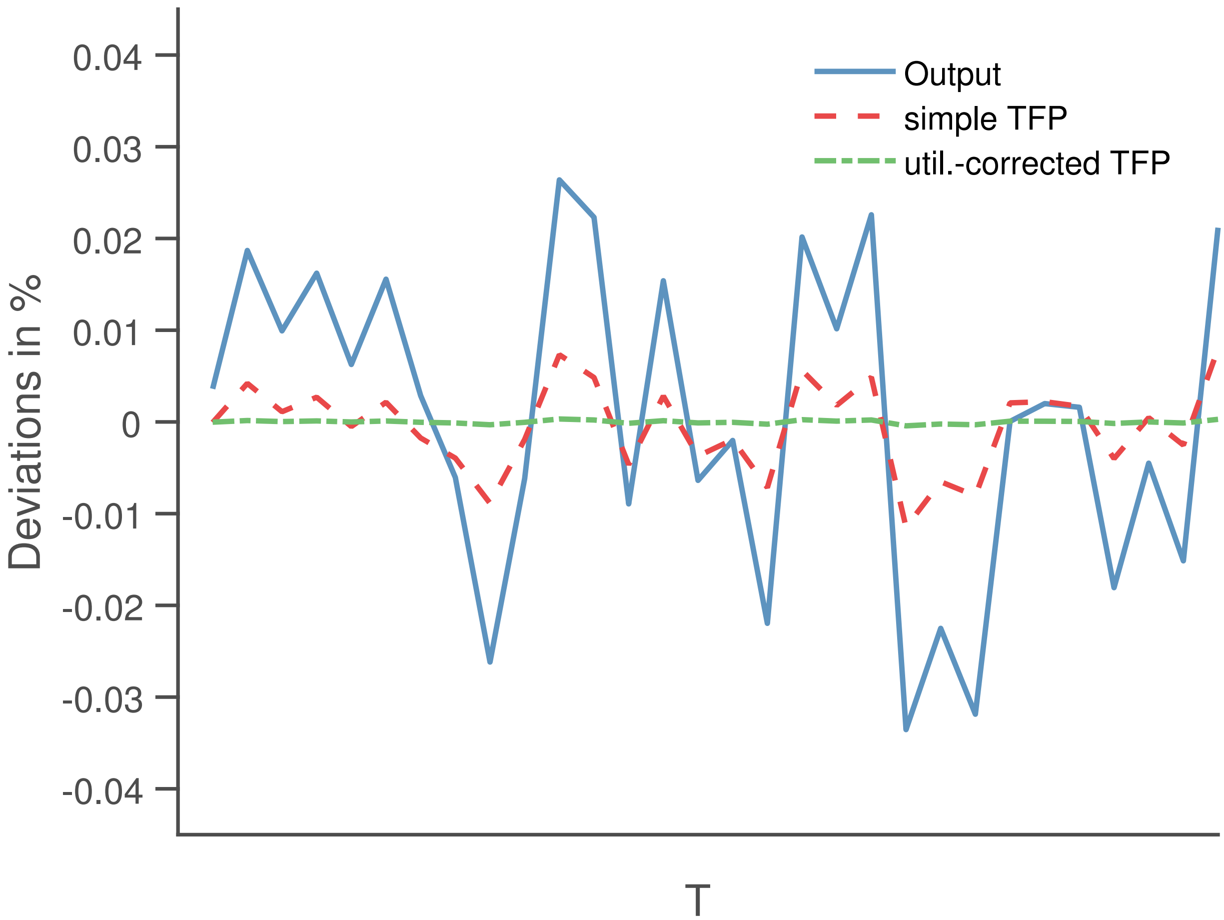
Using different empirical strategies, Bloom et al. (2012) find that recessions are associated with a 23% higher standard deviation of output (compared to normal times), and Bachmann and Bayer (2013) obtain a difference of around 35% between booms and recessions, in line with the empirical values found here. The model hence covers between a quarter to a third of observed fluctuations in aggregate output volatility.

5.2.4 Aggregate Solow residual

We construct the aggregate Solow residual in a similar way as its firm-level equivalent. We compute the uncorrected Solow residual as $SR_{\text{simple},t} = Y_t / \left(K_t^\alpha L_t^{1-\alpha}\right)$ using aggregate capital in the denominator, and the corresponding version corrected for utilization as $SR_{\text{corr},t} = Y_t / \left(\tilde{K}_t^\alpha L_t^{1-\alpha}\right)$ where $\tilde{K}_t = \int_i \tilde{k}_{it} \mathrm{d}i$ is defined as the aggregate utilized capital. Figure 6 displays the log deviations from the mean for output as well as both Solow residual for a subset of the simulated periods. As in Basu et al. (2006) and Fernald (2014), the correlation between the simple TFP measure with output is strong with a value of 0.73. On the other hand, in the model utilization-corrected productivity barely moves over the cycle.[11] The standard deviation of growth in the simulated series's Solow residual is 0.46%. The corresponding empirical measure in John Fernald's quarterly dataset is the uncorrected Solow residual which fluctuates with a standard deviation of 0.87%.

[11]The reason that utilization-corrected TFP fluctuates at all over time in the model is because of changes in the composition of input factors and their allocation between firms. Since this corrected TFP has a very small variance in the simulation (it has a standard deviation of 0.00018) even a tiny amount of noise —like measurement error— would render it acyclical.

Figure 6: Output and TFP measures



Notes: Output (solid blue line) is Y_t, simple TFP (red dashed line) is measured as $Y_t/(K_t^{\alpha} L_t^{1-\alpha})$, corrected TFP (green dash-dotted line) is measured as $Y_t/(\tilde{K}_t^{\alpha} L_t^{1-\alpha})$. Y-axis displays log-differences from non-stochastic steady state. X-axis displays a window of 100 periods out of the 10,000 simulation periods.

5.2.5 Fiscal multiplier

Finally, we consider the cyclicality of the contemporaneous fiscal multiplier $\mathrm{d}Y_t/\mathrm{d}G_t$. In constructing it we follow Sims and Wolff (2017) by averaging the state variables over those periods in which production is in its lowest quintile. We compare output in this "average bad state" to output in the same state, but with an additional small positive shock to government spending. More formally, if S is the aggregate state, and $S + \Delta G$ the aggregate state after small fiscal spending shock ΔG, the government multiplier is computed as $\left(Y^{S+\Delta G} - Y^S\right)/\Delta G$. The value of the multiplier when output is in its top quintile is computed the same way. We obtain values of 0.99 for the multiplier in a recession, and 0.87 for a multiplier in a boom. The empirical literature has not settled on a consensus for the size of fluctuations in the fiscal multiplier over the business cycle, and estimates have relatively wide confidence intervals. Prominent examples include Auerbach and Gorodnichenko (2012a) with estimates around 0 and around 1.5 for historic booms and recessions, respectively, and Ramey and Zubairy (2018) who find only small fluctuations over the cycle.

