

Business Cycle Implications of Capacity Constraints under Demand Shocks

Florian Kuhn* Chacko George^{†‡}

October 15, 2015

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: Capacity constraints, asymmetric business cycles, economic volatility, Solow residual, fiscal multiplier

*Department of Economics, Binghamton University, PO Box 6000, Binghamton, NY 13902-6000, email: floriankuhn@utexas.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. Florian Kuhn is grateful to his advisors Matthias Kehrig and Olivier Coibion for their guidance. 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 acyclicity 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 not 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), Bilal 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. (2014), 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 paper is structured as follows: In the next section 2 we review the stylized facts established by recent literature. 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

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 cyclical 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 deviations in output from trend are likely negative 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)). 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. (2014), 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 $E[(x - \mu)^3/\sigma^3]$. 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 3.24% in an expansion and -3.45% 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 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.

Broadly, there have been two, not mutually exclusive, approaches to explain the negative correlation of profitability risk 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 (2014).

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

¹In 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.

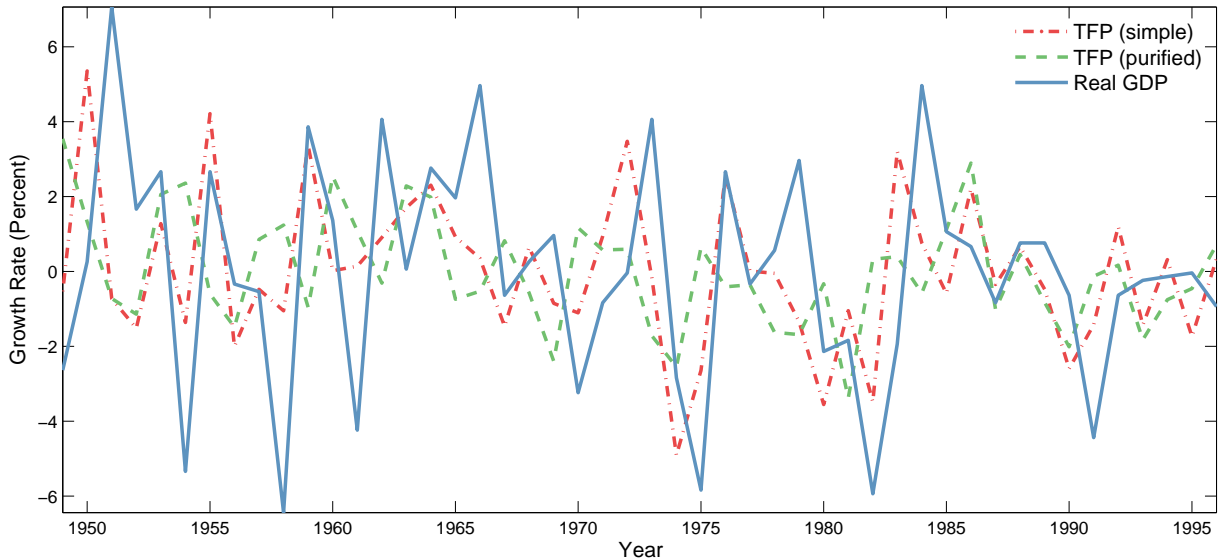
Table 1: Strong recessions larger than strong expansions

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 $E[(x - \mu)^3/\sigma^3]$.

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.

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.

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. Looking at the periods with the largest trend-deviations we find a difference of around 40% between expansions and recessions in the US data, as documented in table 4 in section 5.2.3.

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 Fernald² (see Fernald (2012)) which corrects *only* for intensity of capital and labor utilization. This allows us to check if utilization is indeed responsible for the difference in

²Data available at www.frbsf.org/economic-research/economists/jferald/quarterly_tfp.xls

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 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.

As suggested by Wen (2004) and Basu et al. (2006), demand shocks under variable capacity utilization are a possible explanation of this fact. Alternatively, Bai et al. (2012) provide an example of a search model in which demand shocks can show up as productivity shocks when 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 (2012b) 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 (2012a), 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 (2014) 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 (2015) 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 (2015) 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 Michailat (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.

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

$$\max_k -Rk + \int_0^k b \, dF(b) + [1 - F(k)] k.$$

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 $k_i = k = 1 - R$.

