Capacity Buffers:

Explaining the Retreat and Return of the Phillips Curve

Clement E. Bohr [*]

January 30, 2024

Click here for the most recent version

Abstract

Why did the Phillips curve gradually flatten since the 1960s, and what explained its sudden steepness in the goods sector during the COVID-19 pandemic? Firms face capacity constraints in production and hold excess capital to buffer against fluctuations in demand. The capacity buffer's size influences both the pass-through of demand fluctuations into sales and the sensitivity of a firm's pricing decisions to any realized changes in demand. Over the same time period as the flattening of the Phillips curve, firms' variable cost shares declined, capacity buffers rose, and idiosyncratic volatility of sales rose. This paper argues that the reduction in firms' variable cost shares induced firms to hold larger capacity buffers, which in turn increased the volatility of firm sales and flattened the Phillips Curve. The recent steepness of the Phillips curve in the goods sector arose from a collapse in the size of firms' effective capacity buffers, which was the result of the health precautions that increased the demand for goods and restricted production capacity.

---

[*] Northwestern University. Email: cebohr@u.northwestern.edu.

I am deeply grateful to my advisors Larry Christiano, Matthew Rognlie, and Martí Mestieri. I thank Giorgio Primiceri, George-Marios Angeletos, Martin Eichenbaum, Pooya Molavi, Chen Lian, Michael Cai, Javier Bianchi, Todd Schoellman, Marco Bassetto, Alisdair McKay, Jonathan Heathcote, Juan Pablo Nicolini, Lukas Nord, Gadi Barlevy, François Gourio, Bart Hobijn, Nicolas Werquin, Jing Zhang, Yuriy Gorodnichenko, Pascual Restrepo, Eric Swanson, and seminar participants at Northwestern University, the Federal Reserve Bank of Minneapolis, and the Federal Reserve Bank of Chicago for their useful comments and discussions. All errors are my own.

1 Introduction

At the core of our understanding of inflation is the Phillips curve, which describes a positive relationship between the rate of change in the price level and the level of economic output. Rather than being constant, however, estimates of the sensitivity of inflation to changes in real economic activity have gradually declined over time to the point where inflation is best understood to behave according to a random walk. The puzzle of the flattening Phillips curve then took a sharp turn during the COVID-19 pandemic as inflation returned to levels that had not been seen since the early 1980s. This pandemic inflation was mostly driven by price changes in the goods sector, where the slope of the Phillips curve again appeared steep.

Figure 1 illustrates these movements at the sectoral level by plotting a variation on what Stock and Watson (2020) call the Phillips Correlation. For the goods and services sectors, the inflation rate is plotted and regressed against the deviation in output relative to its trend.[1] The sample is divided into the same three time periods as in their analysis plus a subsequent one for the pandemic recovery to show how this correlation has changed over time. The figures suggest that the slope of the Phillips curve flattened gradually across both sectors between 1961 and 2020, which was then followed by a divergence between the sectors during the pandemic. The correlation in the goods sector increased along with a sharp rise in goods consumption, while the correlation in the services sector remained near zero with a sharp fall in services consumption.[2]

The underlying intuition behind the Phillips curve tends to go through the labor market: a low unemployment rate leads firms to pay their employees higher wages. The firms then pass these higher labor costs on to consumers in the form of higher prices. However, the relationship between wage inflation and the unemployment rate has not suffered from the same degree of flattening, and wage inflation did not underlie the initial surge in inflation during the COVID-19 pandemic. Similar to how inflation is associated with the amount of slack in the supply of labor, as given by the unemployment rate, it also depends on the amount of slack there is in the supply of capital, which is characterized by the size of firms' capacity buffers in production. A firm's capacity buffer measures the additional production beyond its current level that the firm's capital is able to accommodate. As shown in Figure 2, there has been a marked upward trend in the size of firms' capacity buffers. Their size, which is measured as the complement of the utilization rate of capacity, has approximately doubled from just above 10 percent of production capacity in the late 1960s to over 20 percent leading up to the pandemic.

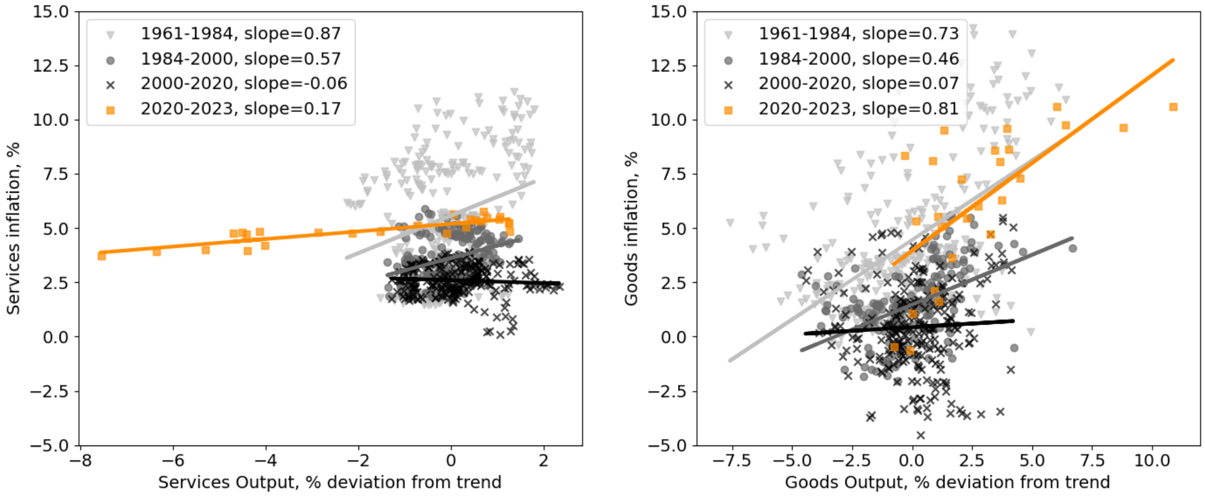
This paper explains the observed flattening of the Phillips curve as a consequence of the increasing size of firms' capacity buffers. Firms face capacity constraints in production and choose to have excess

---

[1] The trend is calculated using the Hodrick-Prescott filter parameterized for monthly data as described in Ravn and Uhlig (2002), and the inflation rate is calculated as the growth rate in the price-level over the following year. The accelerationist versions of the Phillips correlations, where the change in inflation is regressed on the deviation in output from trend, are depicted in Figure 16 in Appendix C. They provide the same general picture.

[2] The flattening of the Phillips curve was recently documented by Hazell et al. (2022) who found that it declined by a factor of two since 1980s, which is a more modest decline than what is suggested by Figure 1. The steepening of the Phillips curve during the COVID-19 pandemic has been documented by Cerrato and Gitti (2022) and Hobijn et al. (2023).

Figure 1: The Unstable Phillips Correlation


(a) Services Phillips Correlation

(b) Goods Phillips Correlation

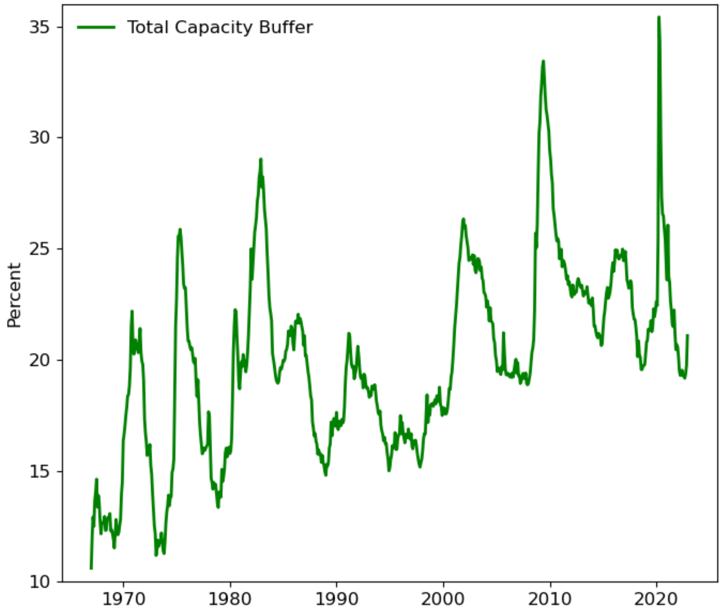
Notes: The panels depict estimates of the Phillips correlations for the services and goods sectors as defined by the regression slope of the sectoral inflation rate on the sectoral deviation of output from trend. The trend is calculated using the Hodrick-Prescott filter parameterized for monthly data as described in Ravn and Uhlig (2002), and the inflation rate is calculated as the growth rate in the price-level over the following year.

capacity on hand as a buffer against fluctuations in demand. The size of these capacity buffers, which are determined by the firm in expectation, dictates how sensitive a firm's pricing and investment decisions are to any temporary changes in demand. If the firm has a large capacity buffer, small variations in demand will easily be met by the firm's excess capacity, and it will have little incentive to adjust its investment and prices. On the other hand, if the firm only has a small capacity buffer, even a small rise in demand may lead the firm to want to substantially raise its prices and investment in order to avoid becoming capacity constrained and forgo additional profits. The gradual rise in the size of firms' capacity buffers made prices less sensitive to fluctuations in aggregate demand, thereby flattening the Phillips curve. Given the difficulty in its estimation, there is little consensus on the exact amount the Phillips curve flattened, though estimates tend to show at least a halving of its slope. The increase in the size of capacity buffers since the 1960s is estimated to have flattened the Phillips curve by up to 26 percent.

Larger capacity buffers also imply an increase in the idiosyncratic volatility of firm sales. Capacity buffers enable idiosyncratic variation in demand to pass through into production and sales, except for those firms whose buffers are insufficient to accommodate it. All else equal, larger capacity buffers imply that fewer firms will become constrained by their production capacity. As a result, more of the variation in demand will pass through into sales, leading to an increase in the volatility of the latter. This is shown to be true at the six-digit NAICS industry level rather than at the firm level, but the estimates are similar to those reported at the firm level in Comin and Philippon (2005).

An increase in price markups may have been what induced firms to hold larger capacity buffers. When

Figure 2: Capacity Buffer for the United States



Notes: Time series of the percent unused capacity for the industrial production sector as measured by the complement of the Federal Reserve's measure of the capacity utilization rate

firms set higher markups and have higher marginal profits, the opportunity cost from being capacity constrained is higher, which leads firms to accumulate larger capacity buffers. The larger markup over variable labor costs also implies a lower labor cost share as a smaller fraction of total revenues is allocated to labor. Consistent with these two implications of rising markups, industries that experienced larger rises in their capacity buffers also experienced larger declines in their labor shares.

An unverified prediction of the theory is that fiscal multipliers should have gotten larger as the capacity buffers grew in size. Fiscal stimulus is more easily accommodated by firms when fewer of them are capacity constrained, and interest rates do not become as elevated because of the weaker response of inflation to the demand shock. The theory predicts that the cumulative fiscal multiplier has increased by about 20 percent since the 1960s.

In addition to the analysis of the structural changes associated with the growing size of capacity buffers, new evidence is provided for the fact that firms' pricing decisions are dependent on the size of their capacity buffers. Using a smooth transition local projection model, the effects of monetary policy on prices and investment are estimated conditional on the size of firms' capacity buffers. It finds that investment and prices were unresponsive to monetary policy surprises when firms had large capacity buffers, which they had about half of the time between 1969 and 2008. In contrast, when capacity buffers were small, which made up about 10 percent of this time period, the response of investment and prices were substantially larger than the standard estimated responses that do not condition on the size of the capacity buffer. The capacity buffers and the responses of prices and investment were of an intermediate size during the remaining parts of the time period.

Lastly, the capacity buffer theory also provides a natural framework for understanding the surge in inflation that occurred during the COVID-19 pandemic. The initial restrictions on production that were imposed to minimize the spread of the virus temporarily reduced firms' production capacities. Despite firms having large capacity buffers on paper, they were reduced in practice since, in order to limit the spread of the virus, firms could not use their capital to the same degree as before. With a more limited effective capacity buffer, firms set higher prices to avoid becoming capacity constrained, and their prices became more sensitive to variation in demand. At the same time, health precautions also led households to switch their expenditures from services to goods, creating a large positive demand shock in the goods sector but a negative one to the services sector. In the goods sector, firms' smaller effective capacity buffers were insufficient to buffer against the large increase in demand, and they became constrained at their temporarily reduced capacities. The result was more upward pressure on prices, which led to a sharp rise in goods inflation along with a steepening in its correlation with output. These two shocks in combination can account for 84% of aggregate inflation at its first peak in the second quarter of 2021 when quantified in a New Keynesian model.

Following the literature review, the rest of the paper is structured as follows. Section 2 provides new motivational evidence from the effects of monetary policy shocks that capacity buffers are a state variable for the dynamics of inflation and investment. Section 3 lays out the theory of capacity buffers in production. Section 4 explains and provides evidence regarding how the secular rise in markups resulted in larger capacity buffers and a flatter Phillips curve. Section 5 embeds the firm theory into a multi-sector New Keynesian model in order to estimate the flattening of the Phillips curve and the growing fiscal multipliers. Section 6 applies the same model to assess the COVID-19 inflation. Section 7 concludes.

1.1 Related Literature

This paper emphasizes the size of a firm's capacity buffer as an endogenous choice, which determines the degree to which a notion of scarcity pricing is relevant, and analyzes its implications for the economy when prices are sticky.[3] To that end, this work builds on the quantity rationing theory of Malinvaud (1977), which features limits to production and uncertainty in demand, resulting in a precautionary motive to hold excess capacity. Fagnart et al. (1997) provided a particularly parsimonious formalization of it, which Fagnart et al. (1999) integrated into a dynamic general equilibrium setting, and Alvarez-Lois (2004, 2006) then extended into a New Keynesian framework. Recently, Boehm and Pandalai-Nayar (2022) applied it to estimate the convexity of industries' supply curves and found robust evidence that their steepness was increasing in their capacity utilization rate.[4]

---

[3] Scarcity (or hockey-stick) pricing is the workhorse model for electricity markets, but it has also been shown to exist in markets with differentiated products and where prices are more sticky, such as for cars (Israeli et al., 2022).

[4] Sun (2023) formulates a complementary theory for why firms operate with excess capacity, where their size increases the amount of demand for a firm's product because households perform partially undirected search for their consumption choices. Gilchrist and

The simplified putty-clay production theory used in these studies is augmented by introducing an explicit investment process with temporal rigidities for the choice of capital intensity—the fixed proportion in which capital and labor are combined. The rigidities in the choice of capital intensity make it an inflexible state variable for the firm alongside that of the capital stock, which allows the theory to better capture the short run dynamics in the size of the capacity buffers. The result is a putty-clay technology, which features a capital-labor elasticity of substitution that is gradually increasing in the duration of time considered, all while remaining highly tractable.[5] Moreover, the production theory is a generalization of and only one step removed from a standard CES production technology, which can be recovered by removing the temporal rigidities in capital intensity.[6]

The fact that capacity buffers have been gradually rising, or equivalently that the share of production capacity that firms are utilizing has been falling, has received little attention. In most cases when the time series is used for economic analysis, its trend has been viewed as a nuisance to be removed rather than considered as structurally important. This paper is the first to provide a structural theory that accounts for the rise in the size of firms' capacity buffers. Bansak et al. (2007) tested whether industries that adopt more high-tech capital possibly operate at lower utilization rates because newer more modular production technologies add greater flexibility in ramping production up and down. They found a small but significant effect of high-tech investment on the decline in capacity utilization rates.[7]

Much empirical work has addressed the price Phillips curve and the degree to which it has flattened. Using regional data back to 1978, Hazell et al. (2022) find that the Phillips curve has always been flat, and that much of its steepness in the early 1980s was due to declining inflation expectations. However, their analysis does not include the 1960s and 1970s when firms operated with small capacity buffers, and their point estimate of the Phillips curve slope does still decline by roughly half. Hooper et al. (2020) and McLeay and Tenreyro (2020) also use cross-regional data to explore how the slope of the Phillips curve has changed. Both find evidence that the wage Phillips curve has remained more pronounced, and that more aggressive monetary policy has contributed to making the Phillips curve appear flatter and identification of it harder. In agreement with those above, Del Negro et al. (2020) use both a SVAR and DSGE model to find that the price Phillips curve has flattened substantially while the wage Phillips curve less so. They also find

---

Williams (2000, 2005) also provide a theory for why firms operate well below their capacity constraints. Similar to here, the notion of capacity arises from a putty-clay production technology, but the underutilization arises due to uncertainty in the productivity of the capital stock, which results in some vintages of capital becoming too expensive to operate. This paper abstracts away from both of these dimensions to focus on the precautionary buffer mechanism. Importantly, these theories are not to be confused with the more common theory of capacity utilization of Greenwood et al. (1988) where the firm is provided with the additional degree of flexibility by allowing them to extract additional capital services out of a given stock of capital, which does not feature a notion of engineering capacity.

[5] The addition of temporal rigidities to the choice of capital intensity bridges the putty-clay production theory with the theory of choosing an appropriate technology for production of Jordà (2005) and Caselli and Coleman (2006). Similar to this paper, León-Ledesma and Satchi (2019) added adjustment costs to the choice of technology in order to create a smoothly increasing capital-labor substitutability over time. The production theory in this paper can therefore also be interpreted as the combination of the capacity utilization theory of Fagnart et al. (1997) with the appropriate technology theory of León-Ledesma and Satchi (2019).

[6] Rather than having the capital intensity be flexible ex-ante and fixed ex-post as in the standard putty-clay framework, here it is fixed ex-ante but gradually more flexible with time ex-post, in what would more appropriately be called clay-putty. Nevertheless, for the rest of this paper, it will be referred to under the standard terminology of putty-clay with which economists are generally familiar.

[7] Other studies that have explored the declining rate of capacity utilization include Pierce and Wisniewski (2018) and Gahn (2023).

that more aggressive monetary policy may have contributed to but cannot explain most of the observed flattening of the price Phillips curve.[8]

A few other structural theories have been proposed as explanations for its flattening. Baqaee et al. (2021) find that the reallocation from low to high productivity firms, which simultaneously occurs from a rise in demand, creates an endogenous positive productivity shock that counteracts the upward pressure on marginal costs. As the productivity distribution of firms has become more dispersed this mechanism has strengthened, leading to a 17 percent flattening of the Phillips curve. Rubbo (2023) finds that the slope of the Phillips curve declines as the cost share of intermediate inputs rises because the price rigidities compound at each step in the supply chain. Fujiwara and Matsuyama (2022) introduce a novel demand structure through which a reduction in the number of competitors raises markups but decreases the pass through rate from costs to prices. Increasing barriers to entry in production then could have led to a flattening of the Phillips curve.[9] Nothing about the rise in capacity buffers precludes these industrial organisation mechanisms from also having taken place.[10] Relative to these studies, however, the capacity buffer explanation is also consistent with other structural changes that occurred over the same time period, including the fall in capacity utilization rates and the rise in the volatility of sales, and the framework provides a seamless description of the rapid steepening of the Phillips curve in the goods sector during the COVID-19 pandemic.

There were many factors that likely contributed to the elevated levels of inflation during the COVID-19 pandemic. An important potential contributor has been a nonlinear Phillips curve, which fall into three camps depending on where the nonlinearity arises. In one, they arise directly from the demand schedule derived from the household household preferences (Harding et al., 2023). In another, they originate in labor market from a nonlinear wage Phillips curve (Benigno and Eggertsson, 2023; Gitti, 2023; Schmitt-Grohé and Uribe, 2022). In the third, they originate from capacity constraints in the product market instead (Comin et al., 2023). This paper pertains to the third, but differs from that of Comin et al. (2023) in that the capacity constraints have non-latent effects on the dynamics of inflation. That is, the capacity constraints create nonlinear effects even when they are not binding. Moreover, while they find that the combination of expansionary monetary policy and shocks to capacity constraints explain the inflationary takeoff of 2021, this paper argues that the sectoral consumption switch from services to goods combined with the temporary reductions in production capacity together explain the surge in inflation.

---

[8] Another prominent hypothesis for the weakening in the domestic relationship between prices and economic slack is globalization, although the evidence on it to date is mixed (Ihrig et al., 2010; Bianchi and Civelli, 2015; Auer et al., 2017). For the United States, it does not appear to have been particularly relevant due to its low degree of openness that only featured a modest rise.

[9] Whether entry costs have risen is debated. Gutiérrez and Philippon (2017) find evidence of rising entry costs from increasing regulations. However, Marto (2023) finds little evidence of higher entry costs being the reason for why markups have risen. Instead he finds that markups rose due to rising income levels. Consonantly, Atalay et al. (2023) find that markups at the product level rose due to a decrease in price sensitivity of consumers despite no increase in concentration at finer product levels.

[10] Most likely they have worked in synergy: for instance a reallocation to more productive firms with larger capacity buffers would amplify the endogenous positive productivity shock. Similarly, the flattening should be amplified if the capacity buffer has expanded at each link in a supply chain.

2 Capacity Buffer Dependent Effects of Monetary Policy

This section provides motivational evidence that firms' capacity buffers are an important state variable underlying the dynamics of prices and investment. When capacity buffers are narrow, either by design or due to unforeseen events, pricing and investment should be more sensitive to variations in demand compared to when capacity buffers are large. To test the potency of this state dependence for the aggregate US economy, the effects of monetary policy shocks are estimated conditional on the size of firms' capacity buffers. This is done using the Logit Smooth Transition Local Projection Model of Auerbach and Gorodnichenko (2012), where the outcome variable depends on a convex combination of two linear regimes. A state variable, in this case the size of capacity buffers, determines the relevant convex combination, i.e. the degree to which one regime is dominating at the time.

The implementation follows closely that of Tenreyro and Thwaites (2016) who estimate the differential effects of monetary policy shocks in expansions versus recessions. It differs in that the analysis is done at the monthly frequency spanning 1969 to 2008, and the regime is determined the size of firms' capacity buffers. Capacity buffers create a state variable that is structurally different from that implied by recessions and expansions. A recessionary state dictates that the economy spends a minority of its time in its slack recessionary regime since recessions are shorter than expansions. A capacity buffer state implies the opposite: the economy spends a majority of its time in the slack regime with large capacity buffers because only on rarely do firms operate with little excess production capacity. Moreover, due to the upward trend in the size of capacity buffers, the time periods where firms operate with small capacity buffers are more prevalent earlier on.[11]

The econometric specification is given by

$$y_{t+h} = \tau t + F(z_t)\left(\alpha_1^h + \beta_1^h m_t + \gamma_1' x_t\right) + (1 - F(z_t))\left(\alpha_0^h + \beta_0^h m_t + \gamma_0' x_t\right) + u_t. \tag{1}$$

The monetary policy shock in period t is denoted by m_t. The impact it has on the outcome variable y_{t+h} that is h months ahead is given by the coefficients β_0^h under the regime with large capacity buffers and β_1^h under the regime with small capacity buffers. A trend t, regime specific intercepts α^h, and a vector of controls x_t are also included as regressors.[12]

The regime transition function $F(z_t)$ takes on the logit specification,

$$F(z_t) = \frac{\exp(\theta z_t)}{1 + \exp(\theta z_t)}, \tag{2}$$

where z_t is a transformation of the state variable. In the baseline specification, $z_t = \bar{B} - B_t$ is the deviation in the size of the capacity buffer relative to a threshold $\bar{B}$, which is set to 17 percent. This level was chosen

---

[11] Figure 26 in Appendix D provides a visuals comparison between the state variables along with a more extended discussion.

[12] The controls include four lags of the outcome variable and one lag of the federal funds rate. Lag lengths were chosen based on the AIC criterion.

because it was near the peak size of the capacity buffers during the 1990s, a time when the Phillips curve was known to have already flattened significantly, but had not completely disappeared yet.[13] As such, this size of the capacity buffer should feature prominently in the transition between regimes. The parameter θ specifies the width of the transition space between regimes around the threshold. A smaller θ implies a larger change in the capacity buffer is needed to transition from one regime to the other. It is set such that 99 percent of the transition happens within 4 percentage points around $\bar{B}$.[14]

The monetary policy shocks used are those of Romer and Romer (2004), which have been updated by Wieland and Yang (2020). In contrast to other series of monetary policy shocks, this one covers the time period spanning back to January 1969 and up to January 2008. The extent to which they extend back in time is necessary for this analysis due to the gradually increasing size of firms' capacity buffers. The measure of capacity buffers, is the complement of the Federal Reserve's measure of total capacity utilization. Industrial production of business equipment is used as the measure for investment, and the measure of consumer prices that is most relevant for the industrial production sector is the goods subcategory of personal consumption expenditures. In addition, the quantity and price indices of personal consumption expenditures, and industrial production are used as broader measures of the economy. Quantity measures are in log levels, prices are in log differences, and the capacity buffer and the federal funds rate are in percent.

2.1 Results

The top row of Figure 3 provides the responses of investment, prices, and the federal fund rate when firms have large capacity buffers.[15] The point estimates of β are in black with 95 confidence intervals in green. In this regime the responses of investment and prices are negligible with neither of them deviating much from their initial levels. In contrast, the second row depicts the responses when firms have small capacity buffers. The responses of investment and prices are substantial and significant at the 5 percent level when firms operate with little excess capacity. Investment decreases by 0.07 log points after two and a half years, and prices immediately begin to decline at a gradual pace to a cumulative amount of 0.08 log points after three and a half years.[16] The shaded areas depict when there is a significant difference at the 5 percent level between the two regimes. There is a significantly differential response in the price level that begins after a year and in the level of investment in the third year after the shock.