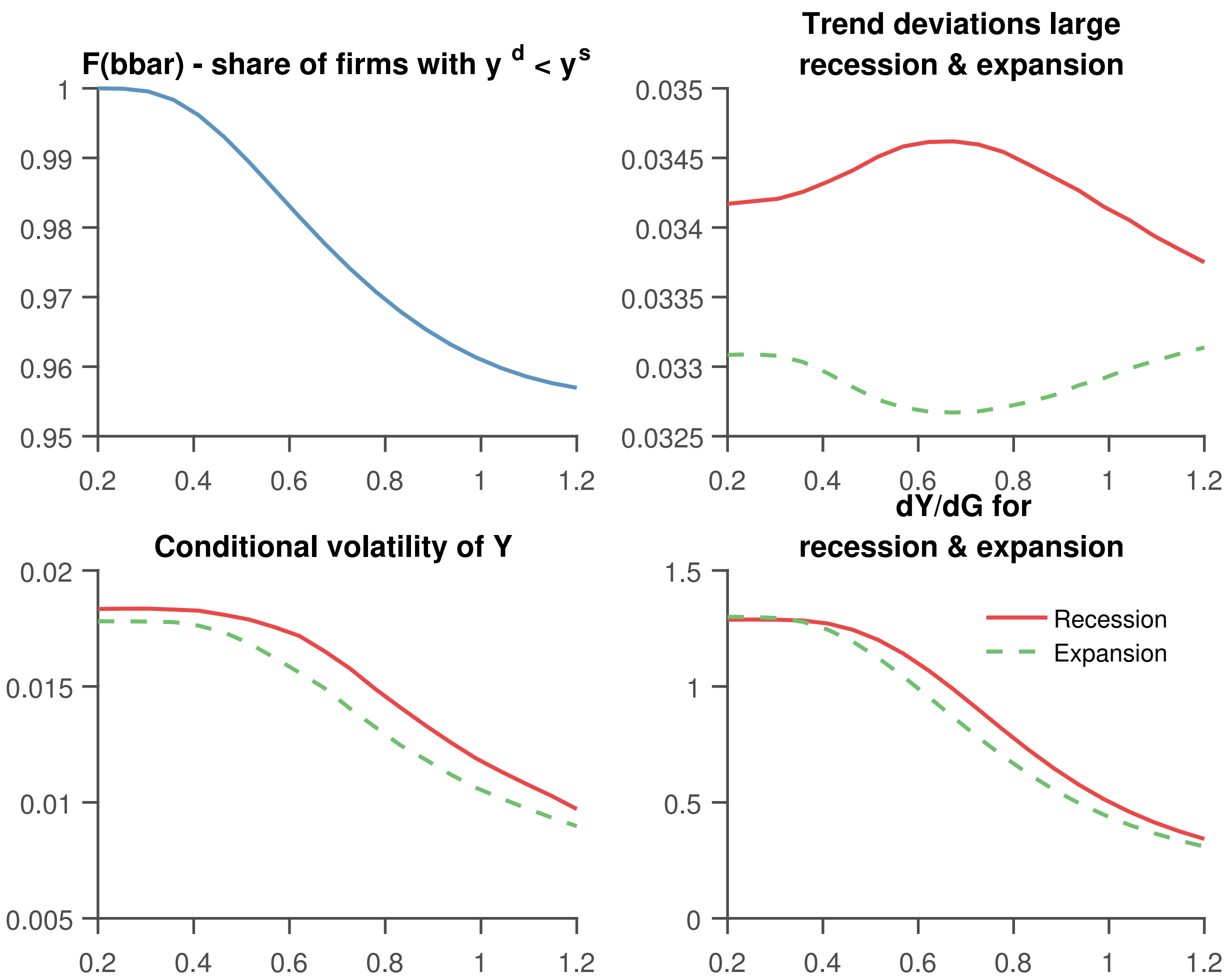
5.2.6 Role of heterogeneity, nominal rigidities, and discussion

Variance of idiosyncratic shocks The variance of idiosyncratic demand shocks, parameterized by σ_b, directly influences how many firms are capacity constrained. It is instructive to consider how model results depend on this parameter. Figure 7 shows this for several outcomes. The graph in upper left displays the share of constrained firms in steady state. Unsurprisingly, the wider the distribution of idiosyncratic demand shocks, the more firms face a level of demand exceeding their capacity. The next two graphs show output deviations from steady state for booms and recessions (top right), and the relative size of these deviations to each other (bottom left), respectively. Notably, output asymmetry is non-monotonic in σ_b. To see why this is, consider that the main driver of the effects in the model is the *change* in the share of constrained firms over the cycle and not its absolute level. Those differences in $F\left(\bar{b}_t\right)$ between expansion and recessions are largest for an interior value of σ_b. At a low value of 0.2 there are practically no constrained firms in equilibrium, and recessions are around 3.5%, or 0.13 percentage points, larger than expansions. (Even when there is no heterogeneity between firms there is some concavity in production through the convex capacity utilization cost and labor supply.) Increasing the standard deviation σ_b to around 0.75 makes recessions more than 6% larger than expansions. For high values of σ_b, output asymmetry is reduced again because, despite a larger share of constrained firms in steady-state, the *change* in this share over the cycle is smaller.

A similar pattern can be observed for the other effects of capacity constraints. Consider for example the fiscal multiplier in the bottom right graph of Figure 7. When virtually no firms are capacity constrained, the timing of government spending does not matter for its effect on output — all firms can increase their production in response to government demand. The cyclicality of the multiplier is strongest when the fluctuations in $F\left(\bar{b}_t\right)$ over the cycle are large. In this case comparatively many firms have idle capacities in a recession and can respond to an increase in government demand.

Summarizing, the firm heterogeneity causing capacity constraints to bind matters in this model because it *generates* cyclicality in the fiscal multiplier and cross-sectional profitability dispersion, and *amplifies* the deepness of recessions and differences in aggregate volatility.

Figure 7: Varying Idiosyncratic Shock Variance σ_b



Notes: X-axes display value for σ_b. On Y-axes: Top left – Fraction of unconstrained firms $F(\bar{b})$ in the non-stochastic steady state. Top right – Absolute log deviations of recessions and expansions from non-stochastic steady state. Bottom left – Log difference between absolute deviations in recession and expansion (i.e. log difference of the curves in top right). Bottom right – Government multiplier in recession and expansion.

Role of nominal rigidities and demand shocks In our model, one-period-ahead pricing by intermediate good firms both creates the upper bound to production y^s that firms are willing to supply, and constitutes the nominal rigidity through which demand shocks have effects on the economy. How important are sticky prices for the effects of capacity constraints? To analyze qualitative differences to the baseline model we set up a flexible-price version with an ad-hoc capacity constraint. In this alternative model intermediate firms' prices adjust freely and technology shocks induce aggregate fluctuations. As we outline below and discuss in more detail in appendix F, in the flexible price case capacity constraints still generate output asymmetry and countercyclical aggregate volatility, however in contrast to the sticky price version (and to the data), cross-sectional profitability becomes *less* dispersed in recessions (i.e. it is procyclical).