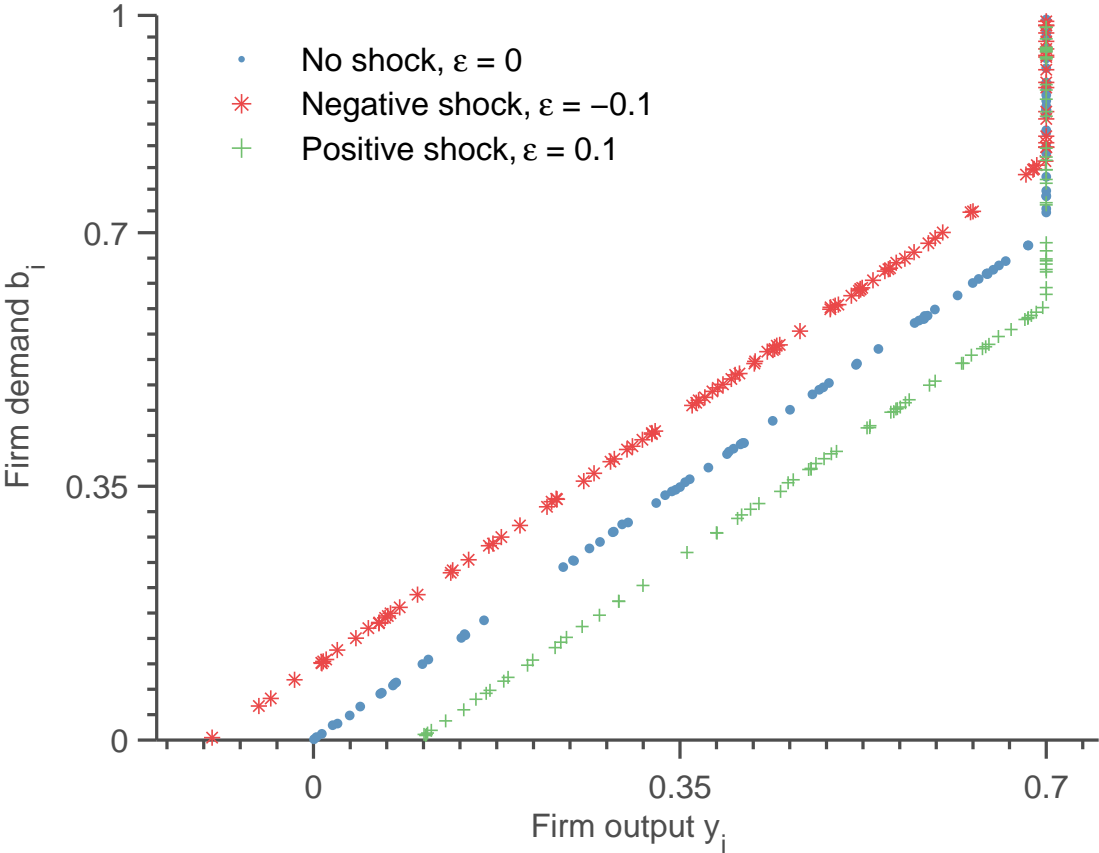
Output fluctuations Aggregate output under a uniform distribution over b between 0 and 1 is $Y = \int_0^{1-R} b \, db + R(1 - R) = \frac{1}{2}(1 - R^2)$. Now there is an unexpected shift in the mean by ϵ , so that b is drawn uniformly from the interval $(\epsilon, 1 + \epsilon)$. Aggregate production 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 $d^2Y/d\epsilon^2 = -\epsilon$ an additional small increase in aggregate demand affects output less, the higher aggregate demand already is.

The government multiplier and the asymmetry in output are therefore closely related. They are not quite measuring the same thing however. 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 $\epsilon = \pm 0.1$.

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 \bullet : 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.

Cross-sectional and aggregate volatility: The example also illustrates that aggregate fluctu-

ations 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 $\text{Var}(y_i)$ 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 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. 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.

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

$$Y = \left(\int b_i^{\frac{1}{\sigma}} y_i^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}.$$

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 $I \geq \int p_i y_i di$, 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 limit⁴ by \bar{y} , it therefore has to consider a continuum of inequality

³Kuhn (2014) 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).

⁴In 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.

constraints $y_i \leq \bar{y} \forall i$. The problem can then be expressed as

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

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

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

with $I_U \equiv \int_{y_i < \bar{y}} p_i y_i di$ the budget spent on unconstrained varieties and $P_U^{1-\sigma} \equiv \int_{y_i < \bar{y}} p_i^{1-\sigma} di$ a price index over unconstrained varieties.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 and p 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

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

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. 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).