The reaction of prices under small capacity buffers is larger than that, which would be implied under the smaller shift in demand that occurred under large capacity buffers. To see this, the change in the price level

---

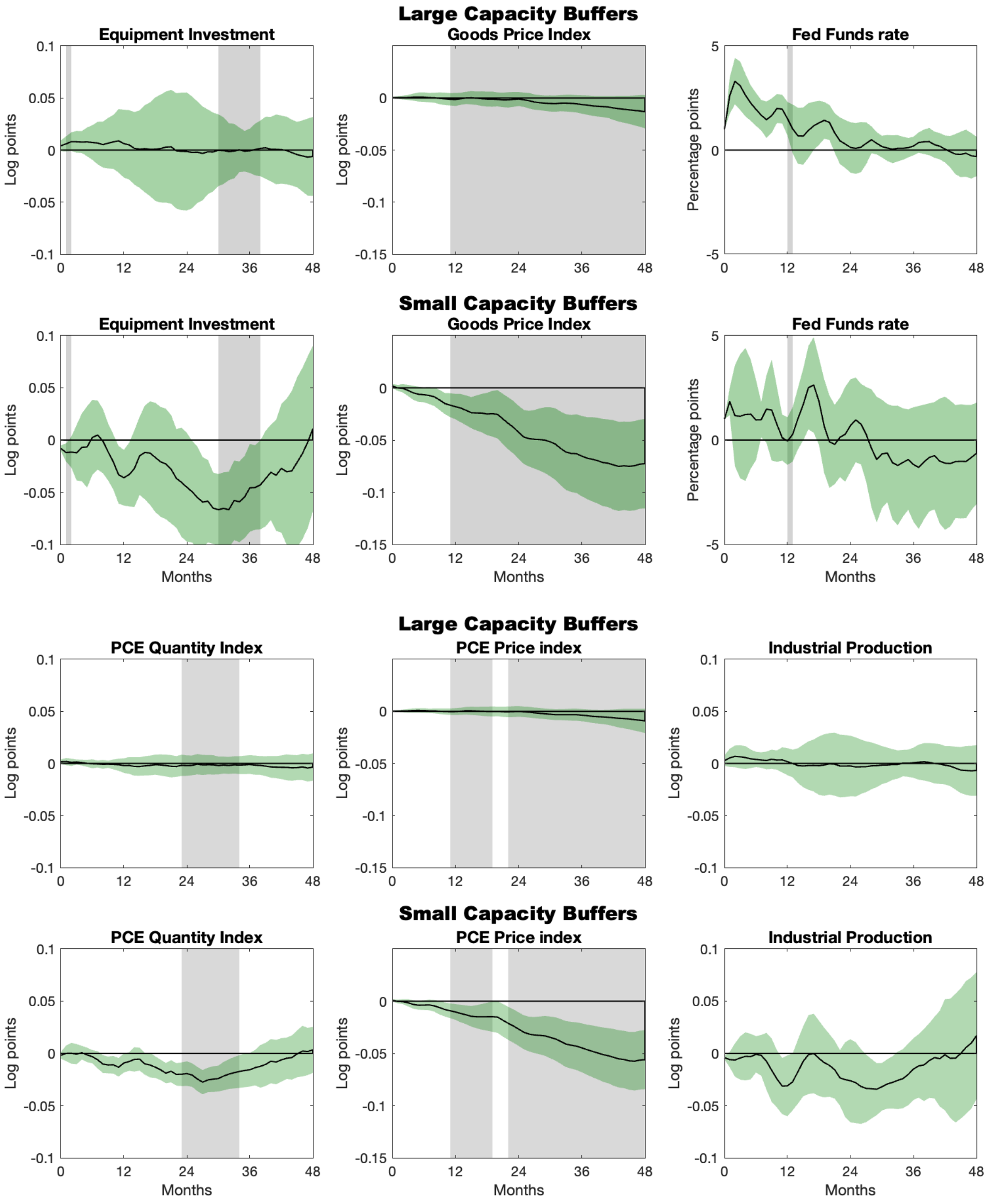
[13] Figure 1 in section 1 is also suggestive of this.

[14] Given the arbitrary degree of choice here, however, robustness exercises are done with the value of θ set such that the 99 percent of the transition occurs within a span of 2 and 8 percentage points of $\bar{B}$, as well as higher and lower values of $\bar{B}$ by 2 percentage points. The finding are similar to the baseline results across regimes that will be discussed next.

[15] The coefficients in (1) are estimated at horizons zero to $H = 48$ as a set of seemingly unrelated regression equations. Standard errors are computed following Driscoll and Kraay (1998) to account for the fact that residuals may be correlated across dates and estimation horizons. The allowed lag length for the autocorrelation in the Driscoll-Kraay standard errors is set to $H + 1$.

[16] That prices react immediately is a notable feature because their reaction is usually delayed by up to two years as shown in the regime-independent impulse responses provided in Figure 27 in Appendix D.

Figure 3: Estimated Capacity Buffer Dependent Impulse Responses



Notes: Estimated impulse responses to a contractionary one percent monetary policy shock in the logit smooth transition local projection model. Row one and three depicts the response in the regime where firms have large capacity buffers, and row two and four depicts the response in the regime where firms have small capacity buffers. Black lines indicate the point estimates of the impulse response while green areas show the 95 percent confidence intervals calculated from the Driscoll-Kray standard errors. Shaded time horizons indicate that the differential response between regimes was statistically significant.

Table 1: Relative response of consumption prices to quantities across horizons

Horizon (months)	12	18	26	30	36
P/C (Any B)	-0.04	-0.13	0.02	0.50	1.21
P/C ($B < 15\%$)	0.69	1.34	1.19	1.35	2.64

Notes: Ratios across horizons of the change in the price level to the change in quantities for personal consumption expenditures. The first row indicates the ratios for the unconditional estimated responses to monetary policy. The second row indicates the ratios for the estimated responses to monetary policy when capacity buffers are small (<15%).

of consumption relative to the change in the quantity of consumption is compared across regimes, whose responses are depicted in the first two columns of the bottom two rows in Figure 3. Since the response under large capacity buffers for consumption quantities and prices is negligible, the ratio of responses when capacity buffers are small is instead compared to those in the regime independent framework, which are provided in Figure 27 in Appendix D. The response of prices relative to that of quantities across horizons are given in Table 1 and are consistently more than twice as large when capacity buffers are small than when not taking the state dependence into account, indicating that prices are more sensitive to variation in demand when capacity buffers are small.[17]

This analysis does not prove that capacity buffers affect the sensitivity of prices and investment to demand shocks. The upward trend in capacity buffers as well as their procyclicality enable it to be a potential stand-in for other secular or cyclical mechanisms. These concerns are addressed at length in Appendix D. When the trend and cyclicality features of the capacity buffer series are isolated and used as the state variable, the responses are weakened or disappear. Similarly, when a recession based state variable is used instead, there is no differential response in investment, and the response of prices is weaker and delayed. The capacity buffer series is thus not a proxy for the business cycle, nor for longer term structural trends. It is the combination of the high and low frequency dynamics of the series that leads to the economically substantial and statistically significant differential results across the two regimes.[18][19]

To summarize, monetary policy surprises had little effect when capacity buffers were large, which constituted about half of the time period between 1969 and 2008. When capacity buffers were small, accounting for about 10 percent of the time period, the effects of monetary policy were not delayed and two to three times that of standard estimates in size. The remaining 40 percent of the time period, the effects were intermediate in size.[20]

---

[17] The analogous ratios for the response of investment relative to output across horizons are given in Table 6 in Appendix D. They similarly indicate that investment is more sensitive to variation in demand when capacity buffers are small.

[18] Another concern is that the monetary policy shocks may be systematically different when capacity buffers are small than when large. This is also addressed in Appendix D and shown not to be the case.

[19] Recently, Gonçalves et al. (2023) point out that the local projection estimates in a state-dependent framework obtains the true population impulse responses only when shocks are small. For larger shocks the estimates do not fully account for the indirect effect that comes from the shock affecting the future states. This bias is unlikely to be a concern for this analysis since monetary shocks tend to be small, especially in contrast to the fiscal shocks which is their focus. Even in their own analysis when it comes to estimating cumulative fiscal multipliers, where the bias compounds, the cumulative bias is small except for in the most extreme cases.

[20] These results are consistent with other recent empirical work that finds capacity utilization as an important state variable. Fazzari et al. (2015) analyze the impact of fiscal policy in a Bayesian Threshold VAR. They conducted a horse-race amongst many state variables including the output gap, unemployment rate, and GDP growth and found that the capacity utilization model had the highest marginal likelihood. Boehm and Pandalai-Nayar (2022) use industry level data to estimate the convexity of supply curves. They find

3 A Theory of Capacity Buffers in Production

This section lays out the objective of the firm, the constraints it faces, and the resulting decisions for investment and pricing that they imply. Time is discrete and perpetual, and the economy is occupied by a unit mass of firms. The objective of firm $i \in [0,1]$ in period t is to maximize its net present value, as measured by the discounted sum of its future profits, given in recursive form by

$$V_{i,t} = \max_{y_{i,t}, l_{i,t}, I_{i,t}, k_{i,t+1}, x_{i,t+1}, p_{i,t+1}} p_{i,t} y_{i,t} - W_t l_{i,t} - P_t^I I_{i,t} - \frac{\chi_p}{2}\left(\frac{p_{i,t+1}}{p_{i,t}} - 1 - \pi_t\right)^2 P_t Y_t + \frac{1}{1+r_t}\mathbb{E}_t V_{i,t+1} \tag{3}$$

where $V_{i,t} = V_t(I_{i,t-1}, p_{i,t}, k_{i,t}, x_{i,t})$. It chooses labor $l_{i,t}$ for use in production of output $y_{i,t}$ in period t, and investment $I_{i,t}$ to accommodate its choices of capital stock, $k_{i,t+1}$, and capital intensity, $x_{i,t+1}$, for the next period. It also chooses at time t its price for period: $p_{i,t+1}$. W_t is the wage rate, P_t^I is the price of investment goods, and r_t is the nominal interest rate, all of which the firm takes as given. χ_p controls the strength of the Rotemberg price adjustment cost, which are relative to the measured aggregate inflation rate, π_t. The cost is in terms of and scaled according to aggregate output, Y_t.

The Putty-Clay Production Technology The firm produces output according to the following technology

$$y_{i,t} = \min\{a^l_{i,t} l_{i,t}, a^k_{i,t} k_{i,t}\} \quad , \quad a^l_{i,t} = F(x_{i,t}, 1) \quad , \quad a^k_{i,t} = F(1, x^{-1}_{i,t}) \tag{4}$$

where $F(\cdot,\cdot)$ is constant returns to scale and characterizes the long run substitutability between capital and labor. In any given period, a firm produces according to a Leontief production function, which is modified by the firm's previous choices of capital stock $k_{i,t}$ and capital intensity $x_{i,t}$. The capital intensity is the fixed proportion in which capital and labor are combined to produce output.[21] It describes how much capital is allocated per unit of labor and thereby governs the productivity of the labor and capital stock. Higher levels of capital intensity mean that more capital is allocated to each unit of labor, making labor more productive. However, given a fixed stock of capital, a higher capital intensity reduces the production capacity,

$$\bar{y}_{i,t} = a^k_{i,t} k_{i,t}, \tag{5}$$

since allocating more capital to each unit of labor reduces the total amount of labor the capital stock can support. While the choice of labor is variable within a period, the fact that the capital stock and intensity are predetermined within a period imposes a limit on production, which is the firm's production capacity.[22]

---

robust evidence that industry supply curves are convex with its steepness rising in the industry's average capacity utilization rate.

[21] This can be seen from the fact that $a^l_{i,t}/a^k_{i,t} = x_{i,t}$.

[22] The motivation behind the production technology is as follows: When a firm purchased its capital, it took on a specific form which can be described in terms of its capital intensity that is how much capital is allocated to each worker. By choosing capital stock that features a higher intensity, the firm makes its employees more productive, but reduces the total amount of employees it can use effectively. Suppose a firm bought five expensive laptops rather than ten cheap ones for the same total cost. Then the amount of capital used per worker, i.e. the capital intensity, is higher because each worker is able to do more work on a better laptop, but they are only able to employ five workers, one for each laptop. If the firm hires less than five workers, some of the laptops will go unused

The technology in (4) is a generalization of the standard constant returns to scale production function in capital and labor. If the firm could flexibly adjust its capital intensity, it would reduce to a standard CES production function in capital and labor. Under no uncertainty, the firm would optimally choose to produce at $a^l_{i,t} l_{i,t} = a^k_{i,t} k_{i,t}$ where it does not waste any resources, which occurs when $x_{i,t} = k_{i,t}/l_{i,t}$. Substituting this level of capital intensity into the production technology above yields

$$y_{i,t} = \min\{F(k_{i,t}/l_{i,t}, 1)l_{i,t}, F(1, l_{i,t}/k_{i,t})k_{i,t}\} = F(k_{i,t}, l_{i,t}).$$

For the remainder of this paper $F(\cdot,\cdot)$ will be assumed to be Cobb-Douglas with parameter α capturing the elasticity of labor productivity $a^l_{i,t}$, with respect to capital intensity $x_{i,t}$. It also governs the steady state capital share after netting out profits per usual, so

$$a^l_{i,t} = x^{\alpha}_{i,t} \quad \text{and} \quad a^k_{i,t} = x^{-(1-\alpha)}_{i,t}. \tag{6}$$

This implies that elasticity of substitution between capital and labor is one over the long run, while it is zero within a period.

Over an intermediate period of time, the elasticity of substitution between capital and labor will take on an intermediate value that is monotonically increasing in the length of the time period that is considered. This is the result of the costs to the capital stock that are incurred from changing the level of capital intensity. The adjustment costs to capital intensity are featured in the law of motion for capital,

$$k_{i,t+1} = \left(1 - \delta_k - \frac{\chi_x}{2}\left(\frac{x_{i,t+1} - x_{i,t}}{x_{i,t}}\right)^2\right) k_{i,t} + \left(1 - \frac{\chi_I}{2}\left(\frac{I_{i,t} - I_{i,t-1}}{I_{i,t-1}}\right)^2\right) I_{i,t}. \tag{7}$$

Adjusting a firm's capital intensity is arguably a type of investment. By changing it, a firm is upgrading its old capital stock so that it features a different capital intensity. This makes the capital stock decline beyond the depreciation rate δ_k as it will in part require replacing some old capital with new. The overall stock of capital is then chosen by the firm's investment choice subject to standard investment adjustment costs of Christiano et al. (2005). The parameters χ_x and χ_I govern the strength of the adjustment costs to capital intensity and to investment, respectively.

Uncertainty in Demand The firm competes monopolistically facing a demand that is derived from the following CES preferences across products,

$$Y_t = \left(\int_0^1 \nu_{i,t}^{\frac{1}{\varepsilon_p}} y_{i,t}^{\frac{\varepsilon_p - 1}{\varepsilon_p}} di\right)^{\frac{\varepsilon_p}{\varepsilon_p - 1}}. \tag{8}$$

---

as a person does not benefit from using two at once. If the firm hired more than five employees, some employees would then stand by idly without a laptop.

Y_t denotes the aggregate demand, ε_p is the own-price elasticity of demand, and $\nu_{i,t} \in [0, \infty)$ are idiosyncratic taste shocks that are distributed i.i.d. across time periods according to the CDF $G(\nu)$ with mean one and variance σ_ν^2. The distribution of idiosyncratic taste shocks is assumed to be log-Normal.

Since the firm's price is preset within a period, it produces to satisfy its demand up to its production capacity in (5). The choice of labor is therefore demand determined and given by $l_{i,t} = y_{i,t}/a_{i,t}^l$. It is impossible to purchase more of a good than the firm is able to produce, even though it may be desired, which results in the following quantity constraints from the demand perspective,[23]

$$y_{i,t} \leq \bar{y}_{i,t} \tag{9}$$

The demand curve for a given firm's good is then derived from the expenditure minimization problem associated with (8) and the quantity constraints (9). This results in the downward sloping effective demand curve,

$$y_{i,t} = \nu_{i,t} Y_t \left(\frac{p_{i,t} + \varphi_{i,t}}{P_t} \right)^{-\varepsilon_p}, \tag{10}$$

with lagrange multipliers $\varphi_{i,t}$ and P_t of the quantity constraints and the CES preferences, respectively. P_t is the shadow price that describes the marginal cost for the bundle of goods demanded, which depends on the degree to which the quantity constraints bind as indicated by $\varphi_{i,t}$. It can be characterized by combining (8) and (10), to get

$$P_t = \left(\int_0^1 \nu_{i,t} (p_{i,t} + \varphi_{i,t})^{1-\varepsilon_p} di \right)^{\frac{1}{1-\varepsilon_p}}. \tag{11}$$

Whenever the firm is capacity constrained $\varphi_{i,t} = P_t \left(\frac{\nu_{i,t} Y_t}{\bar{y}_{i,t}} \right)^{\frac{1}{\varepsilon_p}} - p_{i,t}$, and zero otherwise. We can then rewrite (10) as

$$y_{i,t} = \min \left\{ \nu_{i,t} Y_t \left(\frac{p_{i,t}}{P_t} \right)^{-\varepsilon_p}, \bar{y}_{i,t} \right\}. \tag{12}$$

Capacity Buffers The firm's behavior will depend on whether it is capacity constrained. However, since some of the firm's decisions must be made a period in advance, the decisions are made before the idiosyncratic demand is realized at the beginning of the next period. Given the demand uncertainty, firms will choose to operate with capacity buffers to avoid becoming capacity constrained in periods of high demand. The size of their capacity buffer also influences their behavior as it affects how likely it is that they will become capacity constrained. Given the known distributional shape of the idiosyncratic demand, $G(\nu)$, this probability is captured by the threshold at which a firm becomes capacity constrained. Define this idiosyncratic demand threshold by

$$\bar{\nu}_{i,t} \equiv \frac{\bar{y}_{i,t}}{Y_t} \left(\frac{p_{i,t}}{P_t} \right)^{\varepsilon_p}. \tag{13}$$

---

[23] When demand is greater than capacity, rationing takes place proportional to demand in order to remove any allocative distortions that could have risen from it.

When a firm faces a demand draw $\nu_{i,t} > \bar{\nu}_{i,t}$ it is capacity constrained and will produce at capacity. Conversely, when $\nu_{i,t} \leq \bar{\nu}_{i,t}$, the firm is not and will produce according to demand. The expectation with respect to idiosyncratic component of demand regarding the firm's production and sales in (12) can be written as

$$\mathbb{E}_\nu y_{i,t} = \int_0^{\bar{\nu}_{i,t}} \nu Y_t \left(\frac{p_{i,t}}{P_t}\right)^{-\varepsilon_p} dG(\nu) + \int_{\bar{\nu}_{i,t}}^{\infty} \bar{y}_{i,t} dG(\nu). \tag{14}$$

The idiosyncratic demand threshold impacts the firm's decisions, but it is not observable. It is therefore convenient to characterize the firm's capacity buffer because it features a bijective mapping with the idiosyncratic demand shock threshold. Intuitively, the expected size of the capacity buffer captures the same information as the demand threshold. From it the firm can deduce how large a demand shock it must face before it becomes capacity constrained. Combining (12) and (13), the firm's capacity buffer is defined as the amount of excess capacity relative to the firm's total production capacity,

$$b_{i,t} \equiv \frac{\bar{y}_{i,t} - y_{i,t}}{\bar{y}_{i,t}} = \max\left\{\frac{\bar{\nu}_{i,t} - \nu_{i,t}}{\bar{\nu}_{i,t}}, 0\right\}, \tag{15}$$

with the firm's chosen capacity buffer being defined by its size in expectation,

$$B_{i,t} \equiv \mathbb{E}_{t-1} b_{i,t} = \int_0^{\bar{\nu}_{i,t}} \frac{\bar{\nu}_{i,t} - \nu}{\bar{\nu}_{i,t}} dG(\nu). \tag{16}$$

The expected capacity buffer depends only on the demand threshold $\bar{\nu}_{i,t}$. Because they are strictly monotonic in one another, it can be inverted (16) to get the demand threshold as a function of the capacity buffer size, $\bar{\nu}(B_{i,t})$. The decisions of the firm, which we will cover next, can thus be recast as a function the capacity buffer size, rather than the unobserved idiosyncratic demand threshold.[24]

Symmetric Equilibrium Because the price and capital decisions are made before the realization of the idiosyncratic uncertainty one period later, the theory admits a symmetric equilibrium where all firms have and choose the same price, capital intensity, and capital stock. All firms will have the same demand threshold given by (13), which means that every firm chooses the same size for its capacity buffer in expectation. The relative firm price to the shadow price for the aggregate bundle of goods from (11) is solely a function of the demand threshold,

$$\frac{p_t}{P_t} = \bar{\nu}_t^{\frac{1}{\varepsilon_p - 1}} \left(\int_0^{\bar{\nu}_t} \frac{\nu}{\bar{\nu}_t} dG(\nu) + \int_{\bar{\nu}_t}^{\infty} \left(\frac{\nu}{\bar{\nu}_t}\right)^{\frac{1}{\varepsilon_p}} dG(\nu)\right)^{\frac{1}{\varepsilon_p - 1}}. \tag{17}$$

---

[24] The previous work that this theory builds upon, starting with Fagnart et al. (1997), describe this mapping in terms of the firm's capacity utilization rate, which is the complement to the size of its capacity buffer. Casting it in terms of the capacity buffer, instead, is done for a variety reasons. It emphasises that its size, which is the crucial determinant for the dynamics of prices and investment, is a choice variable of the firm, rather than an outcome of pure circumstance. How the size of it changes over the short, medium, and long run are all at the center of the results in this paper, and these dynamics are best understood from the firm's interest for the capacity buffer. As such, this framing purveys better economic intuition and clarity. It also has the benefit of further distancing the theory from the well known theory of capacity utilization of Greenwood et al. (1988), from which it is conceptually quite different. The capacity buffer feature that arises due to an engineering concept of capacity has no counterpart in that theory.

Despite the per period variation in production, labor demand, and profits across firms, the production side admits a representative firm. The representative firm's sales is equal to the expected amount of sales of any one firm before the idiosyncratic demand is realized in (14). The capacity buffer of the representative firm is therefore also equal in size to the expected size of the capacity buffer of any one firm in (16).

3.1 Optimal Pricing and Investment

Using the expressions derived above, it is convenient to reformulate the firm problem as

$$V_{i,t} = \max_{I_{i,t}, k_{i,t+1}, x_{i,t+1}, p_{i,t+1}} \left(p_{i,t} - \frac{W_t}{a^l_{i,t}} \right) y_{i,t} - P^I_t I_{i,t} - \frac{\chi_p}{2} \left(\frac{p_{i,t+1}}{p_{i,t}} - 1 - \pi_t \right)^2 P_t Y_t + \frac{1}{1+r_t} \mathbb{E}_t V_{i,t+1} \tag{18}$$

subject to (6), (7), (13), and (14). The firm's optimality conditions associated with (18) are summarized in turn.

Price The optimal pricing condition comes from combining the first order condition for $p_{i,t+1}$ with the envelope condition for $p_{i,t}$. This yields the following Phillips curve:

$$(\pi_{i,t+1} - \pi_{i,t})(1 + \pi_{i,t+1}) = \mathbb{E}_t \kappa^p_{i,t} \left(\frac{\varepsilon_{i,t+1}}{\varepsilon_{i,t+1} - 1} \frac{W_{t+1}}{a^l_{i,t+1} p_{i,t+1}} - 1 \right) + \mathbb{E}_t \beta^p_t (\pi_{i,t+2} - \pi_{i,t+1})(1 + \pi_{i,t+2}) \tag{19}$$

where the net inflation rates are denoted by $\pi_{i,t} = p_{i,t}/p_{i,t-1} - 1$ and $\pi^P_t = P_t/P_{t-1} - 1$, and $\beta^p_t = \frac{1+\pi^P_{t+1}}{1+r_t} \frac{Y_{t+1}}{Y_t}$ and $\kappa^p_{i,t} = \beta^p_t \frac{p_{i,t+1} y^{\mathbb{E}}_{i,t+1}}{P_{t+1} Y_{t+1}} \frac{\varepsilon_{i,t+1} - 1}{\chi_p}$. The price elasticity of demand ε_p is not the relevant price elasticity because the sensitivity of actual sales to a change in prices is not dependent on it when the firm is at capacity. When the firm is capacity constrained, the price elasticity of its sales is zero. A higher price would not lower their sales because there is excess demand relative to the constrained supply, so raising the price would not result in a decline in sales. The relevant price elasticity is the effective price elasticity of sales, $\varepsilon_{i,t}$, which is a weighted average of the price elasticity of demand (the price elasticity when not at capacity) and zero (the price elasticity when at capacity),

$$\varepsilon_{i,t} = \eta(\bar{\nu}_{i,t}) \varepsilon_p \quad \text{where} \quad \eta(\bar{\nu}_{i,t}) = \frac{\int_0^{\bar{\nu}_{i,t}} \frac{\nu}{\bar{\nu}_{i,t}} dG(\nu)}{\int_0^{\bar{\nu}_{i,t}} \frac{\nu}{\bar{\nu}_{i,t}} dG(\nu) + \int_{\bar{\nu}_{i,t}}^{\infty} dG(\nu)}. \tag{20}$$

The appropriate weight is the sales weighted probability of having excess capacity, $\eta(\bar{\nu}_{i,t})$, which is fully determined by the demand threshold $\bar{\nu}_{i,t}$ and can in turn be characterized by the capacity buffer using (16). Let the desired unit labor cost markup be defined as

$$\mu_{i,t} \equiv \frac{\varepsilon_{i,t}}{\varepsilon_{i,t} - 1}. \tag{21}$$

Capital Stock The first order condition for $k_{i,t+1}$ and envelope condition for $k_{i,t}$ combine to tell us that the marginal value of capital, given by the Lagrange multiplier $q_{i,t}$ on the capital law of motion in (7), is equal to the discounted value of the increase in revenue net labor costs that the capital would provide from expanding the firm's capacity if the firm were to become capacity constrained next period (in addition to the remaining capital after depreciation and replacement),

$$\frac{q_{i,t}}{P_t} = \mathbb{E}_t \frac{1+\pi^P_{t+1}}{1+r_t} \left(a^k_{i,t+1} \left(\frac{p_{i,t+1}}{P_{t+1}} - \frac{W_{t+1}}{a^l_{i,t+1} P_{t+1}} \right) \int_{\bar{\nu}_{i,t+1}}^{\infty} dG(\nu) + \left(1 - \delta_k - \frac{\chi_x}{2} {g^x_{i,t+1}}^2\right) \frac{q_{i,t+1}}{P_{t+1}} \right). \tag{22}$$

If the firm draws a idiosyncratic demand shock $\nu_{i,t+1} < \bar{\nu}_{i,t+1}$, then the firm will have excess capacity and the additional unit of capital will provide no direct benefit in the next period. Conversely, if it draws $\nu_{i,t+1} > \bar{\nu}_{i,t+1}$ which makes the firm capacity constrained, then the additional capital provides the benefit of increasing the firm's sales, and thus profits, by the amount $a^k_{i,t}$ that the additional capital expands the firm's capacity.