In this extension, we allow intermediate goods firms to sell their varieties on the spot at a flexible price instead of being bound to a pre-set price. This means there is no longer a fixed amount y^s that firms are willing to produce at most, because now higher demand will translate into a higher price for the good which in turn can cover the increasing marginal costs of production. We therefore impose an exogenous capacity constraint, requiring $\tilde{k} \leq k$ or in other words that utilized capital $\tilde{k}$ does not exceed a certain limit proportional to capacity k. Finally we add a common productivity component to the intermediates' production function which is subject to technology shocks. Under the same intuition as before, large deviations from trend output are still larger in recessions than in expansions; and again the variance of output growth is greater in a downturn than in a boom. However, there is a qualitative difference in the behavior of cross-sectional profitability dispersion. Recall that profitability is given by $p_i y_i/(k_i^{\alpha} l_i^{1-\alpha})$ and hence with identical pre-set prices, all capacity constrained firms have the same profitability, compressing the right tail of the distribution in a boom. When prices are flexible they depend on demand b_i even when the firm is capacity constrained, and in the appendix we show that this drives up the affected firms' profitability particularly strongly. As a result, in a boom when many firms become constrained, the right tail of the distribution fans out, leading to *pro*cyclical profitability dispersion. Finally, since the economy is driven by aggregate productivity shocks both measured Solow Residual and true physical productivity are procyclical, and without price rigidities there is no output multiplier from government spending through demand effects.

Effect size Is it possible for the same mechanism to deliver stronger effects? One can think of several factors potentially affecting the results.

First, the model is only solved locally, i.e. any effects of aggregate fluctuations are captured by evaluation of the first and second derivative of the equilibrium conditions at the steady state. Any higher-order concavity in the relation between shock size and output is lost when moving away from the steady-state and could only be recovered through a global solution method.

Second, the model has little internal propagation due to the one-period-ahead choices of prices and capacity. Firms are thus very quick to adjust to aggregate shocks, such that it is hard for individual shocks to "add up" over time. In fact it is predominantly the *innovation* to the aggregate state variable φ_t that matters for chance of binding capacity constraints. Since the model is solved up to a second-order approximation, effect sizes increase linearly in the size of the aggregate shock. As an illustration, if one detrends quarterly GDP since 1949 with a linear filter (instead of the HP(1600) filter used in calibration) this implies a considerably higher standard deviation of the detrended series of 4.7% instead of 1.8%. When this standard deviation is used in the model, the corresponding output deviations in recession and expansion strengthen to -9.05% and 7.80%, respectively. The standard deviations of output growth become 4.4% and 3.6%, cross-sectional profitability becomes 1.12% more dispersed in recessions, and the fiscal multipliers widen to 1.09 and 0.79. Similar increases in effect sizes can be expected from increasing the "time to build" (and price-set) from one period to a longer horizon.

6 Conclusion

This paper includes capacity constraints in a DSGE framework under demand shocks and shows that the model replicates diverse stylized facts of US output: Recessions are deep; they are times of high volatility both in the aggregate and the cross-section; and they are times when fiscal policy is particularly effective. Since firms choose their utilization after capacity has been installed, the mechanism also reproduces an endogenously procylical Solow residual.

A calibrated New Keynesian model yields differences in output between booms and recessions of around 0.20 percentage points, such that the model explains more than a quarter of the 0.7 percentage-point difference we find empirically. The model results cover a similar share of a quarter to a third of the observed increase in aggregate volatility. While the empirical literature has not settled on the size of fluctuations in the government spending multiplier over the cycle, in the calibration we find a multiplier of on average 0.87 in booms and 0.99 in recessions. All effect sizes increase with the severity of recessions.

The model contains a minimal set of ingredients for the mechanism of capacity constraints to qualitatively deliver the stylized facts. An interesting expansion of the model would be to gain more realism in the description of firm behavior. Allowing for richer adjustment frictions to price-setting and investment could allow us to examine whether the mechanism is consistent with additional cross-sectional and intertemporal moments.

References

Adjemian, S., H. Bastani, and M. Juillard (2011). Dynare: Reference manual, version 4.

Álvarez-Lois, P. P. (2006, nov). Endogenous capacity utilization and macroeconomic persistence. *Journal of Monetary Economics 53*(8), 2213–2237.

Auerbach, A. A. J. and Y. Gorodnichenko (2012a, may). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy 4*(2), 1–27.

Auerbach, A. A. J. and Y. Y. Gorodnichenko (2012b). Fiscal multipliers in recession and expansion. *Fiscal Policy after the Financial Crisis* (August).

Auerbach, A. J. and Y. Gorodnichenko (2012c, may). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy 4*(2), 1–27.

Bachmann, R. and C. Bayer (2013, jun). 'Wait-and-See'business cycles? *Journal of Monetary Economics 60*(6), 704–719.

Bachmann, R. and E. R. Sims (2012, apr). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics 59*(3), 235–249.

Bachmann, R. and P. Zorn (2016). The Nature of Capital Expenditures over the Business Cycle.

Bai, J. and S. Ng (2005, jan). Tests for Skewness, Kurtosis, and Normality for Time Series Data. *Journal of Business & Economic Statistics 23*(1), 49–60.

Bai, Y., J.-V. Ríos-Rull, K. Storesletten, J. Ros-Rull, and K. Storesletten (2012). Demand shocks as productivity shocks. *Federal Reserve Board of Minneapolis*.

Basu, S., J. Fernald, and M. Kimball (2006). Are technology improvements contractionary? *American Economic Review 96*(5), 1418–1448.

Bils, M. and J. O. Cho (1994). Cyclical factor utilization. *Journal of Monetary Economics 33*(2), 319–354.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Bloom, N. (2014). Fluctuations in Uncertainty. *Journal of Economic Perspectives 28*(2), 153–176.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2012). Really uncertain business cycles. *NBER* (w18245).

Christiano, L., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*(1), 1–45.

Clower, R. (1967). A Reconsideration of the Microfoundations of Monetary Theory. *Western economic journal 6*(1).

Cooley, T., G. Hansen, and E. Prescott (1995). Equilibrium business cycles with idle resources and variable capacity utilization. *Economic Theory 49*, 35–49.

Corsetti, G., A. Meier, and G. Müller (2012). What determines government spending multipliers? *Economic Policy 27*(72), 521–565.

Decker, R., P. D'Erasmo, and H. Boedo (2015). Market Exposure and Endogenous Firm Volatility over the Business Cycle. *American Economic Journal: Macroeconomics*.

DeLong, J. and L. Summers (1986). Are business cycles symmetric?

Eisfeldt, A. L. and A. a. Rampini (2006, apr). Capital reallocation and liquidity. *Journal of Monetary Economics 53*(3), 369–399.

Fagnart, J.-F., O. Licandro, and F. Portier (1999). Firm Heterogeneity, Capacity Utilization, and the Business Cycle. *Review of Economic Dynamics 2*(2), 433–455.

Fernald, J. (2014). A quarterly, utilization-adjusted series on total factor productivity. *Manuscript, Federal Reserve Bank of San Francisco*.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review 98*(1), 394–425.

Gilchrist, S. and J. C. Williams (2000). Putty-Clay and investment - A Business Cycle Analysis. *Journal of Political Economy 108*(5), 928–60.

Greenwood, J., Z. Hercowitz, and G. Huffman (1988). Investment, capacity utilization, and the real business cycle. *The American Economic Review* (June), 402–417.

Hansen, G. G. D. and E. E. C. Prescott (2005, oct). Capacity constraints, asymmetries, and the business cycle. *Review of Economic Dynamics 8*(4), 850–865.