⁵As noted by Fagnart et al. (1999) the demand function for the constrained varieties is undefined, and y^d denotes demand for the unconstrained varieties.

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

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 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

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

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

and so the cost function as

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

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 it unprofitable to produce. Since output increases with the level of demand shock, the cutoff output quantity will be associated with a cutoff level of 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

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

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 is given by

$$y = \min \left\{ y^d, y^s \right\}. \quad (4)$$

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

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

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

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

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

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

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

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

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

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

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 . There is a random weight φ_t shifting the relative preference of consumption and leisure and which will serve as an aggregate demand shock. This formulation of preferences is consistent with existence of a balanced growth path. Because consumption and leisure are separable, the household's labor supply function is not concave. This allows us to isolate the effects of variable capacity utilization from the mechanism that drives the results of Sims and Wolff (2015).

Separability between consumption and leisure precludes the concavity in the household's labor supply function that drives the results in Sims and Wolff (2015) which helps us isolate the effects of variable capacity utilization on the firm side.

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 $\tilde{\pi}_t \equiv \int \pi_{it} - [R - (1 - \delta)]k_{it} - \frac{\xi}{2} \left(\frac{p_{it}}{p_{i,t-1}} - 1 \right)^2 di$ 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

$$w_t = \varphi_t L_t^\varepsilon C_t, \quad (7)$$

the Euler equation

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

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

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

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:

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

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 $\min\{b_i/\bar{b}, 1\}$. This makes integration over i straightforward and gives aggregate capital utilization costs and labor demand as

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

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

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

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

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 $\bar{r}p_t = p_t/\mathcal{P}_t$, 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

$$\Pi_t^{\text{ppi}} = \Pi_t \frac{\bar{r}p_t}{\bar{r}p_{t-1}} \quad (14)$$

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 $\left\{ R_t, \mathcal{R}_t, w_t, \bar{r}p_t, \Pi_t, \Pi_t^{\text{ppi}} \right\}_{t=0}^{\infty}$, 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 B 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 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 (2015). 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 $\sigma_\varphi = 0.004$ is chosen to match the empirical standard deviation of quarterly postwar US GDP of 1.8% 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 (2015).

The second group of parameters describes the utilization cost function and the variance of idiosyncratic shocks. We assume the distribution of the *iid* idiosyncratic shock b_i to be log-normal and set the parameter σ_b governing its variance to match the variance of innovations

Table 2: Baseline Calibration

Parameter	Value	Meaning	Calibration
<i>Standard parameters</i>			
α	$\frac{1}{3}$	$y_i = \tilde{k}_i^\alpha l_i^{1-\alpha}$	Capital share
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	0.004	Shock variance	$\text{sd}(Y_t) = 1.8\%$
ξ	75	Scale price adj. cost	Ireland (2001)
CB_{rf}	1.75	Taylor rule	Sims and Wolff (2015)
<i>Model-specific parameters</i>			
χ	1	Scale utiliz. cost	See text
σ_b	0.67	sd idiosync. shocks	$\text{sd}(\Delta\text{TFP}_i) = 0.185$

to firm profitability in the data. In particular, both Syverson (2011) and Ilut et al. (2014) find a standard deviation of innovations to the log of firm TFP of around 0.185. We match this to average growth rates in the firms' measured TFP in the model. We are unaware of direct empirical estimates for the parameter χ ; here we set the parameter to 1. We are not very concerned with this issue for two reasons. First, varying the parameter over the admissible range for determinacy implied by the Blanchard-Kahn conditions changes quantitative results only minimally. Second, the fact that the utilization cost parameter is not well identified inside the model can also be observed in other models, see for example Christiano et al. (2005). We show sensitivity of the results to this parameter in section 5.2.6.

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 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 model is 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 here since we want to account for the nonlinearities generating differences between positive and negative shocks. Under linearization these differences would be lost.

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.⁷

Most notable is the strong reaction to the shock on impact in period 1. With prices set one period in advance the usual “New Keynesian” effect of demand shocks via relative prices is fully concentrated in period 1. What remains of the shock in periods 2 and later is primarily driven by the supply side effect of reduced household willingness to work and reduced capital stock from period 1, as well as the fact that firms’ price adjustment costs prevent a full alignment of relative prices in period 2.

As expected, capacity utilization drops along with aggregate output. The share of firm below their capacity constraint $F(\bar{b})$ 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.