Capital Intensity Let $g^x_{i,t}$ denote the growth rate of capital intensity $x_{i,t}$. The first order condition for $x_{i,t+1}$ along with envelope conditions for $x_{i,t}$ combine into the following optimality condition,

$$g^x_{i,t+1}(1+g^x_{i,t+1}) = \kappa^x_{i,t+1} \left(\alpha - (1-\alpha)\left(1 - \eta(\bar{\nu}_{i,t+1})\right) \left(\frac{p_{i,t+1}}{W_{t+1}/a^l_{i,t+1}} - 1 \right) \right) + \mathbb{E}_t \beta^x_{i,t} g^x_{i,t+2}(1+g^x_{i,t+2}), \tag{23}$$

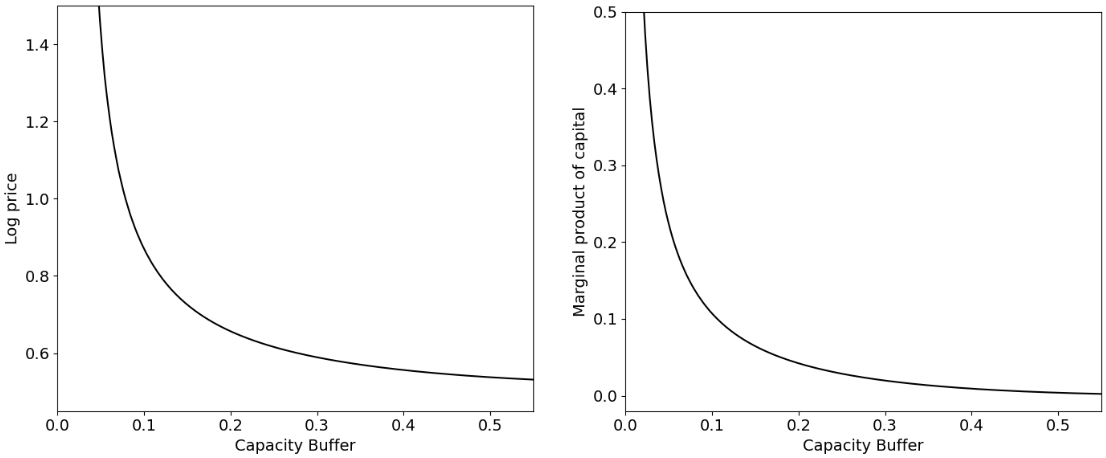
where $\beta^x_{i,t} = \frac{1+\pi^q_{i,t+1}}{1+r_t}\frac{k_{i,t+1}}{k_{i,t}}$ and $\kappa^x_{i,t+1} = \mathbb{E}_t \beta^x_t \frac{W_{t+1} l_{i,t+1}}{q_{i,t+1} k_{i,t+1}} \frac{1}{\chi_x}$.The growth in capital intensity is driven by two forces which pull it in opposite directions. Increasing the capital intensity has the benefit of making workers more productive and reducing production costs by increasing $a^l_{i,t}$. However, increasing $x_{i,t}$ also bears the costs of decreasing the firms overall capacity as it lower $a^k_{i,t}$. While the production costs decreases hold in all states of the world, the negative impact from a decrease in capacity only has an effect in states of the world where the firm becomes capacity constrained and could have benefited from an increase in its capacity.

Investment The first order condition for $I_{i,t}$ and envelope condition for $I_{i,t-1}$ yield the investment condition. Here, it is written in analogous form to the conditions above, where $g^I_{i,t}$ is the growth rate of investment $I_{i,t}$,

$$g^I_{i,t}(1+g^I_{i,t}) + \frac{1}{2}{g^I_{i,t}}^2 = \frac{1}{\chi_I}\left(1 - \frac{P_{I,t}}{q_{i,t}}\right) + \mathbb{E}_t \frac{1+\pi^q_{i,t+1}}{1+r_t} \frac{I_{i,t+1}}{I_{i,t}} g^I_{i,t+1}(1+g^I_{i,t+1}) \tag{24}$$

The growth rate of investment depends on the value of marginal value of capital $q_{i,t}$ relative to the cost of purchasing it P^I_t.

Figure 4: Capacity Buffer Dependent Pricing and Investment Incentives



Notes: The curves trace out the functions given in (25). Calibration is based on the one estimated in section 4. The distribution of idiosyncratic demand, $G(\nu)$ is assumed to be Log-Normal. The capacity buffer is varied via different levels of aggregate demand Y and is defined as its proportion of total capacity.

3.2 Pricing and Investment Incentives in the Short Run

The optimality conditions characterize how the firm's capacity buffer influences its pricing and investment decisions. To gain a deeper intuition, Figure 4 sketches out in a static environment the desired price and associated marginal product of capital depending on the size of its capacity buffer.[25] The curves are traced out ignoring the price adjustment costs in order to understand what price the firm would ideally like to set. The firm's production technology is held constant (i.e. the capital stock and capital intensity which together determine the labor productivity and production capacity). The reason for doing this is that changing a firm's underlying production technology is a costly and slow process. One can therefore gain an understanding for what the firm's incentives are to temporary disturbances by holding them constant.[26] This results in the following functional forms

$$p_i = \frac{\varepsilon_i}{\varepsilon_i - 1}\frac{W}{a_i^l} \qquad \text{and} \qquad mpk_i = a_i^k\left(p_i - \frac{W}{a_i^l}\right)\int_{\bar{\nu}_i}^{\infty} dG(\nu), \tag{25}$$

with the price elasticity of sales and its relation to the capacity buffer size described in (16) and (20).

Given a fixed wage and labor productivity, the marginal cost of the firm is constant. The desired price level is then only affected via changes in the desired markup over labor costs, which depends on the size of the capacity buffer. With the capital stock and intensity are fixed, the size of the capacity buffer is varied by changing the level of aggregate demand Y via (13). A smaller capacity buffer increases the probability

---

[25] The growth rate of investment is determined by the marginal value of capital, $q_{i,t}$, in (24), which in turn is the discounted future sum of its marginal products given by (22). Inspecting how the marginal product of capital depends on the size of the firm's capacity buffer gives insight into the incentive to invest depending on the size of the capacity buffer.

[26] One could alternatively make some assumption about partial adjustment of the variables, which would similarly contain a degree of arbitrariness. In section 5 an exact analysis is done by inspecting the firm's actual response to temporary shocks in a dynamic environment.

of ending up at capacity from the idiosyncratic demand draw ν. This reduces the price elasticity of sales in (20) since the high probability of being capacity constrained implies that there is a high probability the firm can set a higher price without it reducing its sales.

As can be seen in (25), the marginal product of capital is similarly affected by the expected size of the capacity buffer. An additional unit of capital is only beneficial in states of the world that would otherwise have resulted in the firm becoming capacity constrained, which is more likely to occur for under a smaller buffer. The foregone profits from not having the additional capital will also be larger given the higher desired price associated with the capacity constrained states of the world.[27] Conversely, under a large capacity buffer, even a high idiosyncratic demand draw may not be sufficient to hit capacity. In this case, with a low probability of being capacity constrained, the price elasticity of sales will be higher and close to the price elasticity of demand, resulting in a smaller desired markup. The marginal product of capital will also be low since the additional unit of capital is only beneficial if the firm would otherwise have become capacity constrained, which in this case is very unlikely to occur.

Figure 4 also alludes to the fact that the sensitivity of the desired price level and investment are also much greater when the firm has a small capacity buffer. This is because a rise in the level of aggregate demand yields a nearly proportional decline in the size of the capacity buffer. The slopes in the figures are therefore indicative of the sensitivity of prices and the marginal product of capital with respect to changes in aggregate demand. When the capacity buffer is small, even small changes in its size greatly affect the probability of becoming capacity constrained, and thereby the desired price level and the marginal product of capital. When capacity buffers are large, changes in its size have a negligible effect on the probability of becoming capacity constrained, so it has little effect on the firm's investment and pricing decisions.

4 Causes and Consequences of Rising Capacity Buffers

This section provides evidence that capacity buffers rose due to a rise in markups and that the rise in capacity buffers in turn led to an increase in the idiosyncratic volatility of firm sales and a flattening of its supply curve. The rise in markups is modelled as coming from an exogenous decline in the price elasticity of demand. All else equal, under a lower price elasticity of demand, firms set a higher markup and achieve larger profits. This in turn raises the value of having excess capacity. When a firm's profits per unit output are higher, there are larger foregone profits from not having a higher capacity level when demand is greater than capacity. The firm therefore chooses to have a larger capacity buffer to gain the now enhanced profits in the states of the world where it does become capacity constrained.[28]

---

[27] Moreover, if all firms have a smaller buffer, then there is more demand that will spill over from the ones that end up being capacity constrained onto the unconstrained ones.

[28] The degree to which markups have risen has become a hotly debated topic after De Loecker et al. (2020) estimated a rise in markups of 40 percentage points and found much of it to be due to a rise in market power. Foster et al. (2022) conducted a parallel more granular analysis and found that the fall in variable costs shares is more likely due to a change in technology with larger fixed costs of capital, although, depending on the methodology, there remained some evidence for rising market power as well. Importantly, the mechanism in this paper works through the lower variable costs and therefore higher marginal profits, with the mechanism inducing this change

In the steady state the expected marginal product of capital is equal to its user cost,

$$\frac{1+r}{1+\pi^P} - (1-\delta) = \frac{\mu - 1}{\mu}\frac{p}{P}\mathbb{P}\left[b = 0\right] a^k, \tag{26}$$

where μ is the gross markup over labor costs as function of the price elasticity in (21). An exogenous fall in ε_p leads to a rise in the markup μ, which in turn raises the expected marginal product of capital. The user cost of capital has not changed, however, and the firm will expand its capacity buffer, which reduces the probability of being capacity constrained, $\mathbb{P}\left[b = 0\right]$ (and slightly lowers the higher markup), until the marginal product of capital is again balanced with its user cost.

The firm primarily raises its capacity buffer by reducing capital intensity, making the capital it has more productive. Raising its capacity via its capital stock would inefficiently increase the cost share of capital relative to that of labor. In the steady state the firm will have efficiently allocated the costs of production across labor and capital according to the Cobb-Douglas technological frontier. The optimal choice of capital intensity in (23) in steady state necessitates that the relative production cost shares of capital and labor be balanced by the labor productivity elasticity of capital intensity α,

$$(1-\alpha)Rk = \alpha W \mathbb{E}_\nu[l], \tag{27}$$

where R is equal to the marginal product of capital given in real terms in (26). When the rise in markups raises the marginal product of capital, the firm will switch to a less capital intensive production process, spreading out its capital stock among more workers, thereby increasing capital productivity and capacity. The labor productivity a^l falls as a result, but this is compensated for by the lower real wage due to the higher markup. The rise in markups did not lead to a boom in investment. Rather it actually induced firms to employ less capital intensive production techniques reducing labor productivity.[29]

4.1 Labor Shares Fell when Capacity Buffers Rose

One indication as to whether this mechanism has occurred is whether rising markups are correlated with rising capacity buffers in the data. Markups are difficult to measure, however, with one of the difficulties in measuring the rise in markups being that the income accruing to capital versus profits is difficult to disentangle. Measures of the labor share, however, can be more easily separated in the data by counting up

---

not being directly relevant. The rise in markups is modelled as an exogenous fall in the price elasticity of demand because it is the cleanest way to do it without adding additional complexity to the theory.

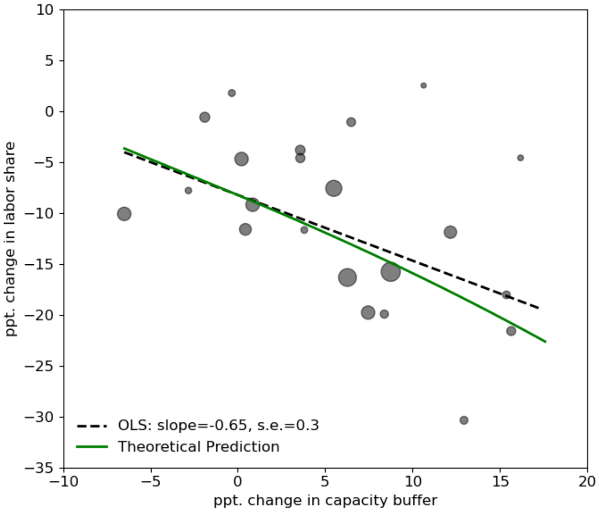
[29] A gradual decline in markups in this framework therefore also implies lower labor productivity growth and a decline in investment, akin to the findings of Crouzet and Eberly (2020, 2021).

all the wages and benefits that go to workers.[30] The income shares in the model's steady state are given by

$$\int_0^1 p_i y_i di = \left(\underbrace{\frac{\mu-1}{\mu}\left(\int_0^{\bar{\nu}} \frac{\nu}{\bar{\nu}}\, dG(\nu)\right)}_{\text{profits}} + \underbrace{\frac{\mu-1}{\mu}\left(\int_{\bar{\nu}}^{\infty} dG(\nu)\right)}_{\text{capital}} + \underbrace{\frac{1}{\mu}\left(\int_0^{\bar{\nu}} \frac{\nu}{\bar{\nu}}\, dG(\nu) + \int_{\bar{\nu}}^{\infty} dG(\nu)\right)}_{\text{labor}} \right) p\bar{y}. \tag{28}$$

If there is an exogenous fall in the price elasticity of demand ε_p and thereby a rise in the markup μ as described by (20) and (21), the labor share will fall. Simplifying (28) also shows that the labor share of income is equal to one over the labor cost markup in the steady state, $S_L = \frac{1}{\mu}$. A rise in markups should induce both a decline in the labor share and a rise in the capacity buffer of a firm.

Figure 5: Capacity Buffers and Labor Shares



Notes: The panel uses the labor shares from the Bureau of Labor Statistics and capacity buffer measures from the Federal Reserve for the three digit NAICS industries underpinning the industrial production sector. Each point depicts the percentage point change in these measures between 1987 and 2021 with a 4 year weighted average used at the start and end of each industry time series. The marker size is proportional to the average value added size of the industry between 1987 and 2021. The dashed line is the associated weighted least squares regression with the HC3 standard errors reported due to small sample size.

There is evidence of this association between markups and capacity buffers in the cross-section of industries. Figure 5 plots the percentage point change in the capacity buffer sizes and labor shares in the cross-section for the 3-digit NAICS industries underlying industrial production. The percentage point changes in the labor share and the capacity buffer for an industry between 1987 and 2021 are given by the vertical and horizontal axes, respectively. Since their values vary substantially at high frequencies the difference is taken between the average of the first four and last four years of the panel for each industry. Each industry is weighted by their average value-added share across the time period. The scatter plot shows that industries which experienced larger declines in their capacity buffers also experienced larger declines in their

---

[30] It is still not a straightforward task to measure as assumptions must be made regarding how proprietors income is split between labor and non-labor income.

labor shares. A linear regression indicates that for every one percentage point decline in the capacity buffer, the labor share declined by about half a percentage point. The estimate is statistically significant at the five percent level.[31]

Figure 5 also depicts the theoretically predicted relationship between the capacity buffer and labor share based on the quantitative calibration of the model that will be performed next. The curve describes the implied changes in the labor share and the capacity buffer that are associated with one another due to changes in the size of the markup from their initial level in the year 1967. Its intercept is then realigned with the intercept in the empirically estimated change in the labor share of the fitted regression line. The theoretically predicted relationship lines up closely with, but is slightly steeper than, the line of best fit while featuring a slight concavity that is qualitatively in line with the appearance of the scatter plot.[32]

4.2 Implications of Rising Capacity Buffers

How a rise in markups leads to a rise in capacity buffers is visualized in the first panel of Figure 6, which again maps out the static price and marginal product of capital curves when the production technology is held fixed and prices are flexible. The exogenous rise in markups shifts upwards the marginal product of capital curve since the value of capital is higher for all sizes of the firms capacity buffer. Given the same user cost of capital, the firms increase the size of their capacity buffers until the marginal product of the capital is again equal to its user cost. The rising capacity buffers has two subsequent implications. One is that its supply curve becomes locally flatter as the firm is operating further from its capacity constraint. The other is that since firms are operating further from their capacity constraints, more of the idiosyncratic variation in demand passes through to variation in actual sales. The cross-sectional variance of sales should therefore have increased.

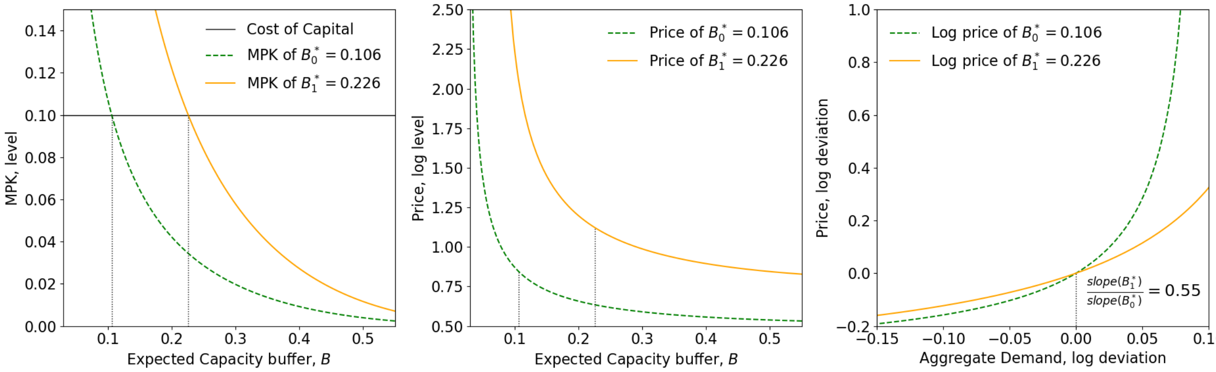
The Flattening of the Supply Curve Define the firm's static short run supply curve as the schedule of prices the firm would flexibly set in response to different realizations of aggregate demand, conditional on its current production technology (its capital stock and intensity). An increase in a firm's capacity buffer leads to a flattening of its short run supply curve. The exogenous rise in markups does shift up and steepen the price curve for all sizes of the firm's capacity buffer as depicted in the middle panel of Figure 6, but this overall steepening is more than outweighed by the local flattening that occurs due to firms choosing to operate with larger capacity buffers. The combined effect is shown in the right panel of Figure 6, which plots the short run supply curve in terms of the log deviation in aggregate output from the its steady state before and after the rise in markups induced a change in the capacity buffer size from $B_0 = 0.106$ to $B_1 = 0.226$.

---

[31] The results are robust to the number of years averaged over at the beginning and end of sample with the coefficient remaining statistically significant and maintaining a slope coefficient smaller than -0.5 up to when the end-point averages are taken over ten years, i.e. when the majority of the sample is used for the averages. The analysis is robust to the choice of weights (e.g. start of sample or no weighting) up to when the end point averages are taken over more than five years.

[32] Appendix B briefly discusses some alternative mechanisms that would also result in a rise in capacity buffer sizes along with their shortcomings. None of them predict a correlation between the change in capacity buffers and labor shares.

Figure 6: The Flattening Phillips Curve Mechanism



Notes: The curves trace out the value of the marginal product of capital and the desired price level (normalized to one in both steady states) for various levels of aggregate demand when the capital stock and capital intensity are fixed at their steady state.

As a result the slope as measured by the log deviation in the desired price levels, declined by 45 percent. Moreover, a larger positive deviation in output is possible without inducing a large increase in the price level under the larger capacity buffer.

The flattening of the supply curve can be analytically characterized by the change in the slope of log desired price level against the log deviation in aggregate output. When the production technology is fixed, variation in aggregate output maps directly into variation in the expected size of the capacity buffer. The desired response in the price depends on the initial capacity buffer size as described earlier. The slope is given by

$$\zeta(\bar{v}) \equiv \frac{\partial \ln p}{\partial \ln Y} = \frac{1}{1-\Omega(\bar{v})} \frac{1}{\varepsilon(\bar{v})-1} \left(\frac{\bar{v} g(\bar{v})}{\int_0^{\bar{v}} \frac{v}{\bar{v}} dG(v)} - (1-\eta(\bar{v})) \right) \tag{29}$$

where $\Omega(\bar{v}) = \frac{\varepsilon_p}{\varepsilon_p - 1} \left(1 - \frac{\int_0^{\bar{v}} \frac{v}{\bar{v}} dG(v) + \frac{1}{\varepsilon_p} \int_{\bar{v}}^{\infty} \left(\frac{v}{\bar{v}}\right)^{1/\varepsilon_p} dG(v)}{\int_0^{\bar{v}} \frac{v}{\bar{v}} dG(v) + \int_{\bar{v}}^{\infty} \left(\frac{v}{\bar{v}}\right)^{1/\varepsilon_p} dG(v)} \right) \in (0,1)$ is an aggregate distortion due to the inefficiencies that arise from rationing. It approaches one as all firms become capacity constrained but generally remains close to zero. The first term in parentheses captures the marginal change in the probability of becoming capacity constrained from a change in aggregate demand which reduces the weighted price elasticity of sales for the firm. The second term captures the counteracting effect from the increase in sales which occur in states of the world where the firm is not capacity constrained, making the firm care more about the higher price elasticity in these states and increasing the weighted price elasticity of sales for the firm.

The Rise in the Variance of Sales The limits to production within each period mean that not all of the idiosyncratic variation in demand passes through into variation in sales. Some firms will become capacity constrained from their idiosyncratic demand draw, and because of this the number sales for that firm will not match the desired demand. The capacity constraints truncate the upper tail of the idiosyncratic demand distribution as it passes through the firm into the distribution of sales. As a consequence, the variance of the sales distribution will be smaller than that of demand with the wedge between the two given by

$$\sigma_{\nu}^{2}-\sigma_{y}^{2}=\int_{\ln\bar{\nu}}^{\infty}\left(\ln\frac{\nu}{\bar{\nu}}\right)^{2}dG(\ln\nu)+\left(\int_{\ln\bar{\nu}}^{\infty}\ln\frac{\nu}{\bar{\nu}}dG(\ln\nu)\right)^{2}-2\mathbb{E}\left[\ln\frac{\nu}{\bar{\nu}}\right]\int_{\ln\bar{\nu}}^{\infty}\ln\frac{\nu}{\bar{\nu}}dG(\ln\nu). \tag{30}$$

Since the idiosyncratic demand was assumed to distributed log normally with mean 1 and variance σ_{ν}^{2}, $\ln\nu$ is distributed normally with mean $-\frac{\sigma_{\nu}^{2}}{2}$ and variance σ_{ν}^{2}. The gap between the two variances only depends on the demand threshold $\bar{\nu}$ or equivalently on the size of the capacity buffer $B(\bar{\nu})$ by (16). When the size of the capacity buffer expands, the demand threshold at which a firm becomes capacity constrained rises, and this increases the pass-through of demand into sales, reducing the size of the variance wedge in (30).[33]

Without data at the firm level, the cross-sectional variation is instead estimated across 6-digit NAICS manufacturing industries spanning 1958 to 2018. The annual log real value added is computed for each industry after which industry-specific levels and trends are removed. The sample is split into a pre and post 1990 period, and the unconditional variance of the residual log real value added is computed for both periods.[34] Using the year 1990 as the split point creates two approximately equal sized samples, with midpoints in the early 1970s and mid 2000s. The last column in Table 2 contains the estimates associated for these two periods despite not exactly matching the reference years of 1967 and 2020. The cross-sectional standard deviation of value added has increase by 0.06 points from 0.25 to 0.31.[35] Estimates for an AR(1) process of the residual log real value added as well as the standard deviation of its growth rate both reaffirmed the increase in it.[36] Reassuringly, the estimates at the 6-digit industry level are similar to those estimated at the firm level in Comin and Philippon (2005). The standard deviation of the residual growth rates, as measured by the log difference, has risen from 0.113 to 0.139, while they estimate a rise from 0.114 and 0.207 at the firm level.

4.3 Quantifying the Structural Changes

From a fall in the price elasticity of demand, the theory predicts a rise in the capacity buffer, a decline in the labor share, a flattening of the supply curve, and a rise in the cross-sectional variance of sales. In order to quantitatively deduce these structural changes, the level and fall in price elasticity of demand need to be derived. Estimating the key parameters underlying the theory of the firm thus serves as a test of the production theory and the rising markups hypothesis. When taken to the data, the theory implies

---

[33] See Kuhn and George (2019) for an analysis of the asymmetric business cycle characteristics that a general equilibrium model with capacity constraints (due to costly utilization) is able to match including countercyclical cross-sectional volatility, aggregate volatility, and fiscal multipliers.

[34] Given the i.i.d period by period draw of the idiosyncratic demand of the theory, the unconditional variance of the residual log real value added is the appropriate empirical moment to match.

[35] This estimate includes aggregate variation as time fixed effects are not removed. Aggregate variation is preserved in the estimate since firm capacity buffers are in reality also chosen with aggregate uncertainty in mind as well. When removing time fixed effects the standard deviation of log real value added similarly grew by 0.05 from 0.217 to 0.266, thus affirming that the increase in variance is cross-sectional.

[36] For the autoregressive process, the estimated standard deviations were 0.107 in the earlier sample and 0.134 in the later sample with persistence parameters 0.853 and 0.885, respectively. They imply unconditional standard deviations of 0.206 and 0.289 via $\sigma_y=\frac{\sigma}{\sqrt{1-\rho^2}}$, which show a similar level and increase. Using sales rather than value added yielded very similar results.

Table 2: Aggregate Empirical and Theoretical Moments

Year	B	S_L	σ_y
1967	0.106	0.644	0.252
2020 (Empirical)	0.226	0.572	0.311
2020 (Theoretical)	0.226	0.550	0.369

Notes: The values of B are its associated estimates for the complement of the federal reserves capacity utilization measure. S_L is taken from the OLS line of best fit for the labor share time series of the non-farm business sector. σ_y is the estimated standard deviation of the residual log real value added across 6 digit manufacturing industries from 1959-1990 and 1990-2018 after removing time fixed effect and industry trends.

reasonable parameter values and predicts changes in the moments of interest from 1967 to 2020 of the right magnitude.

The empirical moments used are the labor share, the capacity buffer size, and the standard deviation of log real value added. Their values associated with the first month of the years 1967 and 2020 are provided in Table 2. The capacity buffer size has risen from 10.6 to 22.6 percent of total capacity while the labor share has declined from 64.4 to 57.2 percent.[37][38] The estimates of the standard deviation of log real value added were discussed in previous paragraph.

The model parameters α, σ_ν, and ε_p^{1967}, are estimated using the empirical moments in the first row of Table 2 pertaining to the year 1967. Given these parameter estimates, of which α and σ_ν are assumed to be time invariant, the empirical estimate for the capacity buffer size in 2020 is used to estimate the necessary fall in the price elasticity of demand ε_p^{2020}. One can then infer what the 2020 values for S_L^* and σ_y are from the theory and check whether they match the empirical estimates. It also provides estimates for the level and rise in the pure profits markup over the time period, μ_{pp}^{1967} and μ_{pp}^{2020}.[39]

First, it is possible to jointly estimate the standard deviation of demand σ_ν and steady state capacity threshold $\bar{\nu}$ using the estimates of the capacity buffer size and standard deviation in value-added σ_y. This is done via the definition of the aggregate capacity buffer (16) and the mapping between sales and demand variances in (30). Second, with σ_ν and $\bar{\nu}$ on hand, it is possible to estimate both α and μ from the size of the labor share S_L using the optimality condition for capital intensity in steady state in (27) which reduces to

$$(\mu - 1)\int_{\bar{\nu}}^{\infty} dG(\nu) = \frac{\alpha}{1-\alpha}\left(\int_0^{\bar{\nu}} \frac{\nu}{\bar{\nu}}\, dG(\nu) + \int_{\bar{\nu}}^{\infty} dG(\nu)\right), \tag{31}$$

and the income shares allocation in steady state that was given in (28). Third, the price elasticity of demand ε_p can be backed out from the steady state markup $\mu(\bar{\nu})$ using (20) and (21).