Ilut, C., M. Kehrig, and M. Schneider (2016). Slow to Hire, Quick to Fire: Employment Dynamics with Asymmetric Responses to News.

Ilzetzki, E., E. G. Mendoza, and C. a. Végh (2013, mar). How big (small?) are fiscal multipliers? *Journal of Monetary Economics 60*(2), 239–254.

Ireland, P. (2001). Sticky-price models of the business cycle: specification and stability. *Journal of Monetary Economics 47*, 3–18.

Kehrig, M. (2015). The Cyclical Nature of the Productivity Distribution. *Working Paper* (January), 1–69.

Knüppel, M. (2014). Can Capacity Constraints Explain Asymmetries of the Business Cycle? *Macroeconomic Dynamics 18*(01), 65–92.

Kuhn, F. (2015). Endogenous profitability dispersion. *Working paper*.

McKay, A. and R. Reis (2008, may). The brevity and violence of contractions and expansions. *Journal of Monetary Economics 55*(4), 738–751.

Michaillat, P. (2014, jan). A Theory of Countercyclical Government Multiplier. *American Economic Journal: Macroeconomics 6*(1), 190–217.

Mittnik, S. and W. Semmler (2012). Regime dependence of the fiscal multiplier. *Journal of Economic Behavior & Organization*, 1–37.

Nakamura, E. and J. Steinsson (2014). Fiscal stimulus in a monetary union: Evidence from US regions. *American Economic Review 104*(3), 753–792.

Ramey, V. and S. Zubairy (2018). Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data. *Journal of Political Economy*.

Salgado, S., F. Guvenen, and N. Bloom (2015). Skewed Business Cycles.

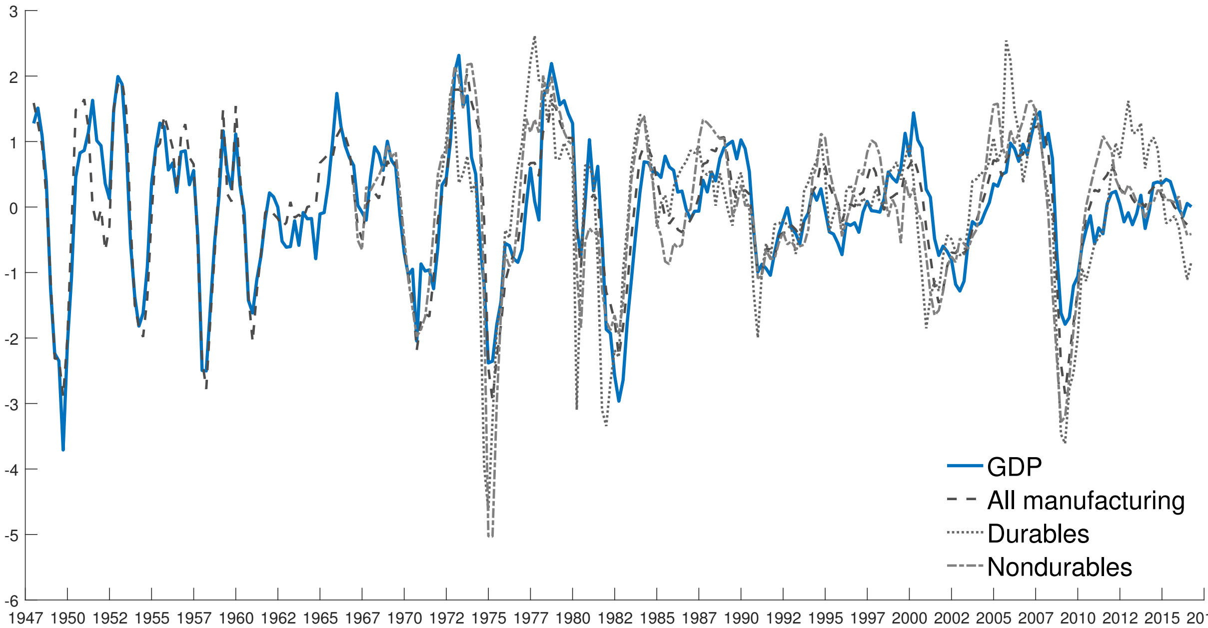
Sichel, D. E. (1993). Business Cycle Asymmetries: A Deeper Look. *Economic Inquiry 31*(2), 224–236.

Sims, E. and J. Wolff (2017). The Output and Welfare Effects of Government Spending Shocks over the Business Cycle.

Syverson, C. (2011, jun). What determines productivity? *Journal of Economic Literature 49*(2), 326–365.

Wen, Y. (2004). What Does It Take to Explain Procyclical Productivity? *Contributions in Macroeconomics 4*(1).

Figure 8: Capacity utilization in manufacturing, durables and non-durables are procyclical



Notes: Blue line is log quarterly GDP. Grey lines are the Federal Reserve's capacity utilization indices for all manufacturing (dashed), manufacturing of durables (dotted), and manufacturing of non-durables (dash-dotted). All series detrended with an HP(1600) filter and normalized with their respective standard deviation over time.

A Indirect evidence for procyclical capacity utilization by firms

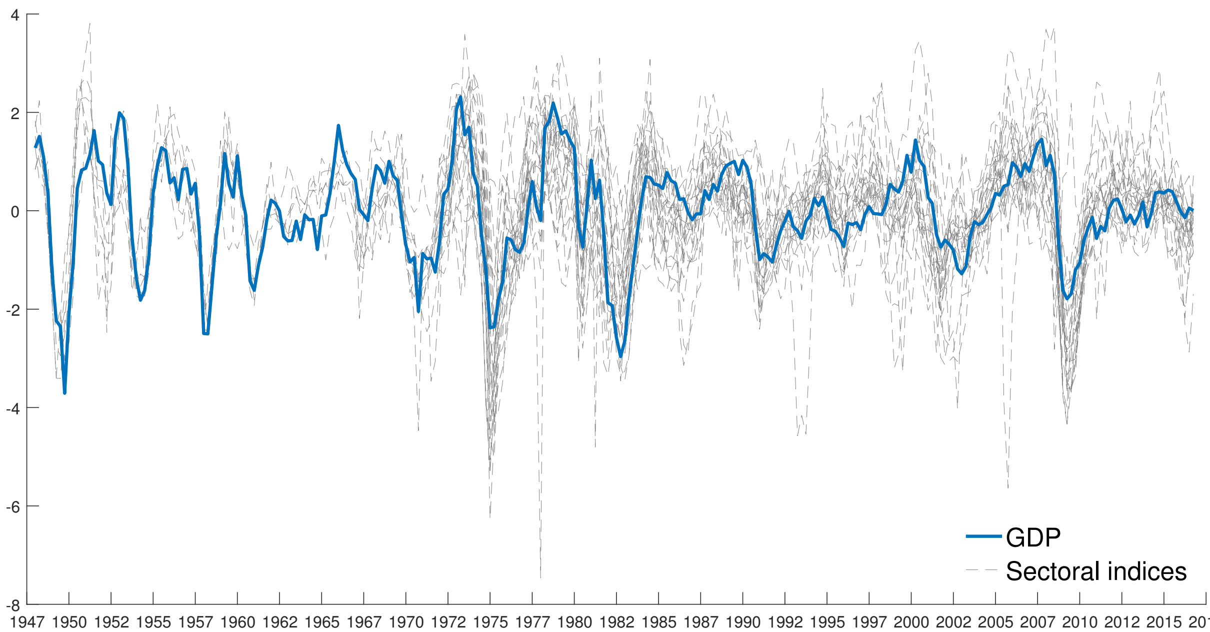
In this section we discuss the data sources and derived statistics referred to at the end of section 2 in more detail.