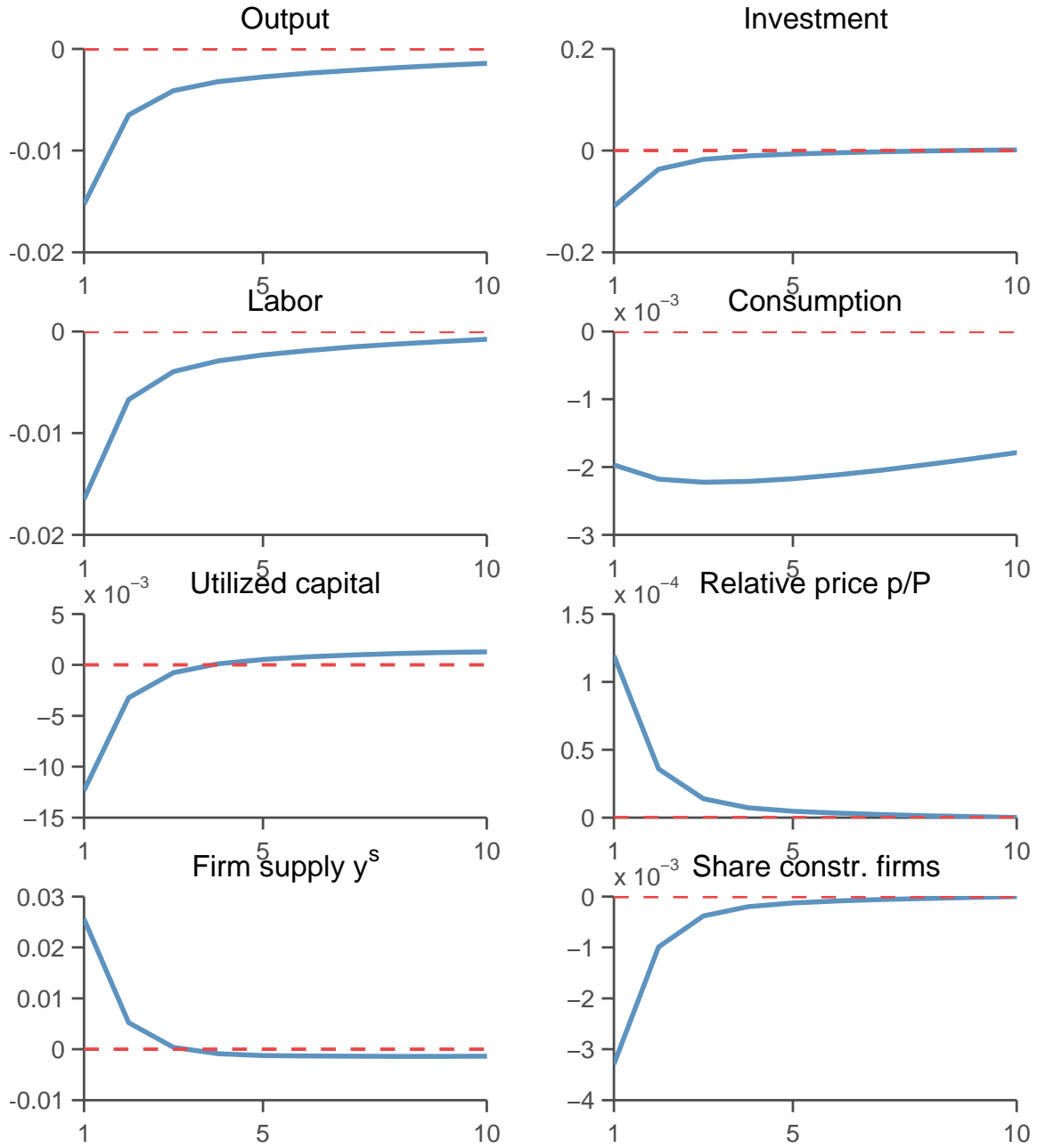
5.2.2 Output asymmetry

We now turn to an assessment of the implications for the stylized facts in general equilibrium. Quantitatively, the model explains around 1/4 of the observed asymmetry in output, and explains fluctuations in the fiscal multiplier of around 0.12.

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

⁷More 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 7.