---

[37] Assuming January 2020 as the steady state capacity buffer size has the benefit of seamlessly integrating the flattening Phillips curve analysis with the COVID-19 analysis, while January 1967 is the earliest entry in the series for total capacity utilization series for industrial production. Both are also conveniently reasonable candidates for the steady states given that they both represent the buffer size about a year after a peak in the capacity utilization time series (this can be inferred for January 1967 by inspecting the capacity utilization series for SIC manufacturing industries which peaked in early 1966 at 91.6%).

[38] The labor share values are derived from the Bureau of Labor Statistics' headline series for the non-farm business labor share. The reported values are the fitted ones from its trend estimated via ordinary least squares.

[39] The profits which are derived from μ also contain the return to capital and are therefore not the pure profits.

Table 3: Parameter and Slope Estimates

σ_v	α	ε_p^{1967}	ε_p^{2020}	μ_{pp}^{1967}	μ_{pp}^{2020}	ζ^{1967}	ζ^{2020}
0.529	0.278	9.25	4.20	1.12	1.31	3.68	2.04

Notes: Estimated model parameter values, pure profits markups, and supply curve slopes based on the empirical moments from 1967 in Table 2 and the 2020 estimate of $\hat{B}$. $\zeta = \frac{\partial \ln p}{\partial \ln Y}$ is the slope of the short run supply curve without price rigidities.

Table 3 depicts the parameter estimates using all of the empirical moments from 1967 plus the capacity buffer moment for 2020. With all of the parameters solved for 1967, the implied rise in the markup can then be determined by how much of a fall is needed in the price elasticity of demand ε_p for the size of the capacity buffer to rise to 0.226. These parameter estimates along with that of the pure profits markup are depicted in the table. The pure profits markup was estimated to be 1.12 back in 1967 with an associated price elasticity of demand of 9.25. The price elasticity of demand then proceeded to fall to 4.20, implying a rise in the pure profits markup of 19 percentage points to 1.31.[40] The parameter estimates fit within the ranges of commonly used values for them. The estimates of the level and rise in markups are also in the middle of those in the literature.

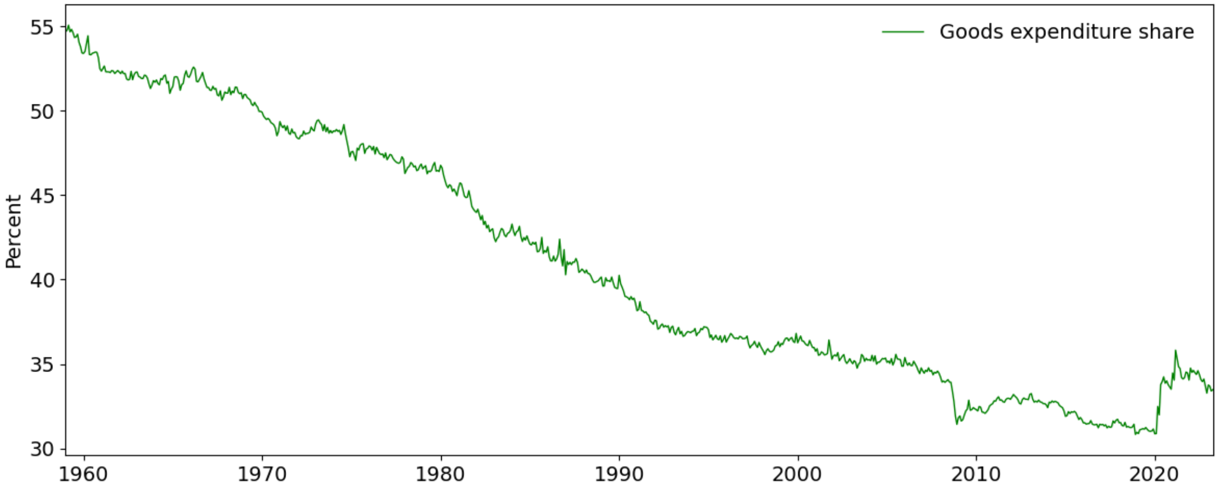
The only empirical moment used from the 2020 period is the aggregate capacity buffer size, which was used to estimate the fall in the price elasticity of demand. Theoretical versions of the other 2020 moments can therefore be compared to those estimated in the Table 2 where the 2020 moments implied by theory are also provided. The theoretical change in capacity buffer is identical by construction. The labor share decline to 55 percent while standard deviation of sales increases to 0.37. That is, both moments changed by slightly more than the their empirical counterparts, but are broadly consistent with the empirical ones. The fact that the theory predicts are larger rise in the cross-sectional standard deviation in value added is consistent with the larger increase that Comin and Philippon (2005) estimated at the firm level.

To quantify the flattening of the short run supply curve, the slope is determined at its 1967 and 2020 levels based on the estimated parameters. Estimates of these based on (29) are given in Table 3. The slope was 3.68 in 1967 and fell by 45 percent to 2.04 in 2020 due to the rise in the size of the capacity buffer. Two comments are in order regarding the slope of this supply curve. First, this is the local decrease in the supply curve, that is, in response to small variations in aggregate demand. The convexity of the supply curve in Figure 6 implies that the change in the desired price level to a large positive demand shock is larger, which is relevant in the case of the pandemic, for instance. Second, this slope is derived under a static setting with a fixed level of capital intensity and capital stock. In a dynamic setting these will invariably also adjust to help offset the changing pressures from demand. Due to their slow pace of adjustment, however, they do not fully offset the decline in the slope as is explored next in section 5.

---

[40] The pure profits markup is derived from the profit share of income as $\mu_{pp} = \frac{1}{1-S_{\text{Profits}}}$. It is also equal to the unconstrained markup, i.e. the markup that would occur if capacity constraints did not exist and the effective price elasticity of demand is equal to the price elasticity of demand, $\mu_{pp} = \frac{\varepsilon_p}{\varepsilon_p - 1}$.

Figure 7: Goods Sector Consumption Expenditures



Notes: The percentage share of personal consumption expenditures spent on goods.

5 The Flattening of the Phillips Curve

In this section the theory of the firm is put into a two sector general equilibrium model in order to estimate the flattening of the Phillips curve. The following section then uses the same framework to analyze its steepening in the goods sector during the COVID-19 pandemic.

Making the distinction between goods and services is necessary for both analyses as there was a large reallocation between goods and services during both episodes. Since prices are substantially stickier in the services sector as documented by Nakamura and Steinsson (2008), the services sector's Phillips curve is flatter than the goods sector's. Figure 7 shows that the share of consumption expenditures that is spent on goods has been declining consistently from a share of 55 percent of expenditures in 1960 to a share of 31 percent at the beginning of 2020. As a result the aggregate Phillips curve will also have declined partly due to the compositional change in expenditures away from the goods sector with its steeper Phillips curve and towawrds the services sector with its flatter Phillips curve.[41]

The gradual decline in the goods expenditure share abruptly ended at the onset of the COVID-19 pandemic. It sharply increased from its 31 percent to 34 percent and later reached a peak of near 36 percent. Partly underlying this was a shock to consumers preferences away from consuming services and towards goods, which resulted in a positive demand shock to the goods sectors that was larger than 10 percent, and a significant amount of inflation followed.

---

[41] The flattening due to growing share of the services sector has also been addressed by Clayton et al. (2018), Cotton and Garga (2022), and Rubbo (2023).

5.1 The Rest of the Model

Since the emphasis is on the novelty of the production side of the economy, the rest of the model uses a fairly standard framework in the New Keynesian literature and is kept as simple as possible. An overview and resulting equilibrium conditions are outlined below while the details of the household problem are provided in Appendix A.

Household Product Demand The economy is populated by a unit mass of households indexed by $h \in [0,1]$. Households smooth consumption across periods by trading a one-period bond that provides a nominal yield of r_t. With an elasticity of intertemporal substitution equal to unity, their felicity function over an index of aggregate consumption is logarithmic, and they discount future time periods geometrically according to the parameter β. Households are identical except for their differentiated types of labor. A representative household euler equation that equalizes their marginal utilities of consumption across time periods exists and is given by

$$\frac{1}{C_t} = \mathbb{E}_t \beta \frac{1+r_t}{1+\Pi_{t+1}} \frac{1}{C_{t+1}}. \tag{32}$$

The aggregate consumption index C_t is a CES composite of the consumption in the goods and services sectors. The sectoral demands for goods and services respectively are given by

$$C_t^g = \theta C_t \left(\frac{P_t^g}{\mathcal{P}_t}\right)^{-\sigma} \quad \text{and} \quad C_t^s = (1-\theta) C_t \left(\frac{P_t^s}{\mathcal{P}_t}\right)^{-\sigma} \tag{33}$$

where θ controls the sectoral consumption shares, and σ is the elasticity of substitution between goods and services. P_t^g and P_t^s are the shadow prices for the goods and services sectors which are each made up of a unit mass of firms as characterized by (17) in a symmetric equilibrium in section 3. $\mathcal{P}_t$ is therefore itself also a shadow price and given by

$$\mathcal{P}_t = \left(\theta P_t^{s\,1-\sigma} + (1-\theta) P_t^{g\,1-\sigma}\right)^{\frac{1}{1-\sigma}}. \tag{34}$$

The growth rate of the aggregate shadow price is denoted by $\Pi_t = \frac{\mathcal{P}_t}{\mathcal{P}_{t-1}} - 1$. The within sector demand is as described by (8) and (12) in section 3 where households have CES preferences with idiosyncratic taste shocks over the unit mass of firms' differentiated products with substitutability given by the price elasticity of demand ε_p.

Household Labor Supply Households each provide a differentiated type of labor and compete monopolistically, but are otherwise identical. They supply their labor to perfectly competitive employment agencies which in turn sells their bundles of labor on to firms. Households set their wage one period in advance subject to a Rotemberg wage adjustment costs in utility with parallel form to the firms' pricing rigidities. In the symmetric equilibrium, all households will set the same wage rate and supply the same amount of labor.

This results in the following aggregate wage Phillips curve

$$(\pi^w_{t+1} - \pi^w_t)(1 + \pi^w_{t+1}) = \mathbb{E}_t \kappa^w_t \left(\frac{\varepsilon_w}{\varepsilon_w - 1} \psi L^{\varphi}_{t+1} \frac{\mathcal{P}_{t+1}}{W_{t+1}} - \frac{1}{C_{t+1}} \right) + \mathbb{E}_t \beta^w_t (\pi^w_{t+2} - \pi^w_{t+1})(1 + \pi^w_{t+2}) \tag{35}$$

where $\pi^w_t = \frac{W_t}{W_{t-1}} - 1$ is the net wage inflation rate, and L_t is the aggregate labor supply. The temporal discount rate is $\beta^w_t = \beta \frac{1+\pi^w_{t+1}}{1+\Pi_{t+1}} \frac{L_{t+1}}{L_t}$, and the slope of the wage Phillips curve is $\kappa^w_t = \beta^w_t \frac{\varepsilon_w - 1}{\chi_w}$. The strength of the wage adjustment costs is controlled by χ_w, and ε_w is the own wage elasticity for the differentiated labor types. ψ controls the strength of the disutility from labor, and φ is the inverse frisch elasticity of labor supply.

Government The government influences the economy in two ways. The central bank sets the nominal yield on the one period bonds according to the Taylor rule

$$r_t = \rho_r r_{t-1} + (1 - \rho_r)(r^* + \phi_\pi(\tilde{\pi}_t - \pi^*)) + \epsilon^r_t, \tag{36}$$

which features persistence ρ_r. The Taylor rule parameter ϕ_π describes the vigor to which the central bank responds to inflation. $r^* = \frac{1+\pi^*}{\beta} - 1$ is the rate of interest in the steady state under the inflation target π^*. ϵ^r_t captures any unexpected deviations from the Taylor rule. The sectoral shadow prices and their aggregate are in practice not observable, and depend on the welfare cost associated with the rationing that occurs. These shadow prices can have large and rapid fluctuations depending on the severity of capacity constraints, and in these cases are not representative of the actual underlying prices. The central bank instead uses the actual underlying sectoral inflation rates measured by ignoring the Lagrange multipliers associated with the prices of capacity constrained firms. These sectoral indices are equal to a sector's firm's price changes. It uses a sectoral consumption weighted average of these for its inflation target

$$\tilde{\pi}_t = \Theta_t \pi^g_t + (1 - \Theta_t)\pi^s_t \tag{37}$$

where $\pi^j_t = p^j_t / p^j_{t-1} - 1$ for sectors $j \in \{g, s\}$ and $\Theta_t = \frac{\int_0^1 p^g_{i,t} y^g_{i,t} di}{\int_0^1 p^g_{i,t} y^g_{i,t} di + \int_0^1 p^s_{i,t} y^s_{i,t} di}$ is the measured good sector expenditure share.

The government also enacts fiscal policy by purchasing goods and services. It funds itself via lumpsum taxes on households, and it purchases goods and services according to the same within sector consumption bundles as the households. When it is not conducting any active fiscal policy, government expenditures occupy a fixed percentage of expenditures in both the goods and services sectors, $\frac{G^*}{Y^*}$. Active fiscal policy is described by deviations in the government expenditure shares in each sector. The impulses to these deviations are given by ϵ_t^G, and they follow an AR(1) process in logs with persistence ρ_G,

$$\frac{G_t^j}{Y_t^j} = \left(\frac{G^*}{Y^*}\right)^{(1-\rho_G)} \left(\frac{G_{t-1}^j}{Y_{t-1}^j}\right)^{\rho_G} \exp(\epsilon_t^G) \tag{38}$$

for sectors $j \in \{g, s\}$.

Equilibrium In equilibrium agents behave optimally under full information and rational expectations, and all markets clear. Firms are able to convert goods and services sector products into capital using the same bundle as is used for consumption and government expenditures. This means that P_t^I in section 3 is equal to the aggregate shadow price index $\mathcal{P}_t$. Total investment demand by firms is then given by $I_t = I_t^g + I_t^s$. Market clearing is expressed at the sectoral level by

$$Y_t^j = C_t^j + i_t^j + G_t^j + \frac{\chi_j}{2}\left(\pi_{t+1}^j - \pi_t^j\right)^2 Y_t^j \tag{39}$$

where i_t^j denotes the demand for investment goods or services at the sector level.[42] The last term in each of the market clearing conditions captures the firms' costs incurred from adjusting their prices.

Given the symmetric firm equilibrium in each sector, the sectoral labor demand by firms is equal to the expected labor demand of an individual firm before the idiosyncratic demand is observed, $l_t^j = \mathbb{E}_\nu l_{i,t}^j$. Market clearing for each household's labor supply can then be expressed at the aggregate level as

$$L_t = l_t^g + l_t^s. \tag{40}$$

Lastly, bonds are in net zero supply.

Putty-Putty Comparison The production side of both the goods and services sectors are assumed to be identical, and for the baseline model, they are described in detail in section 3, which is referred to as the putty-clay technology. In order to better understand the implications of the theory, it will be compared to that of a production framework without the temporal rigidities in capital intensity. The firm theory then reduces to what is most commonly used in the New Keynesian literature, which is a Cobb-Douglas production function. It will be referred to as the putty-putty technology. In this case, the idiosyncratic demand uncertainty has a negligible impact on the pricing and investment dynamics as a firm's production technology is log linear, and they can always adjust production to meet demand, and the capital intensity reduces to the capital to labor ratio, $x_{i,t} = k_{i,t}/l_{i,t}$, always. The investment optimality condition remains the same as before in (24). The marginal benefit of capital changes, however. Investing in an additional

---

[42] Investment was chosen to be comprised of both goods and services, in order to not introduce any asymmetries that would obscure the underlying mechanism related to capacity buffers. It has also become a more accurate description than assuming investment is only comprised of products from the goods sector. By 2020 investment expenditures on Intellectual Property Products had surpassed that on equipment.

unit of capital, lowers the marginal cost of production for the next period by increasing labor productivity. The marginal benefit of capital is therefore the future costs savings associated with labor. The optimality condition for choice of capital stock becomes

$$\frac{q_t^j}{\mathcal{P}_t} = \mathbb{E}_t \frac{1+\pi_{t+1}}{1+r_t} \left(\Sigma \frac{\alpha}{1-\alpha} \frac{W_{t+1}}{\mathcal{P}_{t+1}} \left(\frac{Y_{t+1}^j}{k_{t+1}^j} \right)^{\frac{1}{1-\alpha}} + (1-\delta_k) \frac{q_{t+1}^j}{\mathcal{P}_{t+1}} \right), \tag{41}$$

where $\Sigma = \exp\left(-\frac{\alpha}{(1-\alpha)^2}\frac{\sigma_\nu^2}{2}\right)$ is constant associated with the idiosyncratic demand uncertainty.

Of more direct relevance for the analyses of inflationary dynamics is the change in form of the Phillips curves for each sector $j \in \{g, s\}$, which becomes

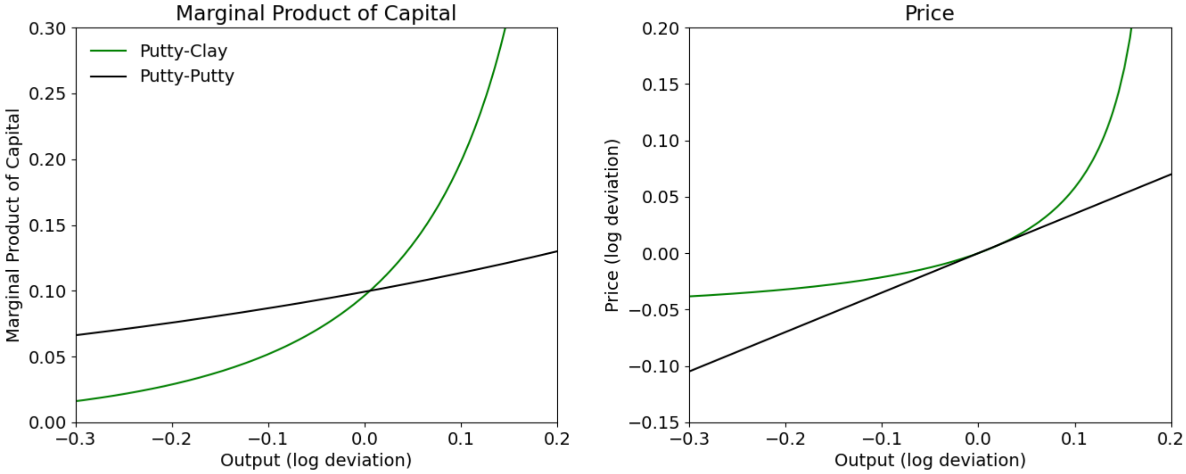
$$(\pi_{t+1}^j - \pi_t^j)(1+\pi_{t+1}^j) = \mathbb{E}_t \kappa_t^{j,pp} \left(\Sigma \frac{\varepsilon_p}{\varepsilon_p - 1} \frac{1}{1-\alpha} \frac{W_{t+1}}{P_{t+1}^j} \left(\frac{Y_{t+1}^j}{k_{t+1}^j} \right)^{\frac{\alpha}{1-\alpha}} - 1 \right) + \mathbb{E}_t \beta_t^{j,pp} (\pi_{t+2}^j - \pi_{t+1}^j)(1+\pi_{t+2}^j), \tag{42}$$

where $\beta_t^{j,pp} = \frac{1+\pi_{t+1}^j}{1+r_t}\frac{Y_{t+1}^j}{Y_t^j}$ and $\kappa_t^{j,pp} = \beta_t^{j,pp}\frac{\varepsilon_p - 1}{\chi_j}$. The amount of sectoral labor demand l_t^j can be determined from the aggregated sectoral production function $Y_t^j = k_t^{j\,\alpha}\left(l_t^j\right)^{1-\alpha}$. Comparing the Phillips curves from the different production structures (19) to (42) shows that the inflationary pressures arise via different avenues. Under the putty-putty framework of (42), prices will rise following an increase in demand due to higher marginal costs in production because the marginal product of labor declines under the fixed supply of capital, while markups remains constant. Under putty-clay, the price pressure arises mainly from a rise in the desired markups as a smaller capacity buffer lowers the price elasticity of sales.

Figure 8 shows a comparison of the short run supply and MPK curves computed in same manner as in section 3 with the capital stock (and intensity when its rigidities are present) are held fixed while allowing the price to vary flexibly. The levels of the desired price and associated marginal product of capital are then traced out for varying levels of aggregate demand. While the curves as were described earlier are highly log-convex in the putty-clay case, they are log-linear in the putty-putty setup when the rigidities in capital intensity are removed. Comparing the putty-clay with the putty-putty model highlights the importance of rigidities in capital intensity. In a dynamic setting, the firm's marginal product of capital and price would only behave exactly as the static putty-clay curve if the their capital intensity was completely fixed. The more flexibly the firms can change their capital intensity in production, the closer their marginal product of capital and desired price level will be to the putty-putty curves.

Calibration The full calibration is provided in Table 4. The key parameters are those related to the production theory, and their estimation strategy was covered in section 4. Most of the other parameters are set to commonly used values for them. The elasticity of substitution between goods and services is set to

Figure 8: Short Run Supply and MPK Curves



Notes: The curves trace out the value of the marginal product of capital and the desired price level for various levels of aggregate demand when the capital stock is fixed at their steady state in the putty-putty and putty-clay models. The capital intensity is also fixed at its steady state in putty-clay case. The slope of the log price curve for the putty-clay case has been globally scaled down to be tangent to the putty-putty slope in the steady state as is done in the NKPC slope calibration via ζ^{pp}/ζ^{pc}.

0.5 to remain as agnostic about its value while acknowledging that the two sectors are complementary.[43] ψ is calibrated such that aggregate consumption is equal to one in the steady state. There remain three less conventional parameters to be calibrated, the adjustment costs associated with prices in each sector and capital intensity, that warrant further discussion.

The strength of χ_x affects how quickly a firm is willing and able to substitute capital for labor or vice versa in the production process. Having assumed a Cobb-Douglas production frontier and a Leontief instantaneous production function implies that the capital labor elasticity of substitution will be unity if given an infinitely long duration of time to adjust and zero if there is no time to adjust. For a given finite duration of time, the capital labor elasticity of substitution will be between zero and one with the shorter the duration the closer to zero it will be. León-Ledesma and Satchi (2019) is the seminal and only paper to have estimated χ_x which yielded a value of approximately 20, which will be used here, under the assumption of an instantaneous capital labor elasticity of 0.2 rather than zero.[44] Given that their model differs in many details to this one, the results were checked for their sensitivity with respect to the value of χ_x to an order of magnitude above and below.

The sectoral price adjustment cost parameters χ_j for $j \in \{g, s\}$ are calibrated to match estimates for the slope of the sectoral New Keynesian Phillips curves (NKPCs). However, such estimates do not exist and must first be calculated. With knowledge of the aggregate NKPC slope and the relative price stickiness

---

[43] Estimates of the elasticity of substitution between goods and services span the range between 0 and 1. A consensus on its value does not seem to have emerged.

[44] León-Ledesma and Satchi (2019) also use a different functional form for the adjustment costs given by $(1 - \exp(-\frac{\chi_x}{2} g_{x,t}^2))$ on production rather than $\frac{\chi_x}{2} g_{x,t}^2$ on capital used here. They note that both forms yield similar results in their paper.

Table 4: Putty-Clay Calibration for January of 2020

Parameter	Description	Matched moment	Value
α	Net capital share	Capacity buffer	0.278
ε_p^{2020}	Price elasticity	Labor share	4.20
σ_v	Std. of idiosyncratic demand	Std. of value-added	0.529
θ	Goods sector share	Goods Share of PCE	0.31
β	Discount factor	2 pct nominal rate	0.9999
σ	Sectoral elasticity		0.50
π^*	2% inflation target		0.005
ϕ_π	Taylor rule		1.5
G^*/Y^*	Govt. output share		0.2
EIS	Log utility		1
δ	Depreciation rate		0.025
φ	Inverse Frisch		2
ρ_r	Nominal rate persistence		0.8
ρ_G	Fiscal shock persistence		0.8
ρ_Q	Capacity shock persistence		0.7
ρ_θ	Sectoral shift shock persistence		0.9
ε_w	Wage elasticity		6
χ_I	Investment adj. costs		4
χ_x	Production rigidites	LLS	20
χ_w	Wage NKPC slope	$\kappa_w = \kappa_p = 0.018$ (ACEL)	277.8
χ_s	Services NKPC slope	$\kappa_s = 0.004$ (ACEL & NS)	1884
χ_g	Goods NKPC slope	$\kappa_g = 0.043$ (ACEL & NS)	181.2

Notes: Contains the parameter symbols, descriptions, values, and method of calibration used for the quantitative analysis of the pandemic preference switch shocks. The same calibration is used for the flattening of the Phillips curve except that the analysis requires varying the values of the θ and ε_p. ACEL refers to the high-markup specification in Altig et al. (2011), NS refers to Nakamura and Steinsson (2008) ε is the steady state price elasticity of sales, and LLS refers to León-Ledesma and Satchi (2019). $\kappa_w = \kappa_p = 0.018$ from ACEL. For $j \in \{g, s\}$, $\kappa_j = \frac{\beta}{1+\pi^*} \frac{\varepsilon^* - 1}{\chi_j} \omega_j^* \zeta^{pc} / \zeta^{pp}$. Here, ζ^{pc} is equal to (29), $\zeta^{pp} = \alpha/(1-\alpha)$, and $\omega_j^* = \frac{p_j^* y_j^*}{P_j^* Y_j^*}$.

across sectors, the sectoral NKPC slopes can be deduced from the fact that the aggregate slope is the expenditure weighted average of the sectoral slopes. The aggregate NKPC slope estimate is taken from Altig et al. (2011) with a value of $\kappa_p = 0.018$ under their high markup specification. Their New-Keynesian price and wage Phillips Curves with Calvo type price rigidities and price-indexation are identical at first order to the Rotemberg price adjustment costs relative to inflation used here. Their estimate came from a dataset that spanned from 1982 to 2008, thus the goods sector share of consumption expenditures for year 2000 was chosen as the relevant sectoral shares for this estimate, which was $\theta^{2000} = 0.36$.