Capacity utilization indices by sector We consider log quarterly GDP as output measure, and capacity utilization time series provided by the Federal Reserve in their dataset "Industrial Production and Capacity Utilization". To assess correlation between two indices we HP-filter the series with the standard smoothing parameter of 1600. We then take first differences on the cyclical component of the series to remove first-order autocorrelation (cyclicality is even stronger when considering levels instead of differences).

The aggregate manufacturing capacity utilization index and the sub-indices for durable goods and non-durable goods are strongly procyclical. Out of the 24 indices for three-digit NAICS manufacturing sectors, 22 are significantly procyclical (the two sectors not exhibiting significant procylicality are food and beverages). Of these three-digit sectoral indices, the median correlation coefficient is 0.48 (computers and electronic products). Figures 8 and 9 plot aggregate and sectoral utilization measures against GDP.

Using sectoral utilization and production data we test two additional predictions of our mechanism. According to the mechanism, shifts in the mean of the cross-sectional distribution

Figure 9: Capacity utilization in three-digit NAICS sectors are procyclical



Notes: Blue line is log quarterly GDP. Grey lines are the Federal Reserve's capacity utilization indices for all 24 three-digit NAICS industries for which data is provided. All series detrended with an HP(1600) filter and normalized with their respective standard deviation over time.

of firm level capacity utilization drive the share of constrained firms which in turn causes the observed business cycle effects. A prediction is that, *ceteris paribus*, the larger the variance of aggregate fluctuations, the stronger we expect the share of constrained firms to fluctuate and hence the stronger business cycle effects we expect to see.

To check whether this is the case in the data, we use the Bureau of Economic Analysis' data on sectoral value added. The data is annual, spanning 1947 to 2015. We compare the asymmetry in sectoral value added with the FRB's information on capacity utilization for all manufacturing sectors (without mining and drilling) for which both datasets provide data within the same NAICS code. As measures of asymmetry in sectoral production we use largest absolute deviations from trend as described in section 2 as well as the coefficient of skewness. In table 5 we compare these two measures by sector. We find fairly robust correlation between the size of fluctuations in economic activity of a sector and the extent to which it exhibits asymmetry, with correlation between the measures of 0.51 and 0.73.

Backlog of orders index The Institute for Suply Management's backlog of orders index[12] tracks managers' survey responses about changes in the backlog of unfilled orders in the three qualitative categories "greater", "same" or "less". The ISM collects these data as part of their monthly "Report on Business" in which more than 300 (manufacturing) and more than 375 (non-

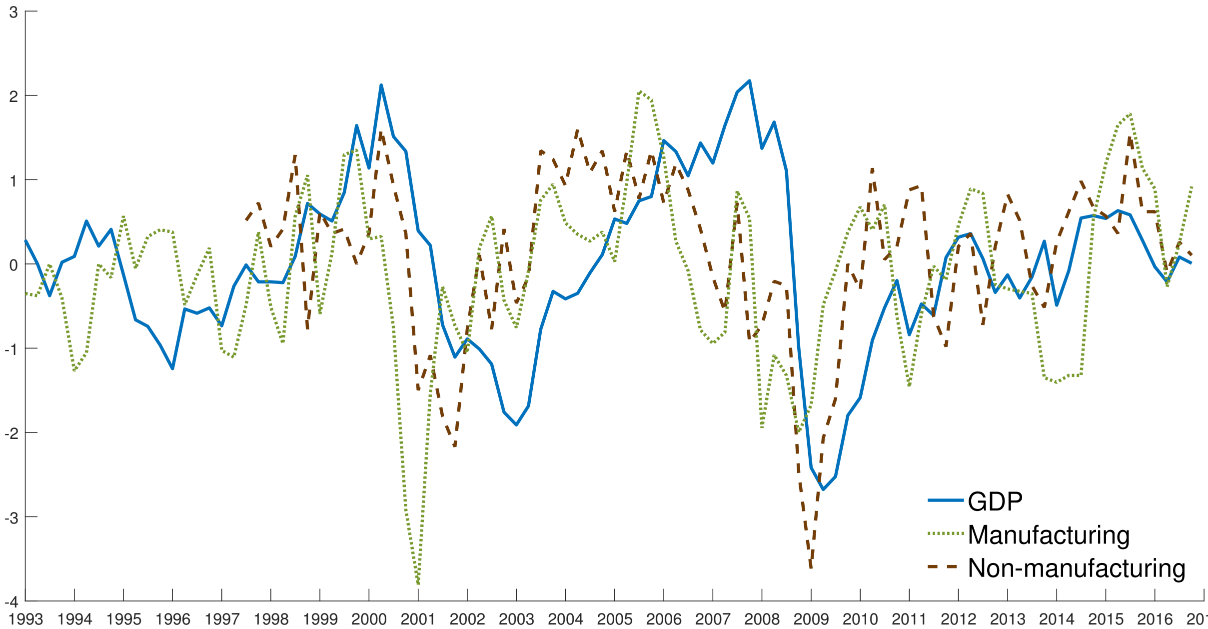
[12]This index was available via the St. Louis Fed's FRED website until 2015, it is currently available at quandl.com (https://www.quandl.com/data/ISM/MAN_BACKLOG-Manufacturing-Backlog-of-Orders-Index)

Table 5: Sectors with larger fluctuations in capacity utilization exhibit stronger asymmetry

Sector (NAICS code)	Standard dev. utilization (in pp)	Asymmetry in sectoral value added: Trend deviation (in pp)	Skewness
Motor vehicles (3361)	8.1	−6.26 (-21.5; 15.2)	−2.32
Primary metal (331)	5.8	−0.30 (-9.4; 9.1)	0.00
Plastics & rubber (326)	4.7	−1.16 (-7.7; 6.5)	−0.01
Machinery (333)	4.6	−1.68 (-10.1; 8.4)	−0.31
Electrical equipment (335)	4.5	0.28 (-7.9; 8.2)	−0.01
Wood product (321)	4.4	−1.50 (-7.6; -6.1)	−0.69
Furniture (337)	4.2	−2.09 (-9.1; 7.0)	−0.62
Textile (313)	4.2	−2.06 (-8.4; 6.3)	−0.98
Fabricated metal product (332)	4.1	−2.42 (-9.0; 6.6)	−0.83
Nonmet. mineral product (327)	3.7	−2.33 (-8.6; 6.3)	−1.14
Computer & electronics (334)	3.5	0.45 (-6.1; 6.6)	1.10
Apparel (315)	3.1	−0.89 (-4.8; 3.9)	−0.71
Paper (322)	3.0	−0.43 (-7.1; 6.7)	−0.50
Chemical (325)	2.8	0.24 (-6.6; 6.9)	−0.20
Petroleum & coal product (324)	2.5	0.92 (-22.9; 23.8)	1.02
Miscellaneous (339)	2.2	−1.34 (-6.3; 5.0)	−0.66
Printing (323)	2.2	−0.07 (-3.8; 3.7)	−0.04
Food & beverage (311)	1.2	0.73 (-5.2; 5.9)	0.43
Correlation coefficient	1	−0.73	−0.56
Rank-order correlation	1	−0.51	−0.24