Table 3: Asymmetry of Output Levels and Other Variables

Variable	Model		Data	
	Mean pos vs neg	Skewness	Mean pos vs neg	Skewness
Levels				
Output Y	3.24% vs -3.45%	-0.11	2.73% vs -3.43%	-0.46
Labor L	3.35% vs -3.54%	-0.12	3.11% vs -3.99%	-0.40
Investment I	18.97% vs -23.12%	-0.43	13.08% vs -17.12%	-0.53
Consumption C	1.46% vs -1.47%	-0.02	2.23% vs -2.24%	-0.00
Growth rates				
Output ΔY	3.16% vs -3.15%	0.01	2.30% vs -2.25%	0.19
Labor ΔL	3.45% vs -3.43%	0.01	1.63% vs -2.03%	-0.55
Investment ΔI	23.32% vs -23.04%	0.02	10.00% vs -11.30%	-0.21
Consumption ΔC	0.36% vs -0.37%	-0.06	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.

highest output to the $N/2$ periods where output is lowest. As shown in Table 3, the average large recession in that sense is -3.45% below trend, whereas the average large expansion is 3.24% above trend. Output is also negatively skewed with a coefficient of -0.11 .

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 0.21 percentage points (or a bit more than 6%) deeper than expansions. As the model was calibrated to match the standard deviation of HP(1600)-filtered, the closest comparable measure is the first row of Table 1 showing a relative difference of 23%, or 0.7 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 $p_i S R_i = p_i y_i / (k_i^\alpha l_i^{1-\alpha})$ 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 $y_i / (\tilde{k}_i^\alpha l_i^{1-\alpha})$ 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 $\min\{b_i, \bar{b}\}$ with

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

(see appendix C).

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 compared to the long-run. If similarly we look at the 1/6 of simulation periods in which output is lowest we find that profitability dispersion increases by 0.51% in recessions, implying that the model captures 18% of the cross-sectional volatility.

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 $\text{corr}(\text{sd}(\log(SR_i))_t, Y_t) = -0.93$ 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 a 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 1.49% 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 in a recession is 39.5% higher in recessions than in booms and thus fluctuates a bit 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

Table 4: Aggregate risk: Conditional Volatilities of Aggregate Variables

Variable	Model		Data	
	Expansion	Recession	Expansion	Recession
Output Y	1.49%	1.66%	0.95%	1.41%
Labor L	1.59%	1.77%	0.68%	1.08%
Investment I	9.66%	11.97%	5.21%	5.91%
Consumption C	0.19%	0.20%	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 $\text{sd}(\log X_{t+1} - \log X_t | X_t \text{ in recession})$. 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).

and Bayer (2013) obtain a difference of around 35% between booms and recessions, in line with the empirical values found here. Based on these estimates, the model covers between a third to a quarter of observed fluctuations in aggregate output volatility.