The relative NKPC slope of the goods sector to the services sector is estimated using the functional form for them in the Calvo model where they depend on the sectors' price hazard rates, which are taken from Nakamura and Steinsson (2008).[45] This yields a goods sector slope that is 10.4 times as steep as the services

---

[45] The relative slope of the goods sector NKPC to the services sector in the Calvo model is $\frac{\kappa_g}{\kappa_s} = \frac{(1-\iota_g)(1-\beta\iota_g)\iota_s}{\iota_g(1-\iota_s)(1-\beta\iota_s)}$ where ι_j is the probability that a firm cannot optimally update its price in any given period. The expected price duration for each sector are taken from Nakamura and Steinsson (2008), which results in the estimates $\iota_g = 0.51$ and $\iota_s = 0.81$. The average price duration are taken from table 2 in their paper and equal 5.27 quarters for the services and 2.03 quarters for the goods sector with $\iota^j = 1 - \frac{1}{\text{Duration}_j}$. The services sector's duration is reported directly while the goods sector's duration is the weighted average of the reported goods sector

Table 5: Capacity Buffers and Expenditure Shares Across Time

	1960	1980	2000	2020
B	0.10	0.12	0.16	0.23
θ	0.53	0.47	0.36	0.31

Notes: The sets of parameter value used the size of capacity buffers and expenditure shares at different moments across time to estimate the gradual flattening of the Phillips curve. The parameter values for the price elasticity of demand, which induce the capacity buffer sizes, are set based on the estimates of it provided in Table 3 assuming that the price elasticity of demand declined linearly since 1960 until 2020. They are 9.91, 8.01, 6.11, and 4.20.

slope. Given the sectoral expenditure weights, the aggregate NKPC slope estimate, and the relative slope estimate, the sectoral NKPC slopes are calculated to be $\kappa_g = 0.043$ and $\kappa_s = 0.004$. The sectoral price adjustment costs under putty-putty model can then be set to match these sectoral Phillips curve slopes.

Two additional adjustments must be made in the calibration of the putty-clay price adjustment costs. First, the fact that the sensitivity of prices with respect to changes in output differs across the putty-putty and putty-clay frameworks must be accounted for. In the putty-clay framework the sensitivity ζ^{pc} is given in (29), while for the putty-putty it is $\zeta^{pp} = \frac{\alpha}{1-\alpha}$. Second, there is a distortion associated from the misallocation due to capacity constraints given by $\omega_j = \frac{p_j y_j}{P_j Y_j}$, which shows up in the putty-clay NKPC slope due to the difference between the adjustment costs which are denoted in terms of the aggregate sectoral good, and the revenue the firm makes that is specified in terms of its own product.

5.2 The Flattening Phillips Curve

As capacity buffers have gotten larger, firms' pricing decisions have gotten less sensitive to variation in aggregate demand, resulting in a flattening of the Phillips Curve. As in Del Negro et al. (2020), the change in its slope will be interpreted from the dynamic responses of inflation to demand shocks, with a weaker response implying a flatter slope. The impulse responses of inflation will be estimated through the lens of the model in response to two types of demand shocks: a fiscal policy shock ϵ_t^G and a monetary shock ϵ_t^r. The magnitude of the shocks will be scaled so to induce a one percent increase in aggregate employment on impact. The flatness of the Phillips curve can then be understood from how much of an increase in inflation is associated with the one percent increase in employment.

In order to determine how much flattening can be attributed to the increasing size of capacity buffers from the rising markups, the impulse responses will be estimated under the different sizes of the capacity buffers, B, and sectoral expenditure shares, θ, that match the US economy in 20 year increments. These values are provided in Table 5. The goods expenditure shares match their measured values as depicted in Figure 7. The capacity buffer is set by recalibrating the price elasticity of demand, based on the estimates of it provided in Table 3 assuming that it declined linearly since 1960 until 2020.[46][47] The model is assumed to

---

sub-categories.

[46] The rise in markups is assumed to occur in both sectors consistent with evidence in Crouzet and Eberly (2020) and Marto (2023) who finds that markups have risen in both sectors with them recently rising more in the service sector.

[47] Note that this also necessitates a recalibration of χ_s and χ_g for each value of the price elasticity of demand in order to keep the

be in its steady state before the shock after which an unexpected deviation from zero in the value of ϵ_0^r or ϵ_0^G occurs.[48]

The Effects of Monetary and Fiscal Shocks Across Time Figure 9 depicts the impulse responses of inflation to the expansionary monetary and fiscal policy shocks for the calibrations of each point in time given in Table 5. Figure 9a and Figure 9c provide the responses under the putty-putty framework when no rigidities in capital intensity are present. A slight dampening in the response of inflation occurs from the increasing share of services in the make-up of the economy with its stickier prices. Between 1960 and 2020 the initial response of inflation in the quarter following the shock declined by 26% and 24% in response to monetary and fiscal shocks, respectively. In contrast, Figure 9b and Figure 9d show a dampening under the putty-clay technology of 53% and 55% for the monetary and fiscal shocks, respectively. In this case the additional dampening came from the increasing size of firms' capacity buffers.

Since the dynamic response of employment is not exactly identical across models as well as for different calibrations of each, a more robust measure of the flattening is the ratio of the integrals of the responses of inflation and employment. For the putty-putty model, this ratio declined between 1960 and 2020 by only 17 and 6 percent in response to the monetary and fiscal shocks, respectively. For the putty-clay model, and therefore with the additional effect of the increasing capacity buffer, this ratio declined by 37 and 39 percent in response to the two shocks.

As can be gauged from the differential effects across the technologies, both the gradual change in sectoral composition and the rise in capacity buffers played substantial roles in the dampening of inflation. In order to disentangle their effects, Figure 10 shows the isolated role of each and when combined in terms of the cumulative effect on the price level in the capacity buffer model. For both types of demand shocks, the combined decline in the rise on the price level from the structural changes was 37 and 39 percent after eight years for the monetary and fiscal shocks, respectively. The increase in the size of capacity buffers was responsible for just over half of the dampening in response to monetary policy shock, while for the fiscal shock it accounted for about two thirds of the dampening.

There is no consensus estimate of how much the slope of the Phillips curve actually flattened due to the difficulty in its estimation and the variability in its structural form. However, estimates generally show at least a halving of its slope.[49] When isolating the role of the capacity buffer, and measuring the amount of flattening it caused from the ratio of the integrals of the responses of inflation and employment, the growing size of capacity buffers played a prominent part in this flattening causing a decline due to them of 20 and 26 percent for the monetary and fiscal shocks, respectively.

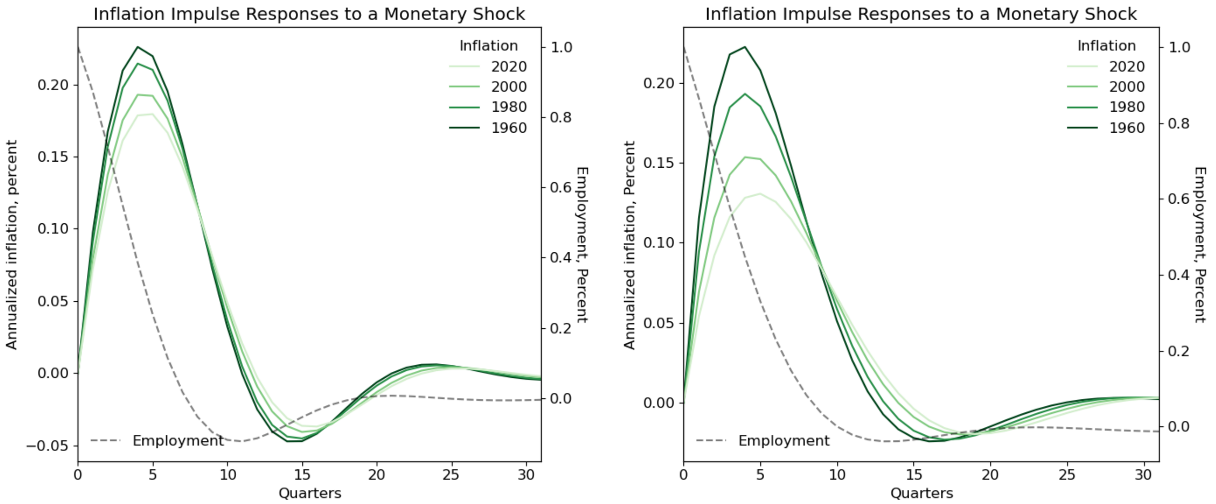
---

slope of the NKPC associated with the average price durations fixed. The fact that the NKPC slope to first order is directly increasing in the price elasticity of demand is viewed as an artifact of the Rotemberg price rigidities (since it does not appear under Calvo price rigidities), and not of economic interest despite that fact that it would imply an even more extreme degree of flattening.

[48] The impulse responses are estimated by performing a first order perturbation around the steady state of the model.

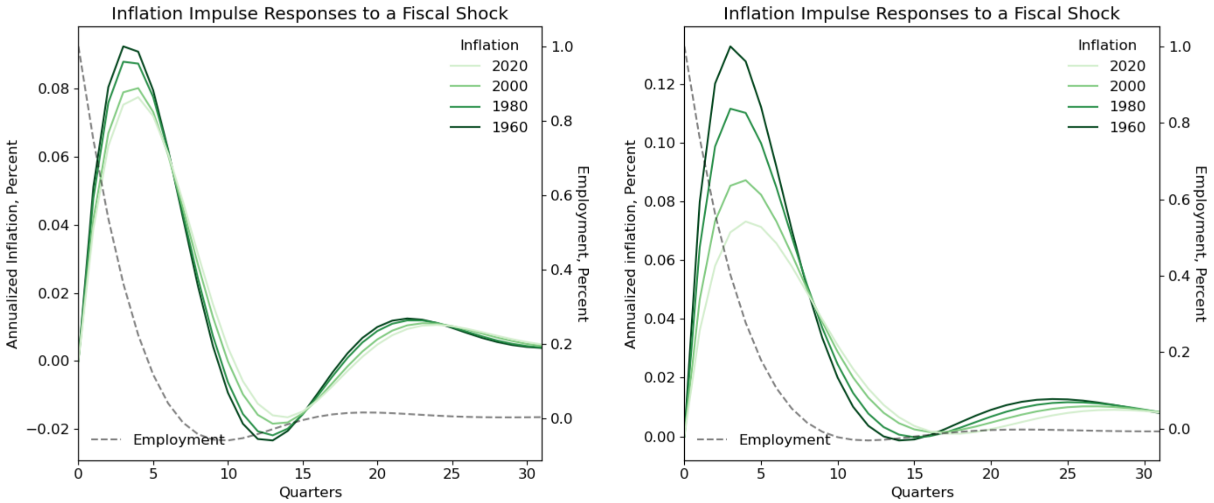
[49] See Hooper et al. (2020), Hazell et al. (2022), and Del Negro et al. (2020).

Figure 9: Inflation Responses to Expansionary Fiscal or Monetary Shocks



(a) MP shock, Putty-Putty Technology

(b) MP shock, Putty-Clay Technology

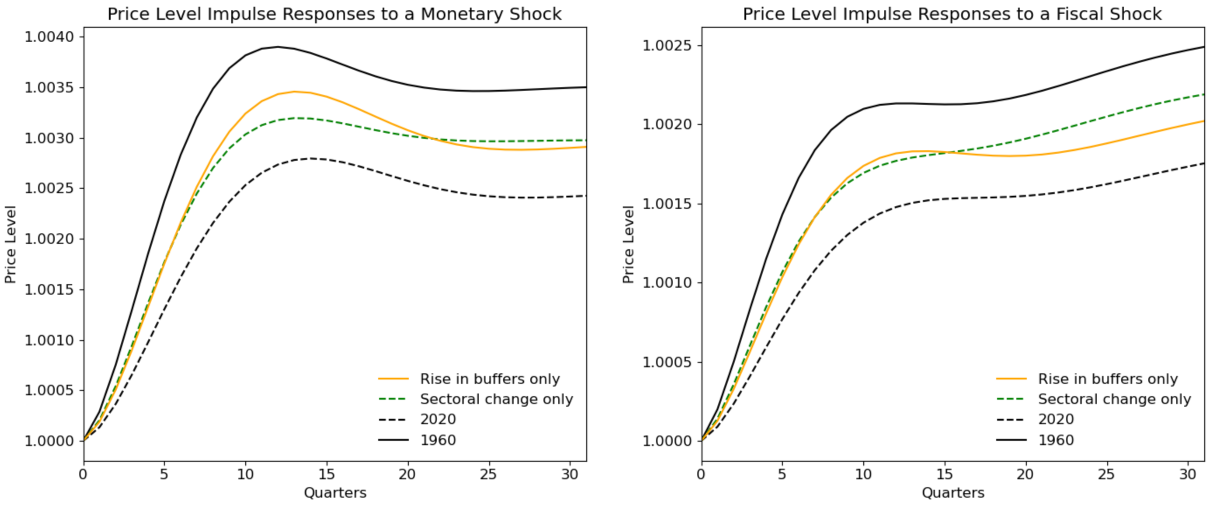


(c) FP shock, Putty-Putty Technology

(d) FP shock, Putty-Clay Technology

Notes: Model impulse responses of inflation and employment to an expansionary fiscal or monetary policy shock that results in a 1 percent increase in employment on impact. Impulses are provided for various sets of parameter values for ε_p and θ according to the their values in 1960, 1980, 2000, and 2020. The left panel depict the response under a putty-putty technology. The right panel depicts it under the putty-clay technology.

Figure 10: Decomposed Response of Price-level to Expansionary Monetary and Fiscal Shocks



Notes: Model impulse responses for the price level to an expansionary fiscal policy shock that results in a 1 percent increase in employment on impact. Impulses are provided for various sets of parameter values for ε_p and θ according to the their values in 1960 and 2020. The "Rise in buffers only" curve keeps the sectoral consumption share fixed at the 1960 level. The "Sectoral change only" keeps the capacity buffers fixed at its 1960 level.

5.3 Larger Fiscal Multipliers

In addition to the flattening of the Phillips curve, the growing size of capacity buffers should also increase the effectiveness of fiscal policy. A fiscal shock is more effective at raising output when firms have large capacity buffers that are able to absorb the fiscal stimulus while having little effect on the price level. Conversely, the effect of fiscal policy on output is more limited when firms have small capacity buffers since the fiscal stimulus crowds out consumption and investment given the more limited feasible increase in production.[50]

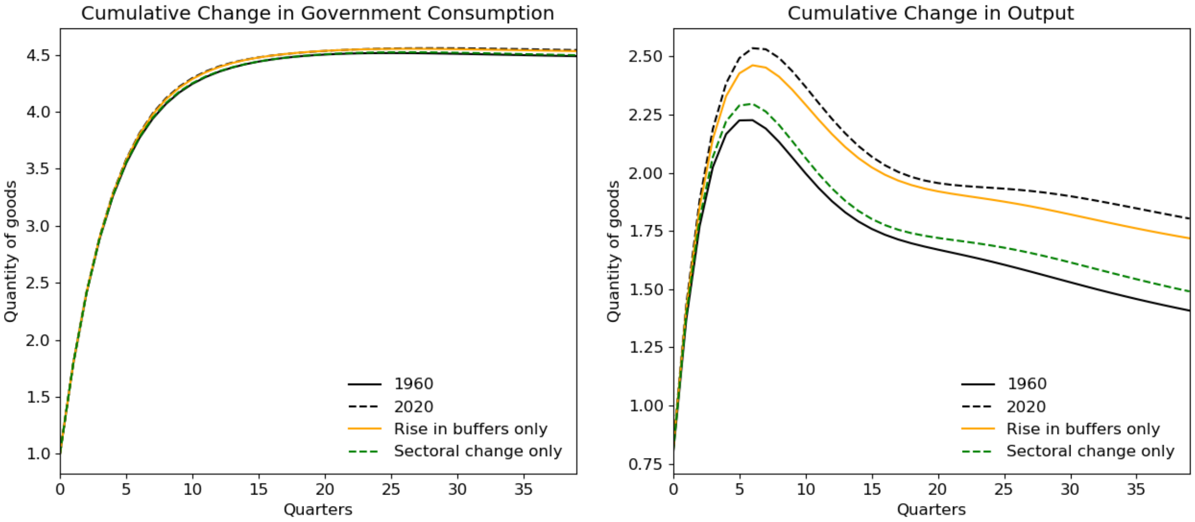
To see this, the responses of sectoral output are estimated to a fiscal shock that leads to a one unit increase on impact in the government consumption for each of them. Figure 11 depicts the cumulative changes in government consumption of goods in the left panel and in the total amount of goods produced in the right panel. The cumulative fiscal multiplier, as measured by the ratio of the cumulative effect on output to that of government consumption, increased by 27 percent from 0.31 to 0.40.[51] Most of the effect came from the rise in the size of capacity buffers. The increasing services share of the economy made a minor contribution since the flattening of the aggregate Phillips curve associated with it dampened the response of aggregate inflation and, as a consequence, the response of interest rates.

---

[50] Kuhn and George (2019) show that the size of the fiscal multiplier should also vary over the business cycle due to capacity constraints.

[51] The effects in the services sector parallel that of the goods sector as shown in Figure 17 in Appendix C where the cumulative fiscal multiplier increase from 0.41 to 0.49. In both cases the effect that is derived from the rise in capacity buffers only is around 20 percent.

Figure 11: Cumulative Response in Goods Output to a Fiscal Shock



Notes: Model cumulative impulse responses for the quantity of goods produced in the right panel and consumed by the government in the left to a fiscal shock. Impulses are provided for various sets of parameter values for ε_p and θ according to the their values in 1960 and 2020. The "Rise in buffers only" curve keeps the sectoral consumption share fixed at the 1960 level. The "Sectoral change only" keeps the capacity buffers fixed at its 1960 level.

6 The COVID-19 Inflation

The events which unfolded around the Covid-19 pandemic disrupted the US economy in more ways than one. These include the initial lockdown with its large rise in unemployment, the required hazard pay to compensate exposed workers, the normalization of remote work, the fiscal and monetary stimulus, and later, the supply chain disruptions and energy price spikes from the war in Ukraine. This section focuses on two of the initial shocks to the economy which when combined can account for the first wave of the COVID-19 inflation.[52] There was a large switch in consumption away from services and towards goods. As shown in Figure 12a this enlarged demand for goods persisted with little reversion to the pre-pandemic trend by 2023. At the same time as households increased their demand for goods, firms' ability to produce were limited due to the restrictions on production that were imposed to limit contagion. Firms' then limited capacity buffers were insufficient for accommodating the larger than 10 percent rise in demand for goods. This resulted in a large increase in prices.

Firms initially became limited in their production because of the pandemic health restrictions that inhibited firms from using their capital stock to its full extent. The same capital did not yield the same level of capacity as before because of the preventative measures to reduce contagion such as social distancing. Production lines in manufacturing plants were forced to slow down as workers needed more space, staggered shifts, and more downtime was needed for sanitation. Restaurants' occupancy rates were reduced as fewer

---

[52] Notably, it does not attempt to model the energy shock which followed the Russian invasion of Ukraine, nor the shortages of intermediate inputs, both of which contributed to inflation in 2022.

people per square foot were allowed.[53] Social distancing, amongst other preventative measures, effectively resulted in a negative productivity shock to capacity.

As part of the Quarterly Survey of Plant Capacity, firms are asked to list the reasons for why they are opearting below capacity where capacity is defined as full use of their capital stock. Figure 12b shows the answers during the COVID-19 pandemic for two of the reasons. The initial health restrictions led to a sharp increase of more than 20 percent in the number of firms reporting "Other" as a reason for not operating at full capacity. The restrictions on production meant that firms were not allowed to operate at full capacity, and this reason was not one of the standard option to tick off on the survey form. At the same time there was a decline from around 10 to 5 percent in firms reporting that equipment limitations were the reason why they were operating below capacity. Both shifts in responses gradually recovered to their pre-pandemic level by 2022, which indicates that the reasons why equipment was not limiting likely being due to it not being the relevant binding constraint. Because of these production limitations, firms' measured capacity buffers did not decline by much during the pandemic recovery despite the large increase in the demand for goods.[54]

Firms are also asked why they reported a change in their capacity from the preceeding quarter if they had done so. Figure 19 in Appendix C provides these survey responses which similarly shows a sharp uptick by almost 20 percentage points in the choice for the reason being "other" since the health restrictions were not a standard choice on the form. This high-frequency shift in capacity does not appear in the Federal Reserves' capacity series for each industry, however, because they intentionally remove high frequency movements in reported capacity by regressing them on measures of changes in the capital stocks. This is done to ensure that their measure is defined with respect to the production capacity that the capital stock can bear, and they find the cyclical variation in reported responses on capacity implausible.[55]

6.1 The Effects on Inflation of the demand shift and capacity Shocks

The shock to sectoral demand for goods versus services is modeled as a shock to the parameter θ which governs the households sectoral preferences for goods and services. The capacity shock is modeled as a

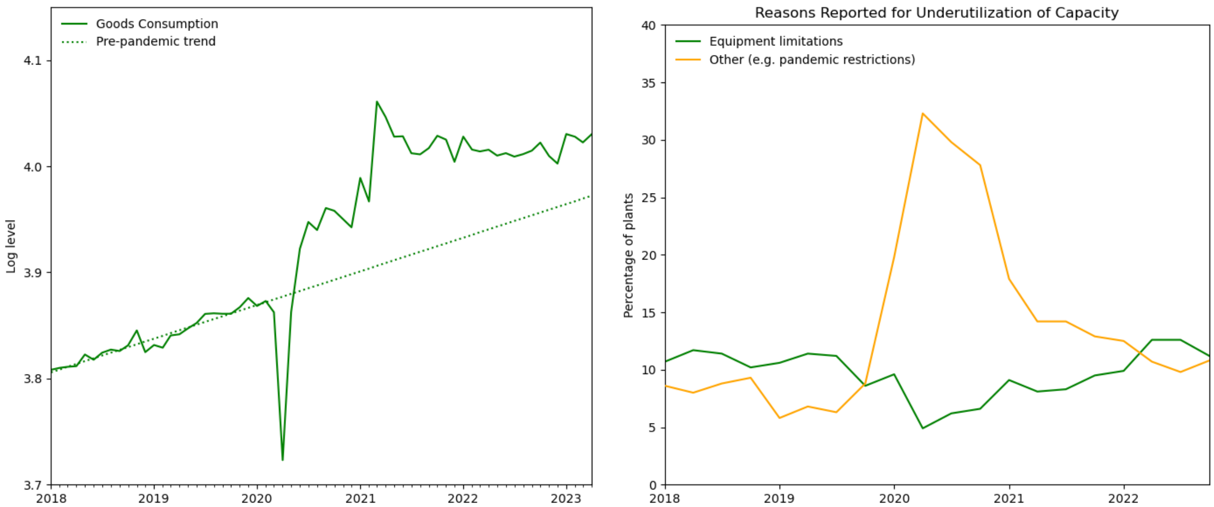
---

[53] Whitehead and Kim (2022) discuss the slow down of production lines in their review of the meat-packing industry during COVID-19. Boudette (2020) discusses the production line slow down at auto plants. Examples of legal impositions include the state of New York's limits to occupancy rates at restaurants and bars (Otterman, 2021), and the executive order in Illinois mandating reduced line speeds in manufacturing (Pritzker, 2020).

[54] Figure 18 in Appendix C reports the other reasons for why firms have excess capacity. Before the pandemic, the main reason for operating below capacity, with more than 70 percent of firms selecting it as one of their reasons, was due to an insufficient amount of orders. According to this theory, this is by design since firms choose to have capacity buffers, so demand should generally be insufficient relative to capacity. A number of disturbances then followed from the COVID-19 pandemic. As the production limitations associated with the health restrictions begun to decline, however, supply chain issues began to limit firm's production capabilities instead. By the year 2022, the percentage of plants reporting that they had insufficient supply of materials or labor rose to above 40 percent. Both of these disturbances also had the effect of temporarily reducing firms production capabilities relative to their capacity as measured by their capital stock and therefore further exacerbated the rise in inflation. Though the variable input constraints certainly played a role in the second wave of inflation in the beginning of 2022, the analysis here focuses on the inflation that ensued from the initial pandemic disruptions. Balleer and Noeller (2023) emphasize the importance of taking variable input constraints into account with the finding that these also lead firms to be more likely to raise prices in response to a monetary policy shock. An analysis of the effect of the intermediate input constraints on inflation in the capacity buffer framework would require adding additional complexity to the theory and is beyond the scope of this paper. However, it would certainly be a worthwhile pursuit.

[55] See Shapiro et al. (1989) for more details on its construction.

Figure 12: Demand Shift for Goods and Excess Capacity Survey Responses



(a) Demand Shift to Goods

(b) Reasons Reported for Excess Capacity

Notes: The left panel depicts the paths of consumption for goods. The right panel reports the main reasons that manufacturing plants indicate for not operating at capacity and the percentage of plants indicating them in the census' quarterly survey of plant capacity.

productivity shock Q_t to a firm i in sector j's capacity at time t,

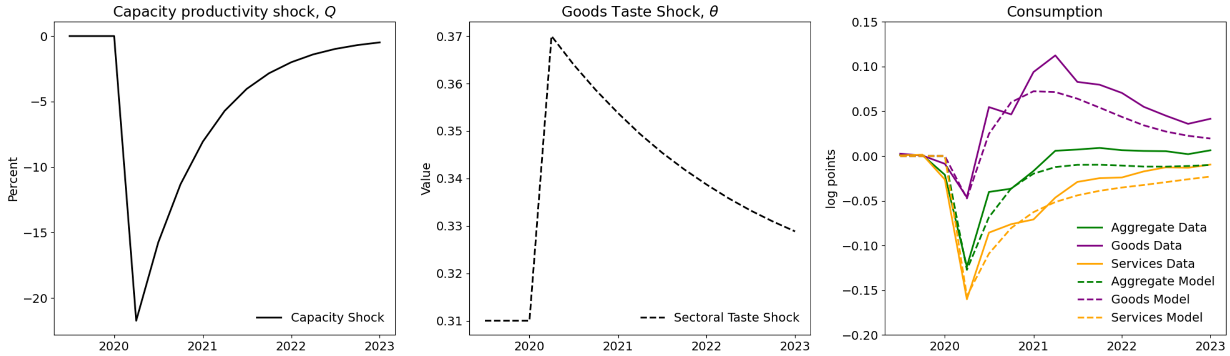
$$\bar{y}^j_{i,t} = a^{k,j}_{i,t} k^j_{i,t} \qquad \text{where} \qquad a^{k,j}_{i,t} = Q_t {x^j_{i,t}}^{\alpha-1} \tag{43}$$

Both shocks follow AR(1) processes. The persistence for the sectoral shift shock is set to $\rho_\theta = 0.9$ considering the marked persistence of the elevated goods consumption share in Figure 12a. On the other hand, a persistence of $\rho_Q = 0.7$ for the log of the capacity shock is chosen in order to have it disappear relatively quickly with 95 percent of it gone after just two years. This matches the dynamic path of the survey responses to the number of firms reporting "other" as a reason for operating with excess capacity.