Notes: Variance in capacity utilization and asymmetry measures by sector. First column contains all the sectors for which both utilization data and value-added data were available. Second column: The time-series standard deviation of the FRB's capacity utilization indices, after removing an HP(1600) trend, in percentage points. Third column: Mean of 5 quarters with largest negative deviations from trend in sectoral value added plus mean of 5 quarters with largest positive deviations from trend (see also description of table 1). Fourth column contains coefficient of skewness for sectoral value added. Correlation coefficient: the correlation of the standard deviation in utilization with asymmetry in value added across sectors, in other words the correlation coefficient of column 2 with columns 3 and 4. Rank order correlation: Spearman's rho for rank-order correlation of column 2 with columns 3 and 4.

Figure 10: Changes in the backlog of orders for manufacturing and non-manufacturing are procyclical



Notes: Blue line is log quarterly GDP detrended with HP(1600) filter. Non-solid lines are backlog of orders indices for manufacturing (green dotted) and non-manufacturing (red dashed). Indices measure difference in the share of respondents reporting increases in backlog of unfilled orders minus share of respondents reporting decrease. Both series for changes in backlogs lead the level of GDP by 3 quarters. All series are normalized by their standard deviation.

manufacturing) firms are surveyed repeatedly, with response rates for backlogs averaging around 87% and 60% respectively. The data range from January 1993 (manufacturing) and July 1997 (non-manufacturing) to December 2016. The index number is reported as the difference between share of respondents indicating an increase in their backlog minus the share of respondents reporting a decrease.

We aggregate index numbers by quarter and correlate the level of the index with first differences in log quarterly GDP detrended with an HP-1600 filter (results are very similar if we keep the monthly frequency and use monthly industrial production as output measure instead). We find significant procyclicality for both manufacturing and non-manufacturing indices with correlations of 0.35 and 0.55, respectively.

We also consider the correlation of the level of the index with the *level* of GDP in order to compare changes in unfilled orders with the current output level. We find that changes in the backlog lead GDP by 3 quarters, and that the correlation of the indices with GDP levels at a lag of 3 quarters is 0.35 and 0.51 for manufacturing and non-manufacturing, respectively. Figure 10 plots the indices together with GDP (at 0 lags).

B Aggregator's demand function

The aggregator's problem is to

$$\max_{\{y_i\}_{i=0}^1} \left[\int b_i^{\frac{1}{\sigma}} y_i^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}} + \lambda \left[I - \int p_i y_i di\right] + \int \mu_i \left[\bar{y} - y_i\right] di$$

such that the first-order necessary conditions with respect to y_i are given by

$$\left[\int b_i^{\frac{1}{\sigma}} y_i^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{1}{\sigma-1}} \left(\frac{b_i}{y_i}\right)^{\frac{1}{\sigma}} = \lambda p_i + \mu_i y_i \qquad \forall i.$$

For any given variety i either we have to consider two cases. If the aggregator is unconstrained in this variety, i.e. $\mu_i = 0$, then

$$\lambda = \left[\int b_i^{\frac{1}{\sigma}} y_i^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{1}{\sigma-1}} (b_j/y_j)^{\frac{1}{\sigma}},$$

whereas the aggregator is limited to purchasing $\bar{y}$ of variety i if

$$\mu_i = \left[\int b_i^{\frac{1}{\sigma}} y_i^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{1}{\sigma-1}} \left(\frac{b_i}{y_i}\right)^{\frac{1}{\sigma}} - \lambda p_i > 0$$

.

For any two varieties i, j with $\mu_i = \mu_j = 0$ then the relationship

$$\frac{y_i}{y_j} = \frac{b_i}{b_j}\left(\frac{p_j}{p_i}\right)^{\sigma}$$

holds. Integrating over all i one then has

$$\begin{aligned} I = \int_0^1 p_i y_i di &= \left(\int_{i\in U} p_i^{1-\sigma} b_i di\right) \frac{y_j}{b_j} p_j^{\sigma} + \int_{i\in C} p_i \bar{y} di \\ &= P_U^{1-\sigma} y_j \frac{p_j^{\sigma}}{b_j} + \int_{i\in C} p_i \bar{y}, \end{aligned}$$

where $U \equiv \{i : \mu_i = 0\}$ and $C \equiv \{i : \mu_i > 0\}$ are index sets over unconstrained and constrained varieties, respectively, and $P_U \equiv \left(\int_{i\in U} p_i^{1-\sigma} b_i di\right)^{\frac{1}{1-\sigma}}$ is a price index over unconstrained varieties.

Demand for an unconstrained variety j is then given by

$$\begin{aligned} y_j &= b_j \frac{\left(I - \int_{i\in C} p_i \bar{y} di\right) P_U^{\sigma-1}}{p_j^{\sigma}} \\ &= b_j \frac{I_U P_U^{\sigma-1}}{p_j^{\sigma}}, \end{aligned}$$

where $I_U \equiv I - \int_{i\in C} p_i \bar{y} di$ are the aggregator's expenses over unconstrained varieties.

C Equilibrium conditions

First-order conditions for p_{it} and k_{it} (i-subscripts suppressed in the following):

$$
\begin{aligned}
E\left[\frac{\xi}{\beta}\left(\Pi_t^{\text{ppi}}-1\right)\Pi_t^{\text{ppi}}+y_t^s\overline{rp}_t\left(\sigma-1\right)\int_0^{\bar{b}_t}\frac{b}{\bar{b}_t}df\left(b\right)\right] \\
= E\left[y_t^s\overline{rp}_t\left\{1-F\left(\bar{b}_t\right)+\sigma\int_0^{\bar{b}_t}\left(\frac{b}{\bar{b}_t}\right)^{\frac{2}{\alpha+2(1-\alpha)}}df\left(b\right)\right\}+\xi\left(\Pi_{t+1}^{\text{ppi}}-1\right)\Pi_{t+1}^{\text{ppi}}\right] \\
R_t-(1-\delta) = \beta E\left[\frac{\alpha}{2}\overline{rp}_t\frac{y_t^s}{k_t}\left\{\left[1-F\left(\bar{b}_t\right)\right]+\int_0^{\bar{b}_t}\left(\frac{b}{\bar{b}_t}\right)^{\frac{2}{\alpha+2(1-\alpha)}}df\left(b\right)\right\}\right]
\end{aligned}
$$

Firm supply y^s:

$$
y_t^s = \frac{\alpha}{\chi}\left(\frac{1-\alpha}{w_t}\right)^{\frac{2(1-\alpha)}{\alpha}}\overline{rp}_t^{\frac{\alpha+(1-\alpha)}{\alpha}}k_t
$$