5.2.4 Aggregate Solow residual

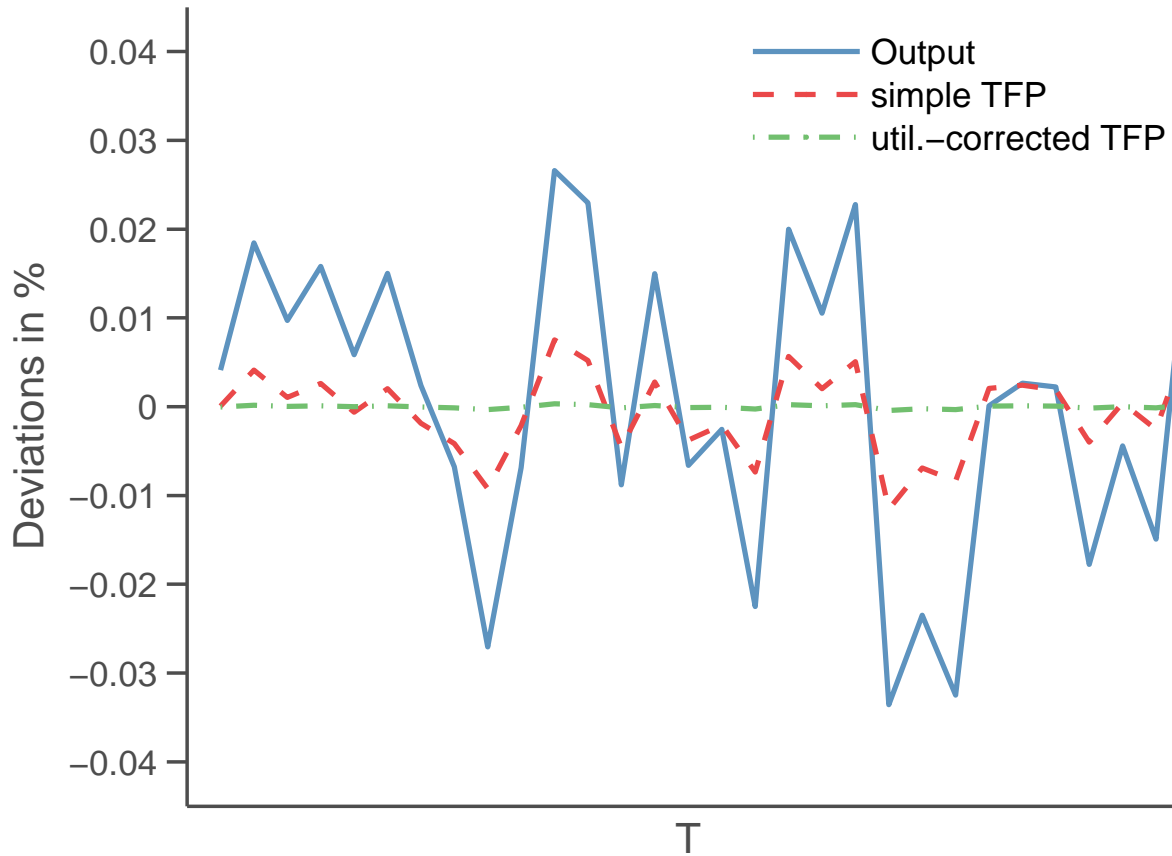
We construct the aggregate Solow residual in a similar way as its firm-level equivalent. We compute the uncorrected Solow residual as $SR_{\text{simple},t} = Y_t / (K_t^\alpha L_t^{1-\alpha})$ using aggregate capital in the denominator, and the corresponding version corrected for utilization as $SR_{\text{corr},t} = Y_t / (\tilde{K}_t^\alpha L_t^{1-\alpha})$ where $\tilde{K}_t = \int_i \tilde{k}_{it} di$ is defined as the aggregate utilized capital. Figure 4 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 (2012), the correlation between the simple TFP measure with output is strong with a value of 0.76. On the other hand, utilization-corrected productivity barely moves over the cycle.⁸ The standard deviation of simple TFP growth in the simulated series is 0.52%. This value is a little smaller than the corresponding measure in John Fernald’s quarterly dataset where the uncorrected Solow residual grows with a standard deviation of 0.87%.

5.2.5 Fiscal multiplier

Finally, we consider the cyclicity of the contemporaneous fiscal multiplier dY_t/dG_t . In constructing it we follow Sims and Wolff (2015) 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 $(Y^{S+\Delta G} - Y^S) / \Delta G$. The value

⁸Strictly, even utilization-corrected TFP fluctuates over time because of changes in the composition of input factors and their allocation between firms. Since corrected TFP has a very small variance (it has a standard deviation of 0.00018), however, even a tiny amount of noise —like measurement error— renders it acyclical.

Figure 4: 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.

of the multiplier when output is in its top quintile is computed the same way. We obtain values of 1.07 for the multiplier in a recession, and 0.95 for a multiplier in a boom. These results are close to what Sims and Wolff (2015) find in their DSGE model using a different mechanism.

5.2.6 Role of heterogeneity, discussion and sensitivity

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 5 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 . Why is this? What matters is the average change in the share of constrained firms over the cycle, and not its absolute level. Those differences in $F(\bar{b}_t)$ between expansion and recessions are largest for an interior value of σ_b . At a low value of 0.3 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.) 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 fiscal multiplier in the bottom right graph of Figure 5. 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 cyclical nature of the multiplier is strongest when the fluctuations in $F(\bar{b}_t)$ 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.