The ensuing inflationary response would imply a counterfacutal contractionary monetary policy with interest rates rising according to the Taylor rule in (36). To obviate this, the nominal interest will be fixed at its steady state rate for 8 quarters following the shock, the same duration as the Federal Reserve kept overnight rates at the zero lower bound. In this way it mimics the fact that in reality the short term interest rates were fixed but without applying the expansionary monetary policy shock associated with lowering it to zero. The impulse responses are solved assuming agents have perfect foresight and that no other shocks occur with the economy eventually reverting to its steady state.

The economy is assumed to be in its steady state leading up to the pandemic after which an unexpected an identical negative productivity shock to capacity and a sectoral preference shock in favor of the good sector hits each sector. The size of the capacity and preference shocks are calibrated to match the initial decline in consumption in both goods and services. Specifically, the size of the capacity shock is set to

Figure 13: Capacity and Sectoral Preference Shocks Calibration and Quantity Responses



Notes: Dynamics paths of capacity and preference shocks along with the empirical and model implied paths of consumption used for calibration.

match the overall decline in consumption on impact, while the size of the preference shock is calibrated to match the differential decline in the services and goods sector declined. The paths of shocks are shown in Figure 13. The magnitude of the capacity shock is 22 percent, and the change in the value of the preference parameter is 0.06.[56]

The calibrated shocks also do well at matching the the dynamics paths of consumption. After the initial impact in the second quarter of 2020, a large amount of the capacity shock has already subsided as the lockdown has ended. The goods sector consumption surged to nearly a peak increase of 10 percent relative to before the pandemic, while the services sector only slowly recovered. By 2023, the levels of consumption are slightly lower in the model than in actuality. This is feature, not a bug, of the exercise. The actual path of aggregate consumption rose to above its pre-pandemic trend by the middle of 2021, most likely due to the expansionary fiscal and monetary policies. The exercise conducted here includes no sort of positive aggregate demand shock to push the model into an expansionary phase above its steady state. Even so, it still produces close to the measured levels of inflation, as will be discussed next.

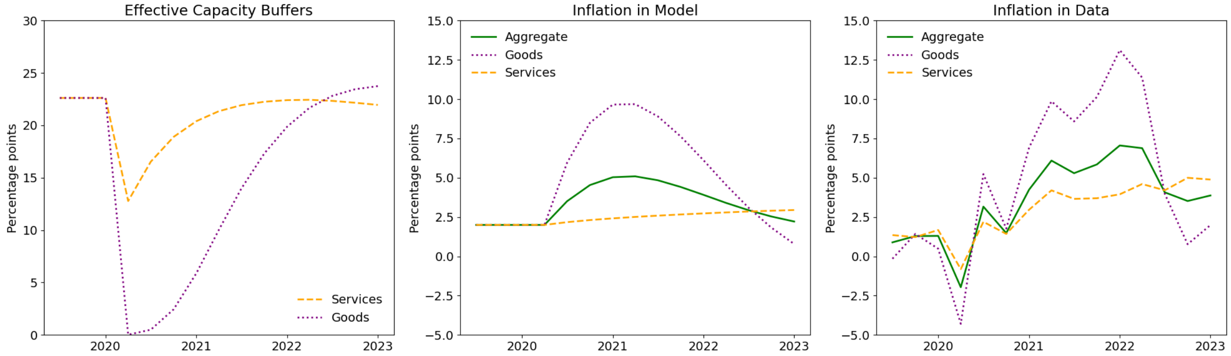
The Effects of the Demand Shift and Capacity Restrictions Figure 14 depicts the impulse responses of inflation across sectors to the pandemic induced decreases in capacity and shift in consumption from services to goods. The middle panel depicts the paths of inflation for goods, services, and the aggregate in the model while the right panel depicts their actual paths in the data.[57] The services sector featured little inflation which only rose slowly. However, the goods sector experienced a substantial amount of inflation

---

[56] The combination of shocks implies that the initial decline in the services sector was a result of of negative supply and demand shocks, while the initial decline in the goods sector came solely from a negative supply shock. This is consistent with the findings of Brinca et al. (2021) who find that in the aggregate most of the initial drop in output was primarily due to a supply shocks. At the sectoral level, they find it was nearly fully a supply shock in goods industries, while being more evenly split between demand and supply shocks in services industries.

[57] The empirical inflation measures used are the those associated with personal consumption expenditures for aggregate, goods and services sectors. The employment cost index is used to calculate the wage inflation. All inflation measures are demeaned and then set to have the same mean inflation rate as the aggregate consumption price index in order to remove differential long run levels of inflation across indices due to labor productivity growth and investment specific technical change, which the model abstracts away from.

Figure 14: Impulse Responses to the Demand Shift and Capacity Restrictions



Notes: Full nonlinear impulse responses in model to the capacity and demand shift shocks while keeping the monetary policy fixed at its steady state for eight quarters.

through 2021, and as a result the aggregate did as well. The amount of inflation the model produces relative to the empirical peak in the second quarter of 2021 was 84% of aggregate inflation, 98% for the goods sector, and 60% in the services sector. In accordance with the data, the model also features the reversal in sectoral inflation rates by the end of 2022 with higher inflation in the services than goods sector.[58]

Why these two shocks resulted in a large rise in inflation in goods sector and only a minor rise in the services is best understood from looking at their impact on firms' capacity buffers depicted in the first panel of Figure 14. The production restrictions in both sectors reduced firms' production capacities and therefore their capacity buffers as well. While the expenditure switch from services to goods further reduced the size of firms' capacity buffer in the goods sector effectively to zero. The decreased demand in the services sector had the opposite effect and partially counteracted the decrease in capacity buffers in the services sector. Consequently, there was a large desire to raise firms prices in the goods sector but only a minor one in the services sector.

The Role of the Sectoral Consumption Shift, Nonlinearities, and Relative Price Rigidities Despite most of the initial drop being due to a decrease in production capacity, most of the inflation that followed came from the shift in consumption away from services and towards goods. Only imposing the sectoral shift in demand still results in large rise in inflation.[59] The main reason for this was the size and persistence of the shift in demand. This demand shift accounted 59 percent of the increase in aggregate inflation at its peak in second quarter of 2021, while the capacity shock accounted for 31 percent, and their interaction accounted for 10 percent.[60]

---

[58] The model does not capture the initial deflation as well as the second peak in inflation during the winter of 2022. However, a large contributor to both of these were changes in energy prices. Oil prices momentarily went negative early in the pandemic, and the invasion of Ukraine in 2022 led to a large increase in energy prices. A quick way to show this is to compare the model response to core PCE inflation and to durable goods inflation as is done in Figure 21 in Appendix C. Most of the initial decline in inflation as well as the inflation of 2022 disappears with the new global peak in inflation being in the second quarter of 2021 as in the model.

[59] The impulse responses under the counterfactual scenario where there was no shock to the productivity of capacity are shown in Figure 22 in Appendix C.

[60] That a majority of the initial surge in inflation in 2021 resulted from changes to demand is consistent with the findings of Shapiro et al. (2022) and Sheremirov (2022) who categorize the contributions to inflation into demand and supply shocks throughout the

Alvarez-Lois (2006) showed that to first order the capacity shock is akin to that of a standard NKPC markup shock as the reduced capacity buffer induces a larger desired markup. What this linearization loses is the fact that the reduced capacity buffer also steepens the firm's supply curve. That is, the sensitivity of prices to variation in demand also increases because of the temporarily smaller capacity buffer, amplifying the effect on prices from the demand shock. But while the contribution to inflation from the interaction of the two shocks is 10 percent, this does not make up the entire nonlinear effect of because the shocks were of such a large size that there is relevant nonlinear component within them as well. To gauge the overall contribution to inflation that is due to the nonlinearities of the framework, the first order responses of inflation to the same combination of shocks is computed and compared to those in the nonlinear setting.[61] The response of aggregate inflation in the linearized analysis is 79 percent of that in the nonlinear one, implying a role of the nonlinearities that is similar in size to that of the direct effect of the negative shock to capacity.

The fact that prices are less sticky in the goods sector relative to services again played a substantial role in amplifying the surge in inflation. This is due to the fact that the sector which experience the large increase in demand was the sector with the less sticky prices, enabling it to raise them by more. Performing the same analysis but without a difference in price rigidities across sectors results in a picture that is qualitatively similar, but with a less vigorous responses in inflation. By the second quarter of 2022 the resulting inflation in the goods sector was 47 percent lower than in the baseline analysis, reaching a peak of nearly 6 percent.[62]

The Role of Monetary Policy The baseline results are in contrast to Comin et al. (2023) and Gagliardone and Gertler (2023) who both find that loose monetary policy played a large role in driving the pandemic inflation. In both of their papers, they argue that when paired with a supply shock, whether it be an energy price shock or negative shock to production capacity, the expansionary monetary policy accounted for much of the rise in inflation. In this study's baseline analysis, the nominal interest rate is held constant at its steady state level in order to limit its role in driving the responses of inflation to the capacity and demand shift shocks. It turns out that even when the Taylor rule is allowed to be active, thereby resulting in elevated interest rates and imposing a contractionary monetary policy, there is still a marked response of increase in inflation in response to the shocks. These figures are provided in Figure 23 in Appendix C.

To test the potency of combing the capacity shocks with the loose pandemic monetary policy, the impulse responses of inflation are estimated to the combination of negative productivity shocks to capacity and a monetary policy shock that lowers the interest rate by 1.5 percentage points and fixes it there for 8

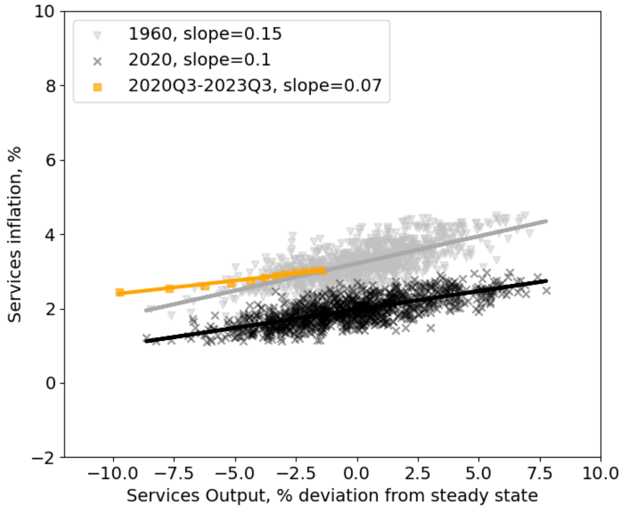
---

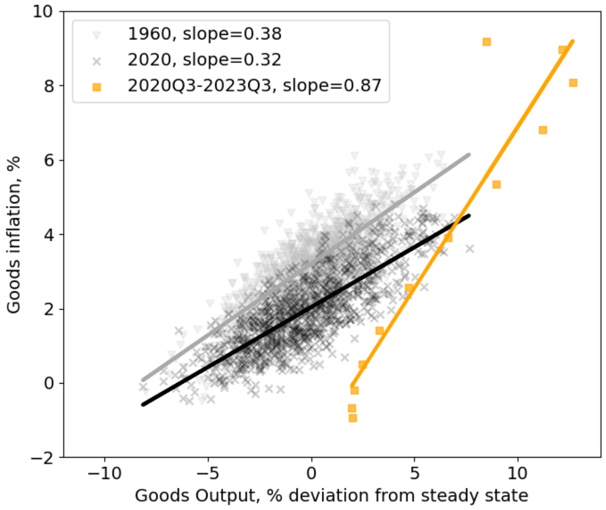
pandemic. While both types of shocks were present throughout the pandemic, demand shocks played a larger role in the rise in inflation of 2021 while supply shocks dominated over 2022.

[61] In order to not have a differential path of interest rates between the linearized and nonlinearized versions of the model, the persistence parameter ρ_r in the Taylor rule is set to 0.999, which inhibits any noticeable response of interest rates in both settings while remaining sufficiently far away from indeterminacy. This alternative monetary policy path results in a slightly less vigorous response of inflation than in the baseline exercise. Thus the size of the nonlinear component in this alternative specification is a lower bound for it in the baseline exercise since the nonlinear component will be larger the larger the responses are.

[62] The sectors' price rigidities were each calibrated to match the aggregate estimate of the slope of Phillips curve in Altig et al. (2011). The impulse responses are depicted in Figure 25 in Appendix C.

Figure 15: Model Phillips Correlations





Notes: Phillips Correlations using simulated data.

quarters. In this case the productivity shocks to capacity are allowed to differ in size across the two sectors in order to calibrate them to match the differential declines in output across the sectors. The impulse responses are provided in Figure 24 in Appendix C. The upper row depicts the shocks in the left panel as well as the path of the nominal interest rate in the model and in the data in the middle panel. The right panel shows that these shocks alone are unable to match the paths of consumption across the sectors. The bottom row depicts the sectoral and aggregate responses of inflation in its middle panel. The responses of inflation are qualitatively correct with the largest response in the goods sector due to its more flexible prices. Quantitatively, they are about half the size relative to the data. It is important to note, however, that this analysis does not consider supply shocks that occurred later on in 2022, which these other studies do.

6.2 Model Phillips Correlations

This study was motivated in part by the sectoral Phillips correlations depicted in Figure 1. These moments are reproduced for the model, and displayed in Figure 15, by constructing simulated data around the steady states that are calibrated to the 1960 and 2020 according to Table 5.[63] Qualitatively, the model Phillips correlations across time and sectors mimic the empirical ones. Quantitatively, the flattening in these slopes are of 33% and 16% for the services and goods sectors, respectively, which when aggregate imply an approximately 26% flattening of the aggregate Phillips correlation due to the rise in capacity buffers alone. This amount of flattening is less than that depicted in Figure 1. However, Hazell et al. (2022) find that a large part of this decline in correlation is due to changing inflation expectations. After controlling for it,

---

[63] Only monetary and fiscal shocks are considered in the simulations. The variance of the shocks are both set to 0.00002 so that the maximum deviations in output from the steady state become just under 10 percent. The steady state inflation rate is also set to equal the average inflation rate between 1960 and 2020, while it is set to the inflation target of two percent for the 2020 period.

the Phillips curve flattens by a factor of two, much closer to the decline that can be attributed to the larger capacity buffers.

The data for the COVID-19 period is taken directly from computed impulse responses to the calibrated shocks in the baseline analysis of this section. This implies the assumption that the demand shift and capacity restrictions were the only shocks to hit the economy during this time period. Despite this reduction in complexity, these two shocks alone are able to reproduce the Phillips correlations in the data, with the goods sector correlation becoming much steeper, while the services sector remains flat. This in line with the earlier findings of this section that these two shocks were the primary drivers of inflation during the pandemic.

7 Conclusion

Firms choose to operate with excess capacity in order to buffer against fluctuations in demand, and the size of firms' capacity buffers impact their investment and pricing decisions. The fact that firms' capacity buffers have gradually risen over the past half century has had a profound effect on the dynamics of inflation resulting in what is known as the flattening of the Phillips curve. Larger capacity buffers have also led to a rise in the cross-sectional variance of value added and that fiscal multipliers should have gradually become larger. One hypothesis which can quantitatively account for the rise in size of firms' capacity buffers as well as the fall in the labor shares across industries is the secular rise in markups.

The inflationary surge during the COVID-19 pandemic is also better understood when viewed through the lens of firms' capacity buffers: the pandemic led to temporary reductions in firms production capacities while households sharply increased their consumption of goods, which overwhelmed the capacity buffers in that sector. These initial two shocks are able to account for a substantial part of the rise and fall in inflation that followed. This paper does not argue that loose monetary and fiscal policy played no role in the high levels of aggregate demand and inflation following the pandemic. Nonetheless, it does suggest that they were not the root causes of the inflationary surge. The fact that expansionary monetary policy is not necessary to account for the majority of the pandemic inflation is consistent with the fact that inflation has been coming down despite continued strong demand and tight labor markets. It suggests that the Federal Reserve did not much play a large role in creating the inflation nor banishing it. Rather, the rise and fall in inflation was a consequence of the health precautions that followed from the emergence of COVID-19.

It is worth stressing the importance that the sectoral makeup, and change in it, had on the inflationary dynamics of the economy. Because prices in the services sector are inherently stickier than in the goods sector, the gradual shift towards consumption of services in the place of goods over the past half century has itself led to a marked flattening of the Phillips curve. The relative stickiness of prices across sectors also greatly exacerbated the surge in inflation during the COVID-19 pandemic because the goods sector, which has the more flexible prices, was the one that experienced the large increase in demand.

The findings from this study indicate that capacity buffers play a central role in determining macroeconomic dynamics. Hitherto, they have not received much attention because firms' production capabilities tend to be modeled as sufficiently flexible to accommodate any level of demand. There remains much to be uncovered regarding the structural implications of production capacity buffers. While this study mainly focused on the implications for inflation, capacity buffers also affect the choices of inventories and investment. This is the topic of ongoing work.

Finally, there are two current deficiencies in the US economic data on capacity buffers. The first is the mismeasurement of firms' effective capacity buffers during the COVID-19 pandemic, which is readily apparent when inspecting the questionnaire responses to the Quarterly Survey of Plant Capacity. The second is that they are only measured for the industrial production sector. Considering their structural importance, it would be beneficial if data on them were to be gathered for all parts of the US economy as well as amended to account for temporary reductions in capacity initially induced by the pandemic health precautions.

References

Altig, D., Christiano, L. J., Eichenbaum, M., and Lindé, J. (2011). Firm-specific capital, nominal rigidities and the business cycle. *Review of Economic Dynamics*, 14(2):225–247.

Alvarez-Lois, P. P. (2004). Capacity constraints, idiosyncratic demand uncertainty and the dynamics of inflation. *Economics Letters*, 83(1):15–21.

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

Atalay, E., Frost, E., Sorensen, A. T., Sullivan, C. J., and Zhu, W. (2023). Scalable demand and markups.

Auer, R., Borio, C. E., and Filardo, A. J. (2017). The globalisation of inflation: the growing importance of global value chains.

Auerbach, A. J. and Gorodnichenko, Y. (2012). Fiscal multipliers in recession and expansion. pages 63–98.

Balleer, A. and Noeller, M. (2023). Monetary policy in the presence of supply constraints: Evidence from german firm-level data.

Bansak, C., Morin, N., and Starr, M. (2007). Technology, capital spending, and capacity utilization. *Economic Inquiry*, 45(3):631–645.

Baqaee, D., Farhi, E., and Sangani, K. (2021). The supply-side effects of monetary policy.

Benigno, P. and Eggertsson, G. B. (2023). It's baaack: The surge in inflation in the 2020s and the return of the non-linear phillips curve.

Bianchi, F. and Civelli, A. (2015). Globalization and inflation: Evidence from a time-varying var. *Review of Economic Dynamics*, 18(2):406–433.

Boehm, C. E. and Pandalai-Nayar, N. (2022). Convex supply curves. *American Economic Review*, 112(12):3941–3969.

Boudette, N. E. (2020). Autoworkers are returning as carmakers gradually crank up factories. *The New York Times*.

Brinca, P., Duarte, J. B., and Faria-e Castro, M. (2021). Measuring labor supply and demand shocks during covid-19. *European Economic Review*, 139:103901.

Caselli, F. and Coleman, W. J. (2006). The world technology frontier. *American Economic Review*, 96(3):499–522.

Cerrato, A. and Gitti, G. (2022). Inflation Since COVID: Demand or Supply. *Graduate Student Bravo Working Paper*.

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

Clayton, C., Jaravel, X., and Schaab, A. (2018). Heterogeneous price rigidities and monetary policy. *Available at SSRN 3186438*.

Coibion, O. (2012). Are the effects of monetary policy shocks big or small? *American Economic Journal: Macroeconomics*, 4(2):1–32.

Comin, D. and Philippon, T. (2005). The rise in firm-level volatility: Causes and consequences. *NBER macroeconomics annual*, 20:167–201.

Comin, D. A., Johnson, R. C., and Jones, C. J. (2023). Supply chain constraints and inflation.

Cotton, C. D. and Garga, V. (2022). The role of industrial composition in driving the frequency of price change. Technical report, Federal Reserve Bank of Boston.

Crouzet, N. and Eberly, J. (2020). Rents and intangible capital: A q+ framework. *The Journal of Finance*.

Crouzet, N. and Eberly, J. (2021). Intangibles, markups, and the measurement of productivity growth. *Journal of Monetary Economics*, 124:S92–S109.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644.

Del Negro, M., Lenza, M., Primiceri, G. E., and Tambalotti, A. (2020). What's up with the phillips curve?

Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics*, 80(4):549–560.

Fagnart, J.-F., Licandro, O., and Portier, F. (1999). Firm heterogeneity, capacity utilization, and the business cycle. *Review of Economic Dynamics*, 2(2):433–455.

Fagnart, J.-F., Licandro, O., and Sneessens, H. R. (1997). Capacity utilization and market power. *Journal of Economic Dynamics and Control*, 22(1):123–140.

Fazzari, S. M., Morley, J., and Panovska, I. (2015). State-dependent effects of fiscal policy. *Studies in Nonlinear Dynamics & Econometrics*, 19(3):285–315.

Foster, L. S., Haltiwanger, J. C., and Tuttle, C. (2022). Rising markups or changing technology? Technical report, National Bureau of Economic Research.

Fujiwara, I. and Matsuyama, K. (2022). Competition and the phillips curve.

Gagliardone, L. and Gertler, M. (2023). Oil prices, monetary policy and inflation surges. Technical report, National Bureau of Economic Research.

Gahn, S. J. (2023). Towards an explanation of a declining trend in capacity utilisation in the us economy: Analysing the nber-ces output–capital ratio. *Review of Political Economy*, 0(0):1–22.

Gilchrist, S. and Williams, J. C. (2000). Putty-clay and investment: a business cycle analysis. *Journal of political Economy*, 108(5):928–960.

Gilchrist, S. and Williams, J. C. (2005). Investment, capacity, and uncertainty: a putty–clay approach. *Review of Economic Dynamics*, 8(1):1–27.

Gitti, G. (2023). Nonlinearities in the regional phillips curve with labor market tightness.

Gonçalves, S., Herrera, A. M., Kilian, L., and Pesavento, E. (2023). State-dependent local projections.

Gormsen, N. J. and Huber, K. (2023). Corporate discount rates.

Granger, C. W. and Terasvirta, T. (1993). Modelling non-linear economic relationships. *OUP Catalogue*.

Greenwood, J., Hercowitz, Z., and Huffman, G. W. (1988). Investment, capacity utilization, and the real business cycle. *The American Economic Review*, 78(3):402–417.

Gutiérrez, G. and Philippon, T. (2017). Declining competition and investment in the us.

Harding, M., Lindé, J., and Trabandt, M. (2023). Understanding post-covid inflation dynamics. *Journal of Monetary Economics*.

Hazell, J., Herreno, J., Nakamura, E., and Steinsson, J. (2022). The slope of the phillips curve: evidence from us states. *The Quarterly Journal of Economics*, 137(3):1299–1344.

Hobijn, B., Miles, R., Royal, J., and Zhang, J. (2023). The Recent Steepening of Phillips Curves. *Chicago Fed Letter*, 0.

Hooper, P., Mishkin, F. S., and Sufi, A. (2020). Prospects for inflation in a high pressure economy: Is the phillips curve dead or is it just hibernating? *Research in Economics*, 74(1):26–62.

Ihrig, J., Kamin, S. B., Lindner, D., and Marquez, J. (2010). Some simple tests of the globalization and inflation hypothesis. *International Finance*, 13(3):343–375.

Israeli, A., Scott-Morton, F., Silva-Risso, J., and Zettelmeyer, F. (2022). How market power affects dynamic pricing: Evidence from inventory fluctuations at car dealerships. *Management Science*, 68(2):895–916.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, 95(1):161–182.

Kuhn, F. and George, C. (2019). Business cycle implications of capacity constraints under demand shocks. *Review of Economic Dynamics*, 32:94–121.

León-Ledesma, M. A. and Satchi, M. (2019). Appropriate technology and balanced growth. *The Review of Economic Studies*, 86(2):807–835.

Malinvaud, E. (1977). *The theory of unemployment reconsidered*.

Marto, R. (2023). Structural change and the rise in markups.

McLeay, M. and Tenreyro, S. (2020). Optimal inflation and the identification of the phillips curve. *NBER Macroeconomics Annual*, 34(1):199–255.

Nakamura, E. and Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics*, 123(4):1415–1464.

Otterman, S. (2021). New york is reopening. here's what that actually means. *The New York Times*.

Pierce, J. and Wisniewski, E. (2018). Some characteristics of the decline in manufacturing capacity utilization.

Pritzker, J. (2020). Illinois executive order 2020-32.

Ramey, V. A. and Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of political economy*, 126(2):850–901.

Ravn, M. O. and Uhlig, H. (2002). On adjusting the hodrick-prescott filter for the frequency of observations. *Review of economics and statistics*, 84(2):371–376.

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American economic review*, 94(4):1055–1084.

Rubbo, E. (2023). Networks, phillips curves, and monetary policy. *Econometrica*, 91(4):1417–1455.

Schmitt-Grohé, S. and Uribe, M. (2022). Heterogeneous downward nominal wage rigidity: Foundations of a static wage phillips curve. Technical report, National Bureau of Economic Research.

Shapiro, A. H. et al. (2022). How much do supply and demand drive inflation? *FRBSF Economic Letter*, 15:1–6.

Shapiro, M. D., Gordon, R. J., and Summers, L. H. (1989). Assessing the federal reserve's measures of capacity and utilization. *Brookings Papers on Economic Activity*, 1989(1):181–241.

Sheremirov, V. (2022). Are the demand and supply channels of inflation persistent? evidence from a novel decomposition of pce inflation. *Federal Reserve Bank of Boston Research Paper Series Current Policy Perspectives Paper*, (94983).

Stock, J. H. and Watson, M. W. (2020). Slack and cyclically sensitive inflation. *Journal of Money, Credit and Banking*, 52(S2):393–428.

Sun, T. (2023). Excess capacity and demand driven business cycles.

Tenreyro, S. and Thwaites, G. (2016). Pushing on a string: Us monetary policy is less powerful in recessions. *American Economic Journal: Macroeconomics*, 8(4):43–74.