Aggregate supply and factor demands from firms:

$$
\begin{aligned}
Y_t &= \bar{b}_t^{\frac{1}{\sigma-1}}y_t^s\left\{\left[\int_0^{\bar{b}_t}\frac{b}{\bar{b}_t}df\left(b\right)+\int_{\bar{b}_t}^{\infty}\left(\frac{b}{\bar{b}_t}\right)^{\frac{1}{\sigma}}df\left(b\right)\right]\right\}^{\frac{\sigma}{\sigma-1}} \\
L_t^d &= \frac{1-\alpha}{w_t}\overline{rp}_ty_t^s\left(\int_0^{\bar{b}_t}\left(\frac{b}{\bar{b}_t}\right)^{\frac{2}{\alpha+2(1-\alpha)}}df\left(b\right)+\left[1-F\left(\bar{b}_t\right)\right]\right) \\
CU_t &= \frac{\alpha}{2}\overline{rp}_ty_t^s\left(\int_0^{\bar{b}_t}\left(\frac{b}{\bar{b}_t}\right)^{\frac{2}{\alpha+2(1-\alpha)}}df\left(b\right)+\left[1-F\left(\bar{b}_t\right)\right]\right)
\end{aligned}
$$

Household optimality conditions (Euler equation, no-arbitrage, labor supply):

$$
\begin{aligned}
\frac{1}{C_t} &= \beta\mathcal{R}_tE\left[\frac{1}{C_{t+1}\Pi_{t+1}}\right] \\
\mathcal{R}_tE\left[\frac{1}{C_{t+1}\Pi_{t+1}}\right] &= E\left[\frac{R_t}{C_{t+1}}\right] \\
w_t &= \varphi_tL_t^{\varepsilon}C_t^{\tau}
\end{aligned}
$$

Definition of producer price inflation:

$$
\Pi_t^{\text{ppi}} = \Pi_t\frac{\overline{rp}_t}{\overline{rp}_{t-1}}
$$

Market clearing conditions:

$$
\begin{aligned}
k_t &= K_t \\
L_t^d &= L_t
\end{aligned}
$$

Taylor rule:

$$\log\left(\mathcal{R}_t\right) = \log\left(1/\beta\right) + CB_{rf}\log\left(\Pi_t\right)$$

Aggregate resource constraint:

$$Y_t = C_t + CU_t + \frac{\xi}{2}\left(\Pi_t^{\text{ppi}} - 1\right)^2 + \left[K_{t+1} - (1-\delta)\,K_t\right] + G_t$$

Aggregator's zero-profit condition $I_t = \mathcal{P}_t Y_t$:

$$Y_t = \overline{rp}_t y_t^s \left(\frac{\int_0^{\bar{b}_t} b df(b)}{\bar{b}_t} + \left[1 - F\left(\bar{b}_t\right)\right]\right)$$

D Variance of firm profitability

A firm's profitability is given as

$$\begin{aligned} p_i SR_i &= \frac{p_i y_i}{k_i^{\alpha} l_i^{1-\alpha}} \\ &= p_i \left(\frac{\tilde{k}_i}{k}\right)^{\alpha} \\ &= \frac{p_i^2}{\mathcal{P}}\left(\frac{\min\left\{b_i, \bar{b}\right\}}{\bar{b}}\right)^{\frac{\alpha}{2-\alpha}}\left(\frac{\alpha}{2}\frac{1}{\chi}\right)^{\alpha}\left(\frac{1-\alpha}{w}\right)^{1-\alpha}. \end{aligned}$$

Since all firms set the same price $p_i = p$, it follows for the variance of log profitability

$$\begin{aligned} \text{Var}\left(\log(p_i SR_i)\right) &= \text{Var}\left(\frac{\alpha}{2-\alpha}\log\left(\min\left\{b_i, \bar{b}\right\}\right) + \log\left[\left(\frac{1}{\bar{b}}\right)^{\frac{\alpha}{2-\alpha}}\frac{p^2}{\mathcal{P}}\left(\frac{\alpha}{2}\frac{1}{\chi}\right)^{\alpha}\left(\frac{1-\alpha}{w}\right)^{1-\alpha}\right]\right) \\ &= \left(\frac{\alpha}{2-\alpha}\right)^2 \text{Var}\left(\log\left(\min\left\{b_i, \bar{b}\right\}\right)\right). \end{aligned}$$

E Results under discount factor shock

We rerun the simulation modeling demand shocks as fluctuations in the household's discount factor rather than her relative preference for leisure. Under this specification, lifetime utility is

$$E\left[\sum_{t=0}^{\infty}\phi_t \beta^t\left(\log C_t - \frac{L_t^{1+\varepsilon}}{1+\varepsilon}\right)\right].$$

Now the exogenous random variable ϕ_t follows a lognormal AR-1 process, where the autocorrelation parameter ρ_ϕ remains at 0.9 and the standard deviation of innovations σ_ϕ is again chosen to match the empirical standard deviation of output around its trend, implying a value of $\sigma_\phi = 0.0073$.

Table 6 compares the results under leisure preference shock and discount factor shock, respectively.

Table 6: Results under discount factor shock

Moment	Baseline	Discount factor	Empirical
Output asymmetry	3.27% vs −3.46%	3.28% vs −3.45%	2.73% vs −3.43%
Cross-sectional volatility	0.48%	0.44%	2.84%
Aggregate volatility	1.50% vs 1.66%	1.30% vs 1.47%	0.95% vs 1.41%
Solow residual	0.46%	0.59%	0.87%
Fiscal multiplier	0.87 vs 0.99	0.87 vs 1.02	–

Discount factor: Model results using the alternative shock specification of appendix E. Baseline: Leisure preference shock as in section 4. Output asymmetry: Average of largest positive and negative output deviations from trend. Cross-sectional volatility: Average increase of cross-sectional dispersion in firm profitability in recessions. Aggregate volatility: Standard deviation of output growth in expansions vs recessions. Solow residual: Standard deviation of the aggregate Solow residual. Moments are explained in more detail in the results section 5.2.

F Capacity constraints in a flexible-price model

To compare the importance of nominal rigidities we qualitatively compare the results of the sticky-price baseline model to a version without pricing friction. We introduce three changes: 1) Firms only set their capacity level k_i one period in advance, but sell their goods under flexible prices after production occurs. From the CES aggregator's demand function it follows that prices are a function of quantity produced y_i and weight b_i, given by $p_i = (b_i Y/y_i)^{1/\sigma}$, as is standard in the monopolistic competition setup with flexible prices. 2) With prices no longer fixed, increasing demand for one firms product will increase both this variety's price and quantity produced without bound, even in the face of the convex utilization costs. To re-introduce capacity constraints in the simplest possible way, we therefore limit firms' utilization $\tilde{k}$ to a fixed proportion of installed capacity k, requiring that $\tilde{k} \leq \gamma k$. The parameter γ determines how tightly the constraint binds and how many firms are capacity constrained in steady state. There will be a new cutoff level $\bar{b}$ for firm-specific demand at which firms produce with their full capacities so that $\tilde{k} = \gamma k$. 3) We replace the aggregate demand shocks with technology shocks. To this end, we modify the intermediate goods firms' production function to $y_i = A\tilde{k}_i^{\alpha} l_i^{1-\alpha}$ where A is the aggregate productivity level common to all firms.