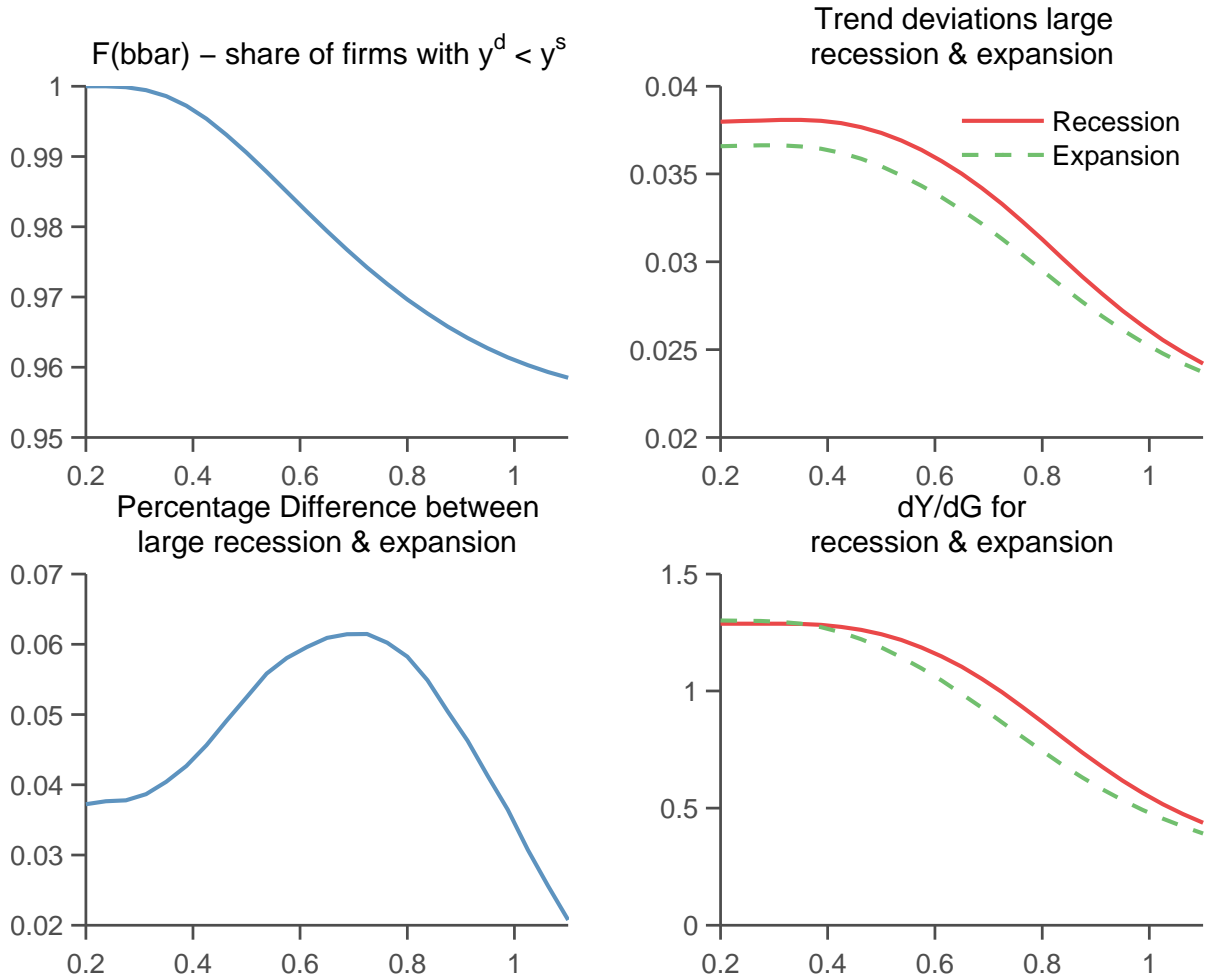
Summarizing, the firm heterogeneity causing capacity constraints to bind matters in this model because it *generates* cyclical nature in the fiscal multiplier and cross-sectional profitability dispersion, and *amplifies* the deepness of recessions.

Sensitivity to utilization cost In the baseline calibration we chose the scale parameter of the utilization cost function to be 1 in lack of more direct empirical estimates. Table 5 shows the specification is not very sensitive with respect to parameter values. The table displays results for recessionary and expansionary output deviations and multipliers for alternative values for χ . Both statistics increase only minimally in the parameter. There is an upper limit for its domain near $\chi = 1.4$ implied by determinacy of the model (otherwise the Blanchard-Kahn conditions are violated).

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.

Figure 5: 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.

Table 5: Sensitivity with Respect to Cost Function

Parameter value	Output asymmetry	Multipliers
$\chi = 0.01$	3.29% vs -3.40%	0.66 vs 0.75
$\chi = 0.1$	3.26% vs -3.41%	0.73 vs 0.84
$\chi = 0.5$	3.25% vs -3.44%	0.85 vs 0.97
Baseline	3.24% vs -3.45%	0.95 vs 1.07
$\chi = 1.4$	3.24% vs -3.44%	1.01 vs 1.13

Notes: Alternative values of utilization cost parameter χ . Baseline: $\chi = 1$. All remaining parameters are held constant, except for the aggregate shock variance σ_φ which needs to be adjusted to match its targeted moment. The value of $\chi = 1.4$ is close to the largest admissible value to not violate determinacy of the model.