Whitehead, D. and Kim, Y. H. B. (2022). The impact of covid 19 on the meat supply chain in the usa: A review. *Food Science of Animal Resources*, 42(5):762.

Wieland, J. F. and Yang, M.-J. (2020). Financial dampening. *Journal of Money, Credit and Banking*, 52(1):79–113.

APPENDIX TO CAPACITY BUFFERS

A Household Side of Model

Households and Employment Agencies The economy is populated by a unit mass of households indexed by $h \in [0,1]$ who are identical except for their differentiated labor supply. They set their own wages one period in advance subject to Rotemberg adjustment costs in utility. Given their preset wages, labor is supplied to meet its demand. The demand for labor comes from employment agencies that bundle the differentiated household labor supplies into a CES composite,

$$L_t = \left(\int_0^1 {l_t^h}^{\frac{\varepsilon_w - 1}{\varepsilon_w}} \, dh \right)^{\frac{\varepsilon_w}{\varepsilon_w - 1}} . \tag{44}$$

They supply the labor bundle on to firms in the goods and services sectors for an aggregate amount of labor supplied $L_t = l_t^s + l_t^g$. The employment agencies are perfectly competitive and earn zero profits. Conditional on hiring labor type l^h for a wage rate w^h, the demand for it is given by

$$l_t^h = L_t \left(\frac{w_t^h}{W_t} \right)^{-\varepsilon_w} \tag{45}$$

where W_t is the wage rate for the bundle of labor supplied to firms,

$$W_t = \left(\int_0^1 {w_t^h}^{1-\varepsilon_w} \, dh \right)^{\frac{1}{1-\varepsilon_w}} . \tag{46}$$

The households' sectoral and firm level expenditures in the face of quantity constraints are characterized in the main body of the paper. The problem of an equivalent individual household with an aggregate consumption decision is considered here,

$$V(w_t^h, b_t^h) = \max_{C_t^h, b_{t+1}^h, w_{t+1}^h} \ln C_t^h - \psi \frac{{l_t^h}^{(1+\varphi)}}{1+\varphi} - \frac{\chi_w}{2} \left(\frac{w_{t+1}^h}{w_t^h} - 1 - \pi_t^W \right)^2 \frac{W_t}{P_t} L_t + \beta V(w_{t+1}, b_{t+1}) \tag{47}$$

$$\text{s.t.} \qquad \mathcal{P}_t C_t^h + b_{t+1}^h = w_t^h l_t^h + (1 + r_t) b_t^h + T_t \tag{48}$$

$$l_t^h = L_t \left(\frac{w_t^h}{W_t} \right)^{-\varepsilon_w} , \tag{49}$$

where $\mathcal{P}_t$ is provided in (34). With a constant EIS of unity, they smooth consumption C_t^h over time using a one-period bond b_t which pays interest r_t. They incur a disutility scaled by ψ from supplying labor with a inverse Frish elasticity of φ. χ_w governs the strength of the wage adjustment costs which are proportional to aggregate real labor income. Households geometrically discount the future according to β. It sets its choices for consumption, savings, and wage rates while satisfying its budget constraint and the demand for is labor. T_t encompasses the lumpsum transfers including payments for government expenditures and receipts from firm's profits.

B Alternative Explanations for Rising capacity Buffers

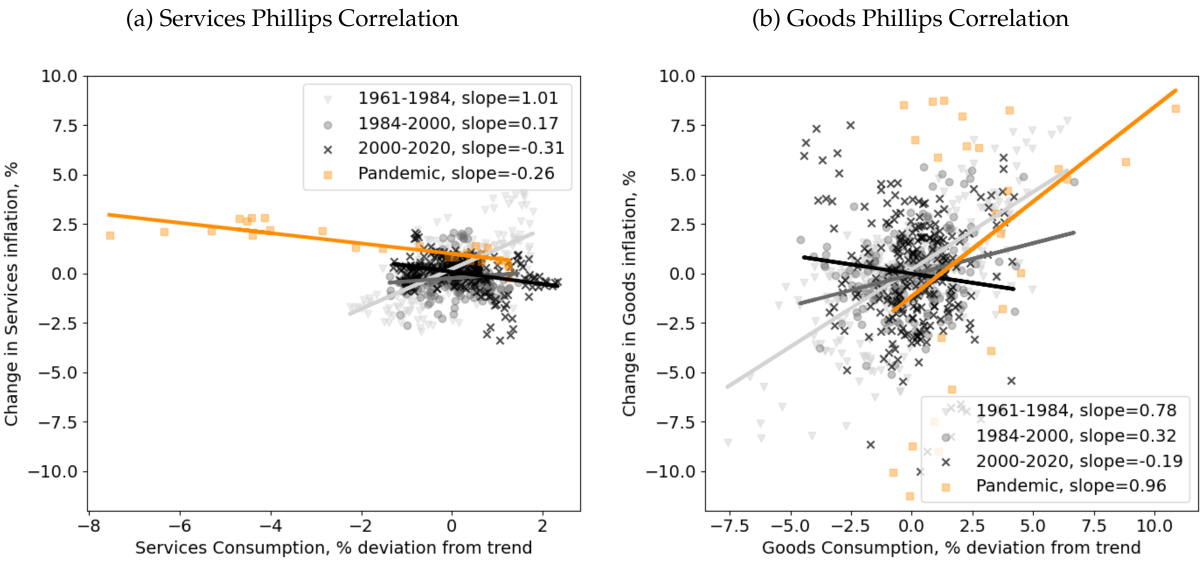
There are a couple alternative mechanisms which are worth mentioning as they may also play some role in inducing firms to operate with larger capacity buffers. They are briefly discussed here. Each of them is inconsistent with the data in at least one dimension, however. That does not rule them out necessarily. It means that they alone are not able to explain all of the changes in the listed moments (at least according to this theory) and therefore some other structural changes must also be occurring with the mechanisms working in synergy to explain the facts. Such multiple mechanism hypotheses are not considered in this paper.

The Decline in the Cost of Capital An alternative mechanism which could also induce firms to increase the size of their capacity buffer and thereby flatten the Phillips curve in this framework is the decline in the risk free rate. From Figure 6 it can be seen that lowering the risk-free rate would appear as lowering of the horizontal black line in the left panel. This would induce firms to accumulate larger capacity buffers, since their cost has decreased, which would then also result in flattening of the Phillips curve. Though this may have contributed to its flattening since 1980 since real rates have been declining since then, real rates also used to be low preceding this time period implying that firms should have had larger capacity buffers and a flatter Phillips curve back in the 1960s, which was not the case. Moreover, Gormsen and Huber (2023) show that firms with more market power have not lowered their discount rates (their hurdle rates for investment) despite the cost of capital coming down for them, which would mean that this real reduction in the cost of capital has occurred, but it has not been taken advantage of by firms, especially for those with more market power, as they document.

Rising Idiosyncratic Demand Uncertainty Rising demand uncertainty would lead to firms accumulating larger capacity buffers as a precaution, and it would also be consistent with the rise in sales variances. However, the larger capacity buffers alone would lead to fall in markups and therefore an increase in the labor share in contrast to the evidence put forth in Figure 5. It is plausible that rising demand uncertainty along with some other mechanism such as automation have together brought about the changes. However, one would not expect to find the positive correlation between the change in labor shares and utilization rates depicted in Figure 5.

C Additional Figures

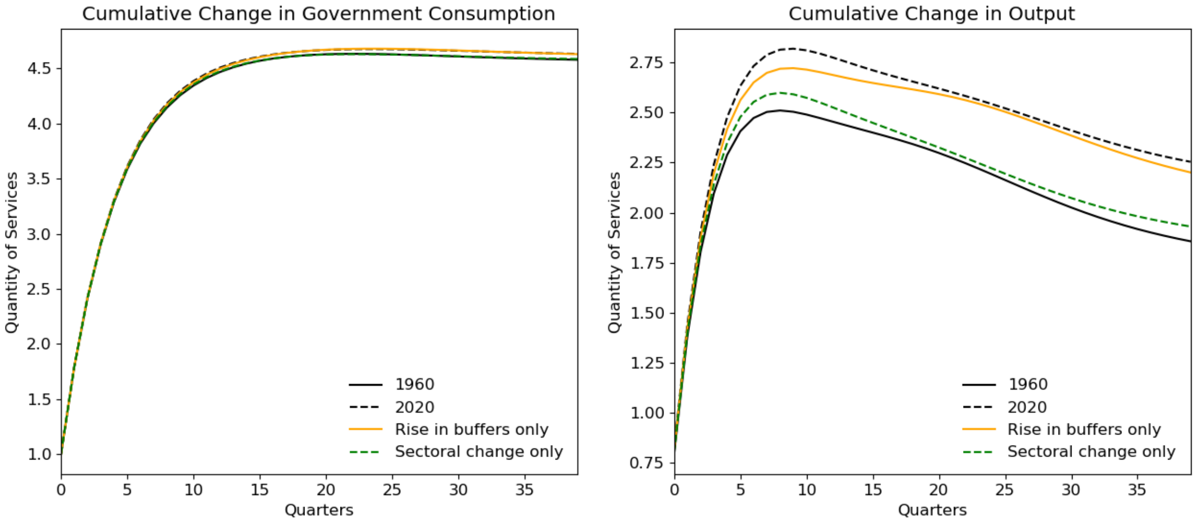
Figure 16: The Changing Slope of the Accelerationist Phillips Curve


(a) Services Phillips Correlation

(b) Goods Phillips Correlation

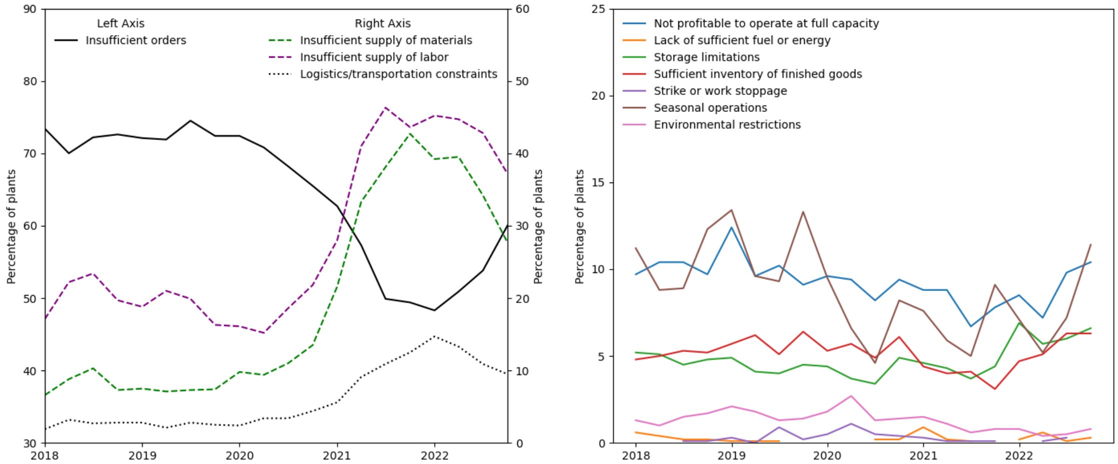
Notes: The panels depict estimates of the NKPCs for the services and goods sectors under the assumption that inflation expectations are equal to the previous periods inflation rate. The trend is calculated using the Hodrick-Prescott filter parameterized for monthly data as described in Ravn and Uhlig (2002), and the inflation rate is calculated as the growth rate in the price-level over the following year.

Figure 17: Cumulative Response in Services Output to a Fiscal Shock



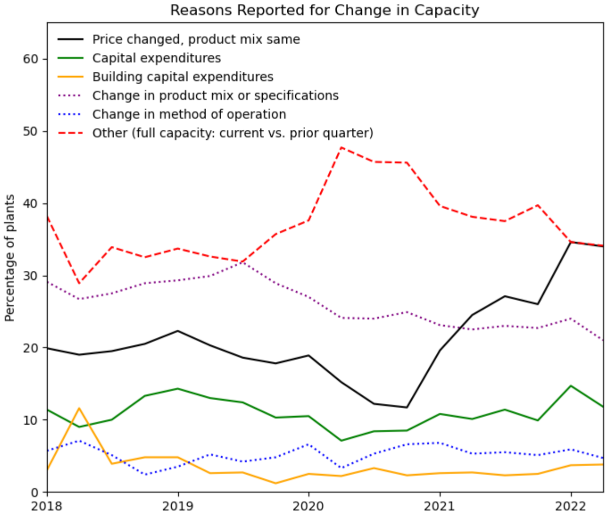
Notes: Model cumulative impulse responses for the quantity of Services produced in the right panel and consumed by the government in the left to a fiscal shock. Impulses are provided for various sets of parameter values for ε_p and θ according to the their values in 1960 and 2020. The "Rise in buffers only" curve keeps the sectoral consumption share fixed at the 1960 level. The "Sectoral change only" keeps the capacity buffers fixed at its 1960 level.

Figure 18: Other Reasons for Operating Below Capacity



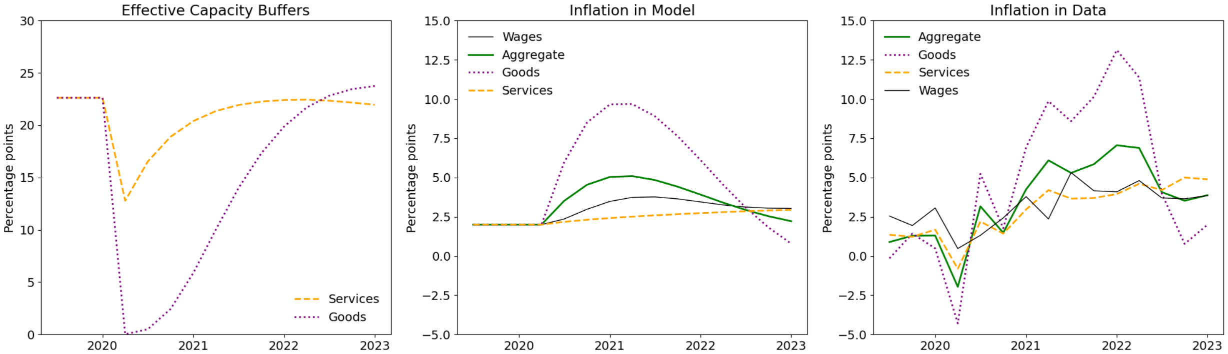
Notes: Survey responses for other less common reasons that manufacturing plants indicate for not operating at capacity and the percentage of plants indicating them in the census' quarterly survey of plant capacity.

Figure 19: Reasons for a Change Capacity



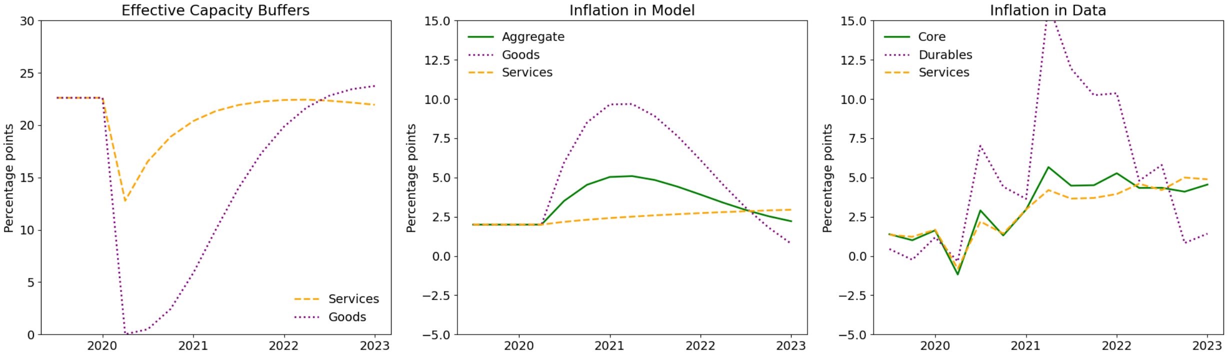
Notes: Survey responses for reasons that manufacturing plants indicate for why their reported level of capacity changed since last quarter in the census' quarterly survey of plant capacity.

Figure 20: Wage Impulse Responses to the Demand Shift and Capacity Shocks



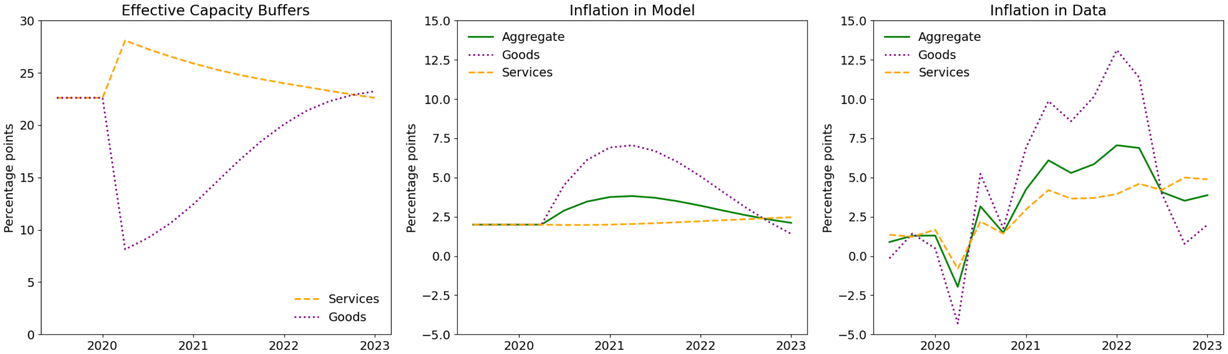
Notes: Full nonlinear impulse responses in model to the capacity and demand shift shocks while keeping the monetary policy fixed at its steady state for eight quarters.

Figure 21: Impulse Responses to the Demand Shift and Capacity Shocks



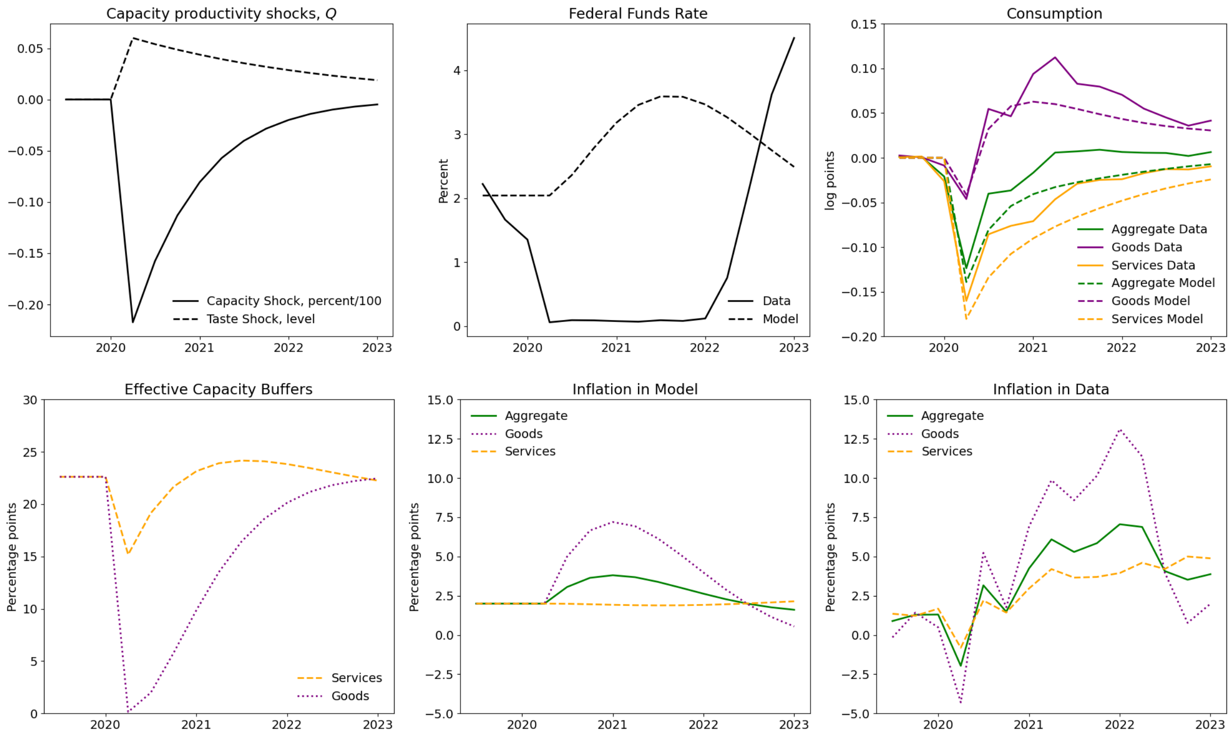
Notes: Full nonlinear impulse responses in model to the capacity and demand shift shocks while keeping the monetary policy fixed at its steady state for eight quarters.

Figure 22: Impulse Responses to Demand Shift Shock only



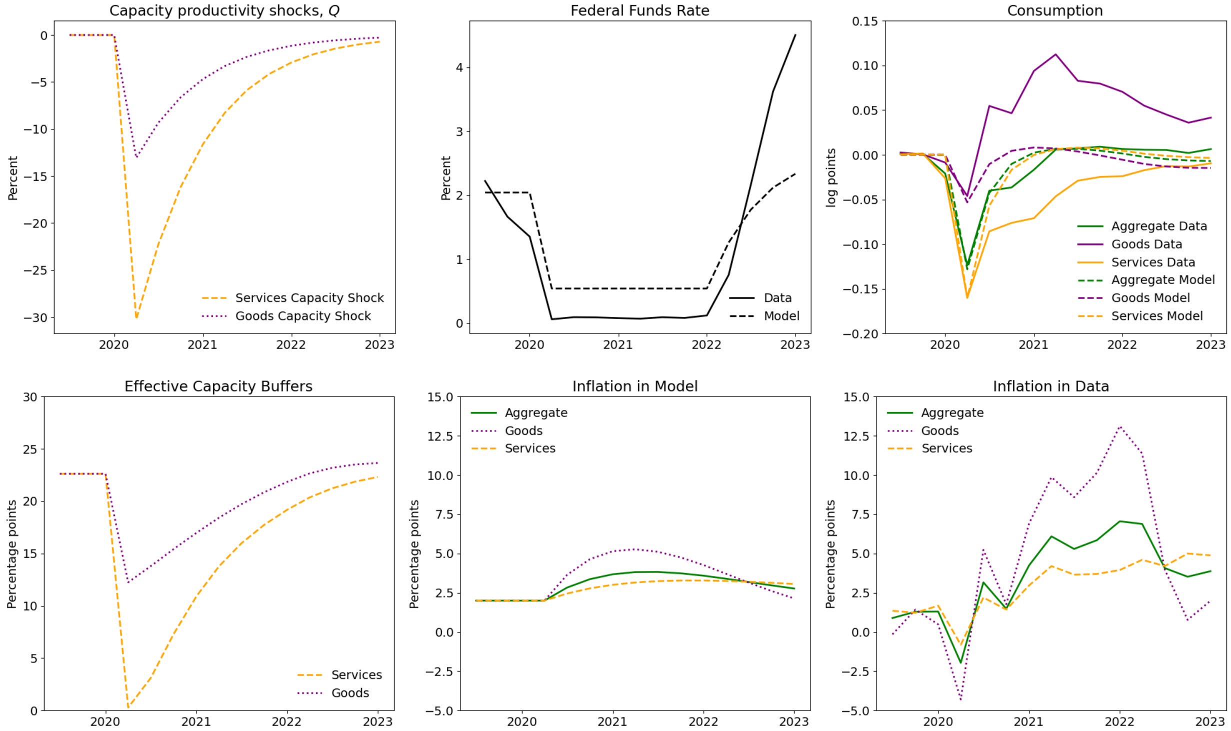
Notes: Full nonlinear impulse responses in model to the demand shift shock while keeping the monetary policy fixed at its steady state for 8 quarters.

Figure 23: Paths of Monetary Policy and Shocks under Active Taylor Rule



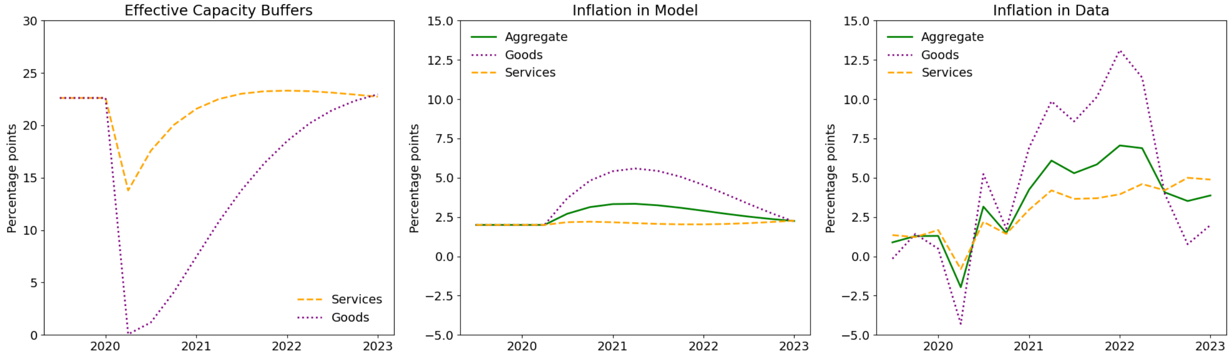
Notes: Full nonlinear impulse responses in model for the capacity and demand shift shocks along with the empirical and model implied paths of consumption and inflation while the interest rate is allowed to behave according to the specified Taylor Rule.

Figure 24: Calibration of Monetary Policy and Capacity Shocks along with Quantity Responses



Notes: Full nonlinear impulse responses in model to the capacity shocks while the interest rate is fixed at a level 1.5 percentage points below its steady state for 8 quarters. Dynamics paths of productivity shocks and monetary policy along with the empirical and model implied paths of consumption used for calibration.

Figure 25: Impulse Responses without Differential Price Rigidities



Notes: Full nonlinear impulse responses in model to the capacity and demand shift shocks while keeping the monetary policy fixed at its steady state for 8 quarters, but without differential price rigidities across sectors.