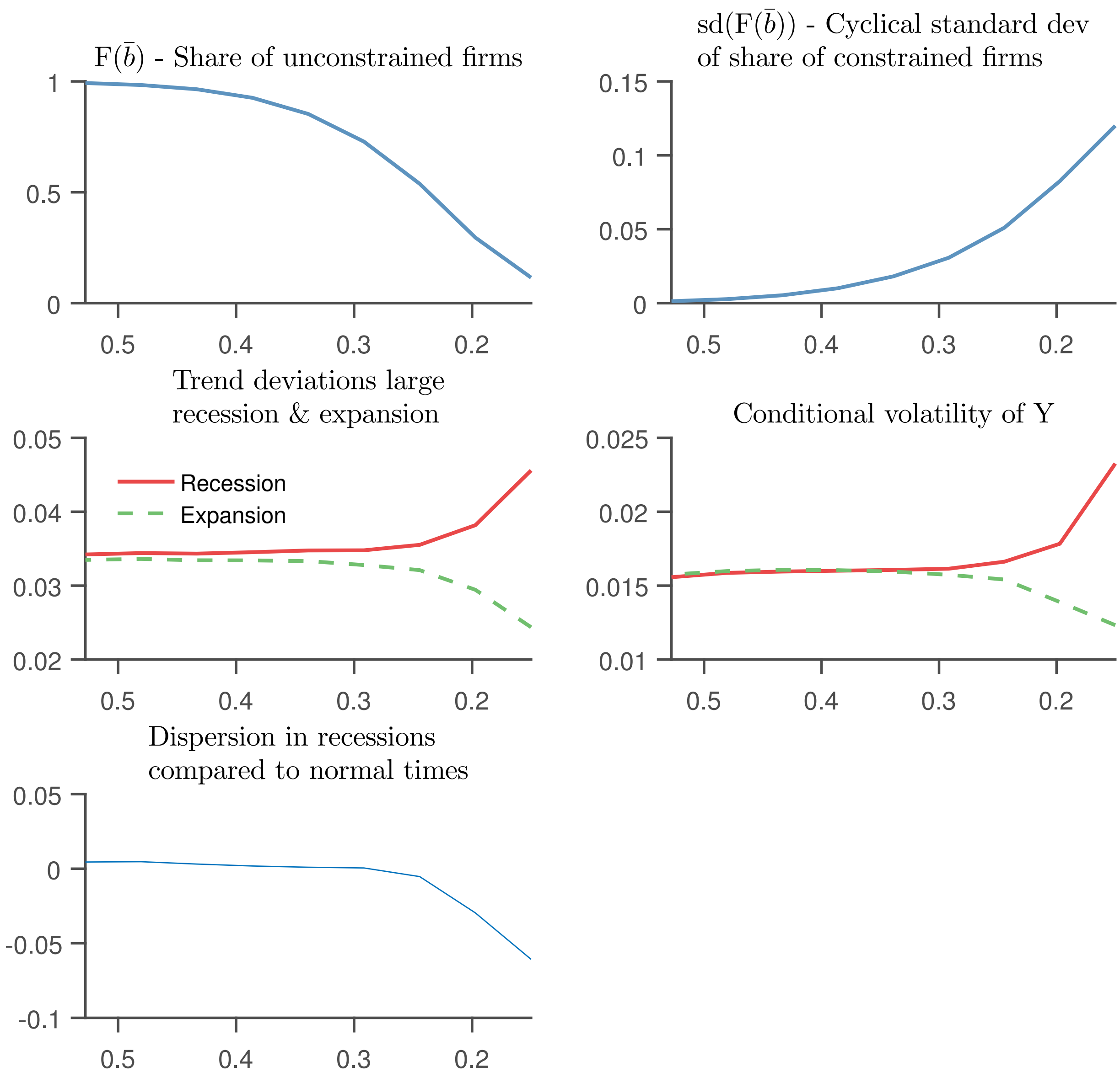
We keep all parameters the same as in the baseline model where applicable, and choose the process for A so that output has the same standard deviation and persistence as in the baseline. This leaves the parameter γ which affects how costly it is to hold capacity – with high γ firms will have a lot of room for adjusting their utilization leading to a low probability of the constraint binding, whereas with low γ firms are more likely to be constrained.

To see the effects as capacity constraints gradually begin to bind, in figure 11 we plot different model outcomes as a function of (decreasing) γ. As γ decreases, more firms become constrained (top left panel) and the number of firms moving in and out of the constraint over the cycle increases (top right panel). As discussed in the baseline model, fluctuations in the share of constrained firms drives the effects of constraints. The same intuition applies here: output asymmetry (center left panel) and conditional output volatility (center right panel) increase with the cyclical variation in capacity constraints. There is, however, an important qualitative difference in the behavior of profitability dispersion across firms, which becomes *pro*cyclical, in contrast to the results in the baseline model and in the data. Recall that profitability is given as $p_i y_i / (A k_i^{\alpha} l_i^{1-\alpha})$. With flexible prices, p_i will still increase as a function of firm demand b_i even if the firm is in the constrained region $b_i > \bar{b}$, implying that profitability increases even as y_i and l_i are constant. In fact, profitability increases more steeply in b_i once a firm becomes constrained.[13] Therefore, in a boom when more firms become constrained, the right tail of the profitability distribution fans out and cross-sectional dispersion increases. The bottom left panel shows the difference in dispersion between recessions and normal times under flexible prices. We see that as the relevance of constraints increases, profitability in recessions becomes more and more concentrated. Finally, since the model is driven by technology shocks, true TFP is procyclical, and due to the flexible prices there are no demand effects of government spending on output.

Summarizing, the price rigidity has two main effects in the main model (beyond endogenizing the capacity constraint): 1) It generates the empirically observed countercyclicality in firm-level profitability dispersion, and 2) it allows aggregate demand shocks to move the economy, which in turn generates an output multiplier of government spending as well as introduces a difference in the cyclicality between measured Solow Residual and true TFP. The price rigidity does not qualitatively affect the impact of the constraints on aggregate output asymmetry and aggregate volatility.

[13]This is because for $b_i \leq \bar{b}$ profitability is proportional to $b_i^{\frac{2-\alpha}{\sigma(2-\alpha+\alpha\sigma)}}$, whereas if $b_i > \bar{b}$ it is proportional to $b_i^{\frac{1}{\sigma}}$ and $\frac{1}{\sigma} > \frac{2-\alpha}{\sigma(2-\alpha+\alpha\sigma)}$.

Figure 11: Flex-price model outcomes as function of γ



Notes: Top left panel: Share of constrained firms in steady state $F(\bar{b})$. Top right panel: sd($F(\bar{b})$) over time measures how many firms move in and out of the constraint over the cycle. The following outcome measures are as defined in the baseline model: Center left panel: Output asymmetry (mean absolute value of deviation trend output for 1/6th of periods with highest and lowest output, respectively). Center right panel: Conditional volatility (standard deviation of output growth conditional on being in the 1/6th of periods with highest and lowest output, respectively). Bottom left panel: Extra dispersion in recessions (mean profitability dispersion in 1/6th of periods with lowest output relative to average).