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 -9.57% and 8.12%, respectively, and the recessionary and expansionary fiscal multipliers become 1.17 and 0.85, respectively. Similar effects 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 procyclical Solow residual.

A calibrated New Keynesian model yields differences in output between booms and recessions of around 0.21 percentage points, such that the model explains more than a quarter of the 0.7 percentage-point difference we find empirically. While the empirical literature has not settled on the size of fluctuations in the government spending multiplier over the cycle, in our basic model we find a multiplier of on average 0.95 in booms and 1.07 in recessions. The multiplier increases 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 this approach would be to gain more realism in the description of firm behavior. A stronger intertemporal dimension

could be added to the problem by including more heterogeneity in price-setting and investment behavior. This could yield new testable implications for the mechanism and make the cross-sectional aspects of the model more accurate.

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. and Y. Gorodnichenko (2012a). Fiscal multipliers in recession and expansion. In A. Alesina and F. Giavazzi (Eds.), *Fiscal Policy after the Financial Crisis*.
- Auerbach, A. J. and Y. Gorodnichenko (2012b, 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.
- 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, 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. Cho (1994). Cyclical factor utilization. *Journal of Monetary Economics*.
- 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.
- 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’Erasmus, 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? (September).

- 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. (2012). 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*.
- Gilchrist, S. and J. C. Williams (2000). Putty-Clay and investment - A Business Cycle Analysis.pdf. *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*.
- Hansen, G. and E. Prescott (2005). Capacity constraints, asymmetries, and the business cycle. *Review of Economic Dynamics*.
- Iltut, C., M. Kehrig, and M. Schneider (2014). Slow to Hire, Quick to Fire: Employment Dynamics with Asymmetric Responses to News.
- Iizetzki, 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.
- Knüppel, M. (2014). Can Capacity Constraints Explain Asymmetries of the Business Cycle? *Macroeconomic Dynamics* 18(01), 65–92.
- Kuhn, F. (2014). Endogenous profitability risk. *Working paper, The University of Texas at Austin*.
- 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 (2014). Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data.
- Sichel, D. (1993). Business cycle asymmetry: a deeper look. *Economic Inquiry*.
- Sims, E. and J. Wolff (2015). 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).

A Aggregator's demand function

The aggregator's problem is to

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

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

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

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

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

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.

B Equilibrium conditions

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

$$\begin{aligned}
& E_S \left[\xi \left(\Pi_t^{\text{ppi}} - 1 \right) \Pi_t^{\text{ppi}} + y_t^s \bar{r} \bar{p}_t (\sigma - 1) \int_0^{\bar{b}_t} \frac{b}{\bar{b}_t} df(b) \right] \\
& = E_S \left[y_t^s \bar{r} \bar{p}_t \left\{ 1 - F(\bar{b}_t) + \sigma \int_0^{\bar{b}_t} \left(\frac{b}{\bar{b}_t} \right)^{\frac{\psi}{\alpha + \psi(1-\alpha)}} df(b) \right\} + \xi \left(\Pi_{t+1}^{\text{ppi}} - 1 \right) \Pi_{t+1}^{\text{ppi}} \right] \\
R_t - (1 - \delta) & = E_S \left[\frac{\alpha(\psi - 1)}{\psi} \bar{r} \bar{p}_t \frac{y_t^s}{k_t} \left\{ [1 - F(\bar{b}_t)] + \int_0^{\bar{b}_t} \left(\frac{b}{\bar{b}_t} \right)^{\frac{\psi}{\alpha + \psi(1-\alpha)}} df(b) \right\} \right]
\end{aligned}$$

Firm supply y^s :

$$y_t^s = \left(\frac{\alpha}{\chi} \right)^{\frac{1}{\psi-1}} \left(\frac{1-\alpha}{w_t} \right)^{\frac{\psi(1-\alpha)}{\alpha(\psi-1)}} \bar{r} \bar{p}_t^{\frac{\alpha+\psi(1-\alpha)}{\alpha(\psi-1)}} k_t$$

Aggregate supply and factor demands from firms:

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

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

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

Definition of producer price inflation:

$$\Pi_t^{\text{ppi}} = \Pi_t \frac{\bar{r} \bar{p}_t}{\bar{r} \bar{p}_{t-1}}$$

Market clearing conditions:

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

Taylor rule:

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

Aggregate resource constraint:

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

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

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

C Variance of firm profitability

A firm's profitability is given as

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

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

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