D Detailed Exposition of STLPM Analysis of Monetary Policy Shocks

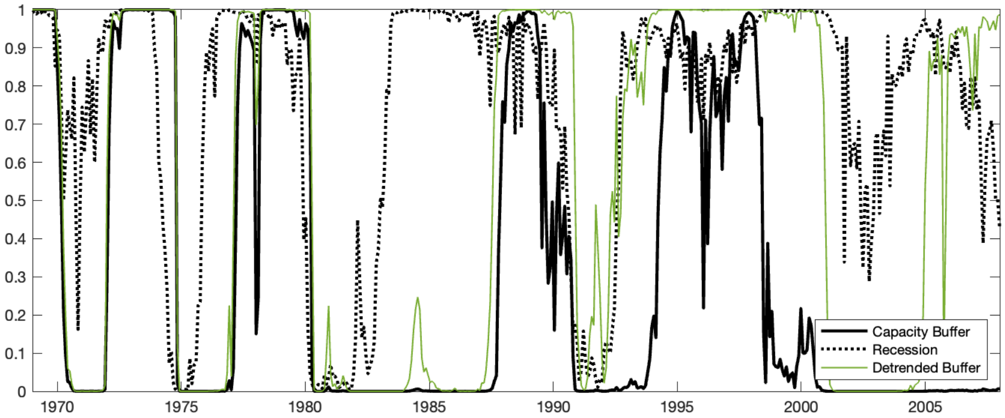
Firm's capacity buffer are an important state variable for the dynamics of prices and investment. Due to the differing incentives it has when capacity constrained than when it is not. When capacity buffers are narrow, either by design or due to unforeseen events such as a strong expansion or pandemic induced production restrictions, pricing and investment should be more sensitive to variations in demand compared to when capacity buffers are large. In fact, when firm capacity buffers are sufficiently large, variation in demand should have a negligible effect on the firm's prices and investment as suggested by Figure 4.

In order to test the potency of this state-dependency for the aggregate US economy, the effect of monetary policy shocks are estimated dependent on the size of firms' capacity buffers. This is done using a Logit Smooth Transition Model, first proposed by Granger and Terasvirta (1993), which estimates a nonlinear model by allowing for the variables to depend on a convex combination of two linear regimes. A state variable determines the convex combination, i.e. the degree to which one regime is dominating at the time. In this analysis the state will be determined by the size of the capacity buffer.

The previous studies that have used a logit smooth transition local projection model did it with the purpose to explore whether the economy might respond differently to shocks in a recession versus an expansion. Auerbach and Gorodnichenko (2012) and Ramey and Zubairy (2018) applied it to study the impact of fiscal stimulus. This paper follows in the footsteps of Tenreyro and Thwaites (2016) to analyse the impact of monetary policy. All three of these papers used a seven quarter moving average of output growth as their indicator for whether the economy is in a recession or not with it adjusted so that the economy is in a recession on average 20 percent of the time. In contrast, having the size of the capacity buffers determine the state is explicitly motivated by the theory in section 3 with firms using a putty-clay production technology and facing demand uncertainty. Moreover, due to strong nonlinearity that arises near capacity as depicted in Figure 4, the economy spends the majority of its time in the slack state where prices and investment are insensitive to variation in demand, which is the opposite of the previous studies where the majority of the time is then spent in the expansionary phase. The contrast in capacity-dependent versus recession dependent states is shown in Figure 26 by the black solid and dotted lines, respectively. While there is comovement between the two state-variables, they diverge in two main ways. One is that the capacity-dependent state spends the majority of its time in the "low" regime (with a value of 0) with large capacity buffers while the opposite is true for the recession-dependent state which spends most of its time in its "high" regime (with a value of 1) of expansion, as discussed above. The other is while fluctuations between the high and low regimes is more stationary throughout the time period for the recession-based state, there is marked downward trend in the capacity dependent one with it spending less time on average in the tight capacity buffer state towards the end of the sample.

The state-dependent impulse responses from monetary policy shocks are estimate using the Jordà (2005) method which accounts for the nonlinear specification by directly projecting the identified monetary policy

Figure 26: Comparison of Capacity Buffer versus Recessionary Based States



Notes: Time series depicting which regime or combination of regimes the economy finds itself in with the pure regimes being defined by 1 and 0 on the y-axis. The capacity buffer state comes from capacity utilization rates associated with industrial production with 99% of the regime switch occuring between of 15 and 19 percent. The recession dependent state takes a 21 month moving average of the growth rate in real personal consumption expenditures. The series is standardized, set to be in a recession 20 percent of the time, and the rapidity to which is switches regimes is parameterized by $\theta = 3$. The detrended capacity buffer state is estimated by removing the linear trend in the capacity utilization series, and using the same transition width of 4 percent around the now mean zero series.

shocks on to future values of the dependent variables in a local projection model. Beyond the standard concerns regarding mispecification of the Structural VAR model, which are likely be exasperated in a smooth transition setup, calculating and conducting inference on their nonlinear impulse responses is not straightforward. The impulses depend on the initial state when the shock occurs as well as by how much the regime transitions along the impulse. On the contrary, as discerned by Auerbach and Gorodnichenko (2012), the local projection method should incorporate the average transition that occurs as a result of the monetary policy shock into the estimated impulse responses as given by the coefficients in the nonlinear framework. The local projection method also enables straightforward by use of standard asymptotic theory to conduct inference.

D.1 Econometric Specification

The logit smooth transition local projection model applied here uses the same specification as that in Tenreyro and Thwaites (2016) which is given by

$$y_{t+h} = \tau t + F(z_t)\left(\alpha_1^h + \beta_1^h m_t + \gamma_1' x_t\right) + (1 - F(z_t))\left(\alpha_0^h + \beta_0^h m_t + \gamma_0' x_t\right) + u_t. \tag{50}$$

The monetary policy shock in period t is denoted by m_t, the impact it has on the outcome variable y_{t+h} that is h periods in advance is given by the coefficients β_0^h for the regime with large capacity buffers and β_1^h for the regime with small capacity buffers. A trend, regime-dependent intercepts, and controls x_t are also included as regressors, where the controls include lagged values of the outcome variable and the federal funds rate. The state transition function $F(z_t)$ is takes on the logit specification,

$$F(z_t) = \frac{\exp(\theta z_t)}{1 + \exp(\theta z_t)}, \tag{51}$$

where z_t is the state variable. In the baseline specification, z_t is simply the size of the capacity buffer relative to some threshold $\bar{B}$. The parameter θ specifies the width of the transition space between regimes around the threshold. In line with all three previous papers that have applied this methodology, the state variable and parameters associated with it are preset by the econometrician. This reduces the task to a linear estimation problem. The coefficients in (50) are estimated at horizons zero to H as a set of seemingly unrelated regression equations, which provides the covariances of parameter estimates between the equations. Standard errors can then be computed following Driscoll and Kraay (1998) to account for the fact that residuals will be correlated across dates and estimation horizons.

Data and Calibration The monetary policy shocks used are the narrative shocks of Romer and Romer (2004) which have since been updated and extended by Wieland and Yang (2020). The shocks are derived by using the Federal Reserves historical records leading up to the time of FOMC meetings to derive the committee's intended changes to the federal funds rate. The intended changes are regressed on the Federal Reserve's contemporaneous forecasts for aggregate measures of the economy as well as lagged and concurrent measures. The residuals then capture the intended changes to the federal funds rate purified for any systematic changes in it based on the available information to the Federal Reserve at the time. In contrast to other state-of-the-art estimates of monetary policy shocks, this method covers at the monthly frequency the time period spanning all the way back to January 1969 and up to January 2008. The extent to which they expand back in time is crucial for this analysis due to the gradually increasing size of firms' capacity buffers. As shown in Figure 26, the last time firms' capacity buffers were narrow (excluding the pandemic), was in the 1990s and even then they were not overwhelmingly tight. An analysis which only covered more recent times would not feature enough time periods when firms were particularly capacity constrained.

All other data used is also available at the monthly frequency and taken from FRED. The measure of the state z_t, the size of total economy's capacity buffer, is the complement of the Federal Reserve's measure of total capacity utilization. The two main outcome variables, y_t, are monthly measures of firm investment and prices. Industrial production of business equipment is used as the measure for investment, and the measure of consumer prices that is most relevant for the industrial production sector is the goods subcategory of personal consumption expenditures. In addition, the quantity and price indices of personal consumption expenditures, and industrial production are used as broader measures of the economy. Quantity measures are put in log levels, prices are put in log differences, and the capacity utilization and the federal funds rates are in percent.

The forecast horizon H is set to 48 months, and the allowed lag length for the autocorrelation in the

Driscoll-Kraay standard errors is set to $H + 1$. According to the Akaike Information Criterion, the controls include 4 lags of the dependent variable and 1 lag of the federal funds rate. An additional benefit of having the state variable grounded in theory of capacity buffers is that an empirical measure of it can be used directly, and it can guide the parameterization associated with it. The regime switch is centered around a capacity buffer size of $\bar{B} = 17$ percent such that $z_t = \bar{B} - B_t$. This level was chosen because that was near the peak size of the buffer during the 1990s, a time when the Phillips curve was known to have already flattened significantly, but hadn't completely disappeared yet.[64] As such, this size of the capacity buffer should feature prominently in the transition between regimes. This level also implies that the economy spent nearly 70 percent of its time closer to the regime with large capacity buffers. θ is set so that 99 percent of the transition happens within 4 percentage points around $\bar{B}$.

Is the increasing size of capacity buffers driving the results? One concern may be the positive trend in the size of capacity buffers. This trend is at the heart of the hypothesis in section 4 explaining the flattening of the Phillips curve amongst else, and therefore should be included in the baseline analysis as it should be contributing to the differential response to monetary policy shocks.[65] However, if the monetary policy shocks used to be more effective for reasons other than the mechanism put forth here, the baseline analysis may be partly attributing this alternative mechanism, whatever it may be, to the trend in the size of capacity buffers. The initial analysis is therefore repeated but after removing the trend in the capacity buffer size thereby only using variation in capacity buffer size that is attributable to the business cycle. The results should be worse than the baseline analysis, but given the high variability in the size of firms capacity buffers, it should still be able to account for much of it. The results are given in Figure 29. When firm's are operating closer to their capacity constraint, the impact on investment is initially negative and significant, but quickly becomes too imprecisely estimated. The response in the price-level is very similar but larger than in the baseline exercise with prices beginning fall immediately with a peak effect at -0.125 log points. Again, there is little response of investment and prices when capacity buffers are large.

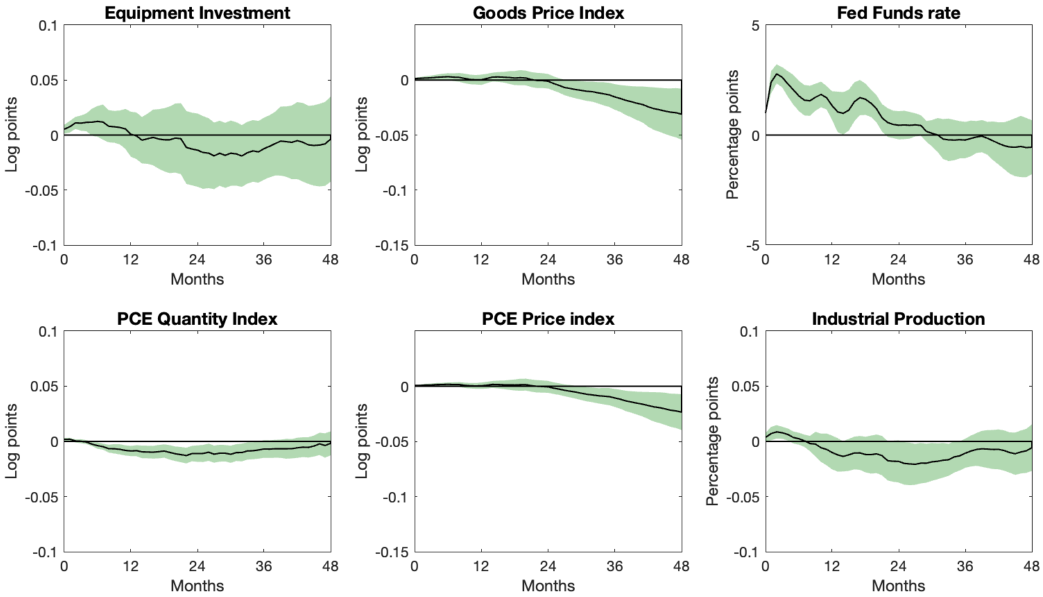
Alternatively, one could estimate the impulse responses before and after year 1990 which is at the middle of the sample. If the baseline results were driven by the trend in capacity buffer size, the responses in the pre-1990 should be substantial but mute in post-1990 sample. Figure 30 in depicts these estimated responses. After 1990 there is a negligible response to a monetary contraction, although they are very imprecisely estimated. On the other hand, The pre-1990 responses are precisely estimated, and they resemble the impulse responses estimated on the full sample without any state-dependencies. While these results do indicate that the effects of monetary policy were more pronounced before 1990, as one would expect from the increasing size of capacity buffers, the differential response just from this trend is small relative to that

---

[64] See figure 1 in Stock and Watson (2020) for instance.

[65] In a unrelated but similar vain, Coibion (2012) points out that the effects the Romer and Romer (2004) monetary policy shocks are sensitive to the inclusion of the Volcker Disinflation time-period around 1980. The monetary policy shocks affecting this period are mostly attributed to the tight capacity buffer state in Figure 26 exactly implying that there should be a larger effect from this period. Rather than being cause for concern, it is exactly consistent with the theory put forth here.

Figure 27: Regime Independent Estimated Impulse Responses



Notes: Estimated impulse responses to a contractionary one percent monetary policy shock in a linear local projection model. Black lines indicate the point estimates of the impulse response while green areas show the 95 percent confidence intervals calculated from the Driscoll-Kray standard errors.

from cyclical variability in the capacity buffers. To conclude, The trend in the size of capacity buffers is contributing to, but not driving the results, further confirming the actual size of firm capacity buffers as the correct state variable.

Estimated Impulse Responses in Linear Framework To use as a benchmark, Figure 27 provide the impulse responses to a contractionary one percent monetary policy shock in a local projection model excluding any capacity buffer dependencies. The response of investment is in the first column, the price level in the second column, and the federal funds rate in the third. In this case, the one percent contractionary monetary policy shock raises the federal funds rate to a peak of 2.8 percent after which it declines to its original level over the course of two and a half years. In response, there is no significant decline in investment. Output declines by 0.02 log points and consumption declines by 0.01 log points after two years. Goods and consumption prices only begin to decline after two years to reach -0.02 and -0.03 log points after four years, respectively.

Is the capacity buffer a proxy for whether the economy is in a recession, or vice-versa? As mentioned in the discussion surrounding Figure 26, the previous studies that have used the logit smooth transition local projection framework have done so while using a moving average of output growth as their state variable with it being a measure of whether the economy is in a recession or expansion. They did this to estimate the differential impact of fiscal or monetary policy when an economy is in a recession or an expansion, yet

Table 6: Relative response of investment to output across horizons

Horizon (months)	12	18	26	30	36
I/Y (Any B)	-0.77	0.34	0.71	0.94	1.17
I/Y ($B < 15\%$)	1.06	230.40	1.62	1.90	2.70

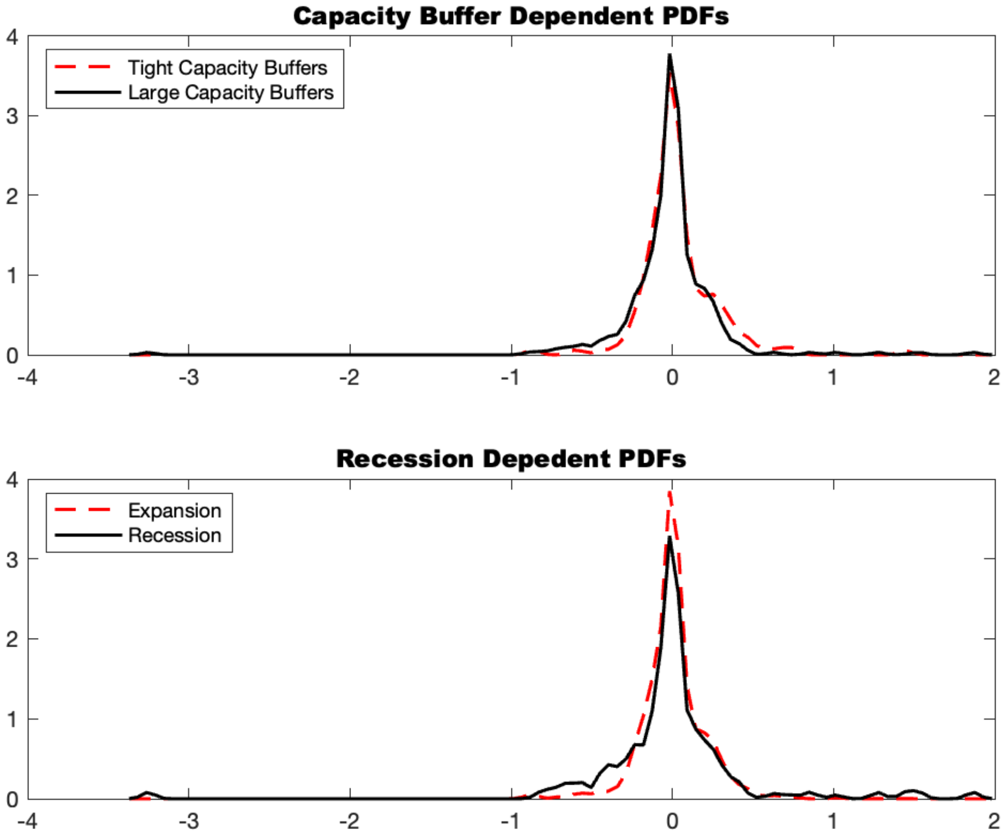
Notes: Ratios across horizons of the change in the investment level to the change in output for industrial production. The first row indicates the ratios for the unconditional estimated responses to monetary policy. The second row indicates the ratios for the estimated responses to monetary policy when capacity buffers are small (<15%).

they did not specify an underlying economic mechanism for why there should be a differential response. The size of firm's capacity buffers is one such mechanism, but one could imagine others that depend on labor market tightness for instance.

For a comparison, Figure 31 provides impulse responses under such a state-variable with it being the 21 month trailing moving average of consumption growth, the regime switch is set at the twentieth percentile such that it spends only 20 percent of its time in the recessionary state. Following Tenreyro and Thwaites (2016), the moving average of the growth rate is then standardized and θ set to 3. When in the expansionary regime, the response of investment is negligible and insignificant at all horizons, and the contractionary effect on the price level only begins to takes affect after two years rather than immediately. The weaker responses take place in spite of a more persistent and contractionary path of the federal funds rate. These weaker results suggest that the expansionary state may be a proxy for the capacity constrained regime with many of the capacity unconstrained periods incorrectly attributed to the expansionary regime. One reason they are so much weaker is that the recession-dependent state attributes most of time periods to being in the expansionary, less slack, state, thereby watering down the impulse reponses in this regime with observations which should be attributed to the more slack regime, which the capacity-dependent state conversely attributes to most of the time periods.

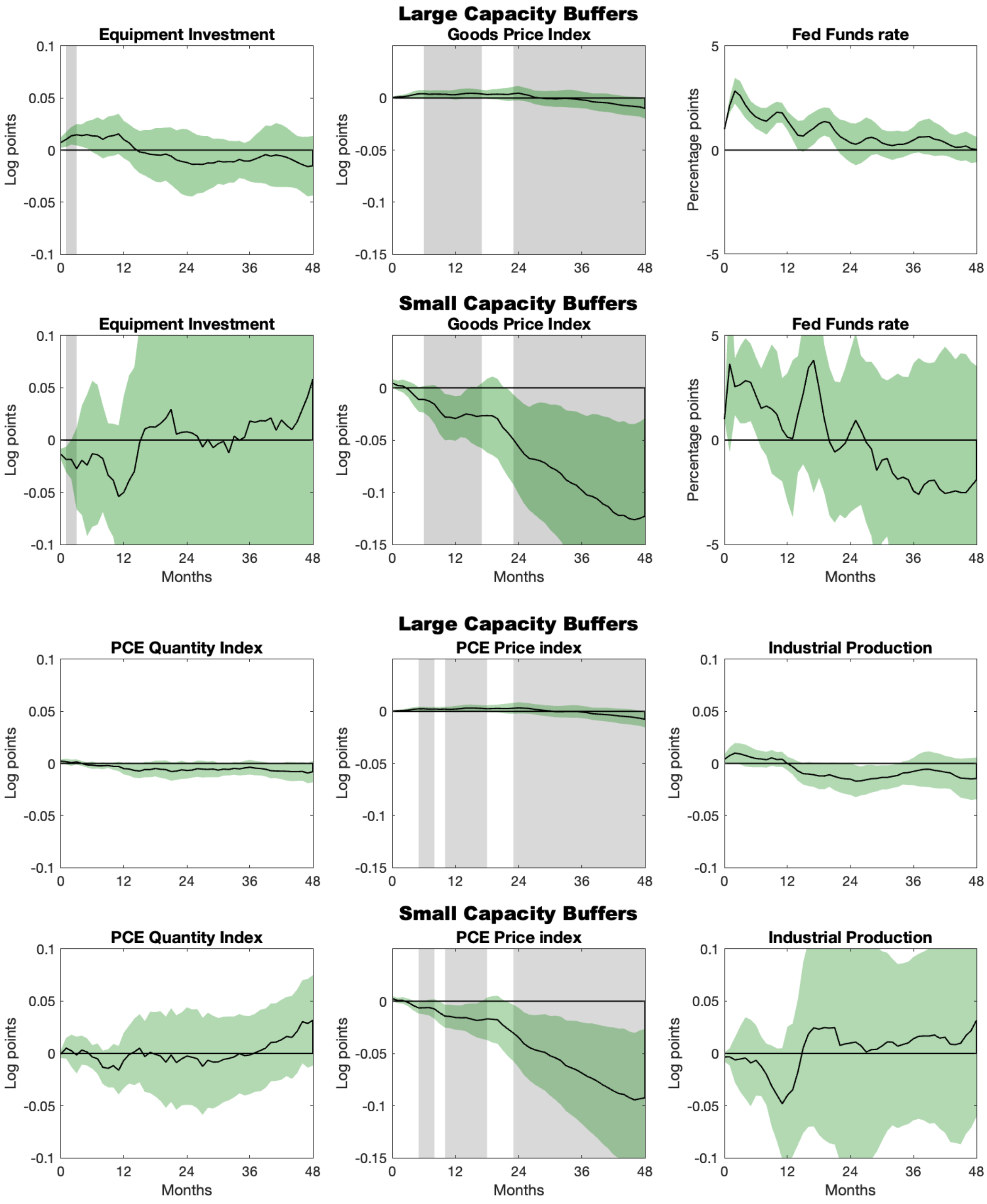
Are Monetary Shocks Systematically Different during periods with small capacity buffers? The top panel of Figure 28 shows that the distributions of monetary policy shocks are fairly similar when firms have large capacity buffers as when they are small. There is a slight skew with the lower tail being fatter during times of large capacity buffers, likely due to the positive correlation with the fact that large capacity buffers are prevalent during recessions. This is corroborated by the lower panel which displays the distribution of shocks when the economy is in a recession versus expansion. Though it is also fairly symmetric, it has a more noticeable fatter lower tail.

Figure 28: Regime Specific Shock Distributions



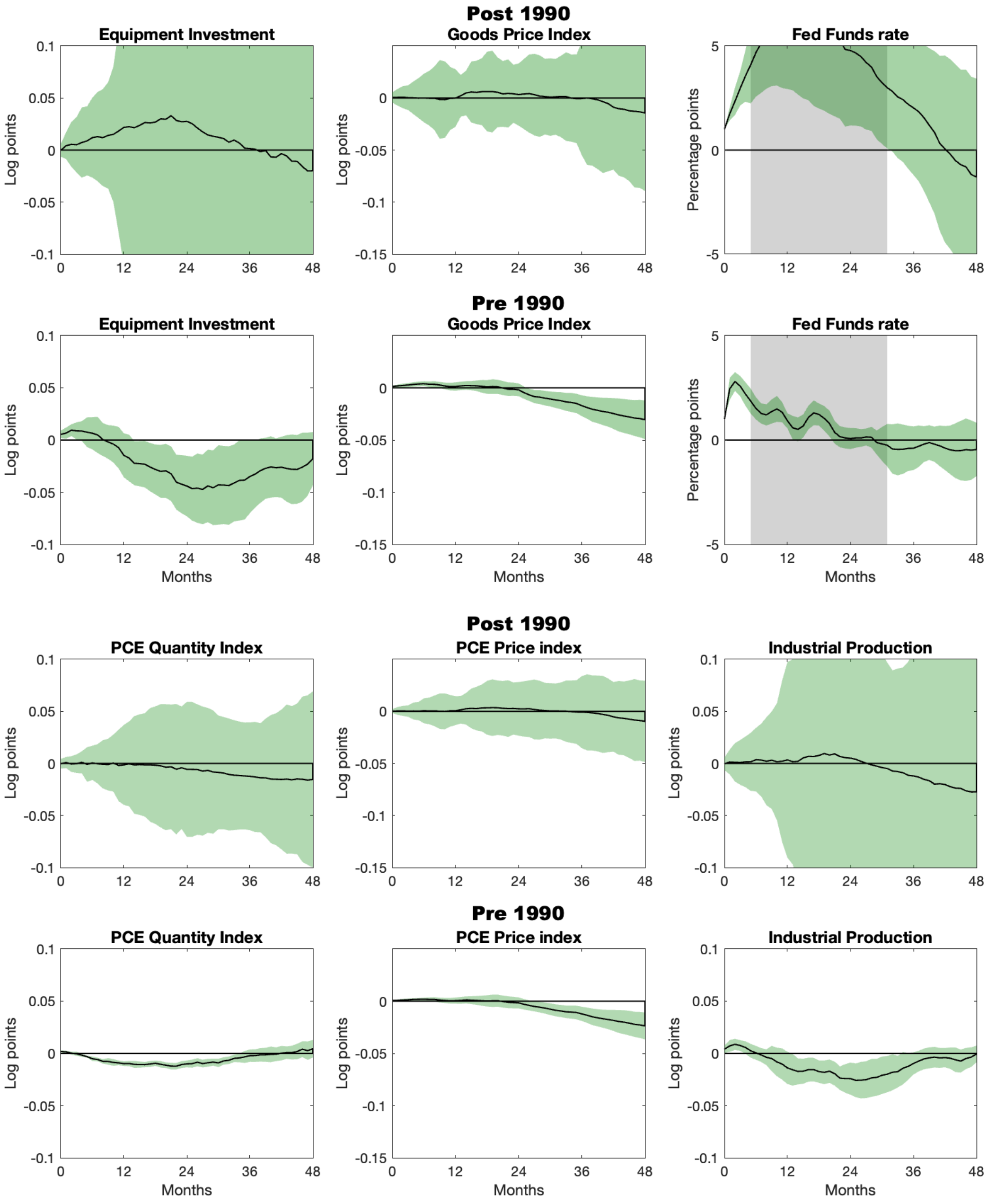
Notes: Distributions of monetary policy shock size conidtional on the regime and state variable underlying the regime.

Figure 29: Capacity-Dependent Impulse Responses with Detrended State



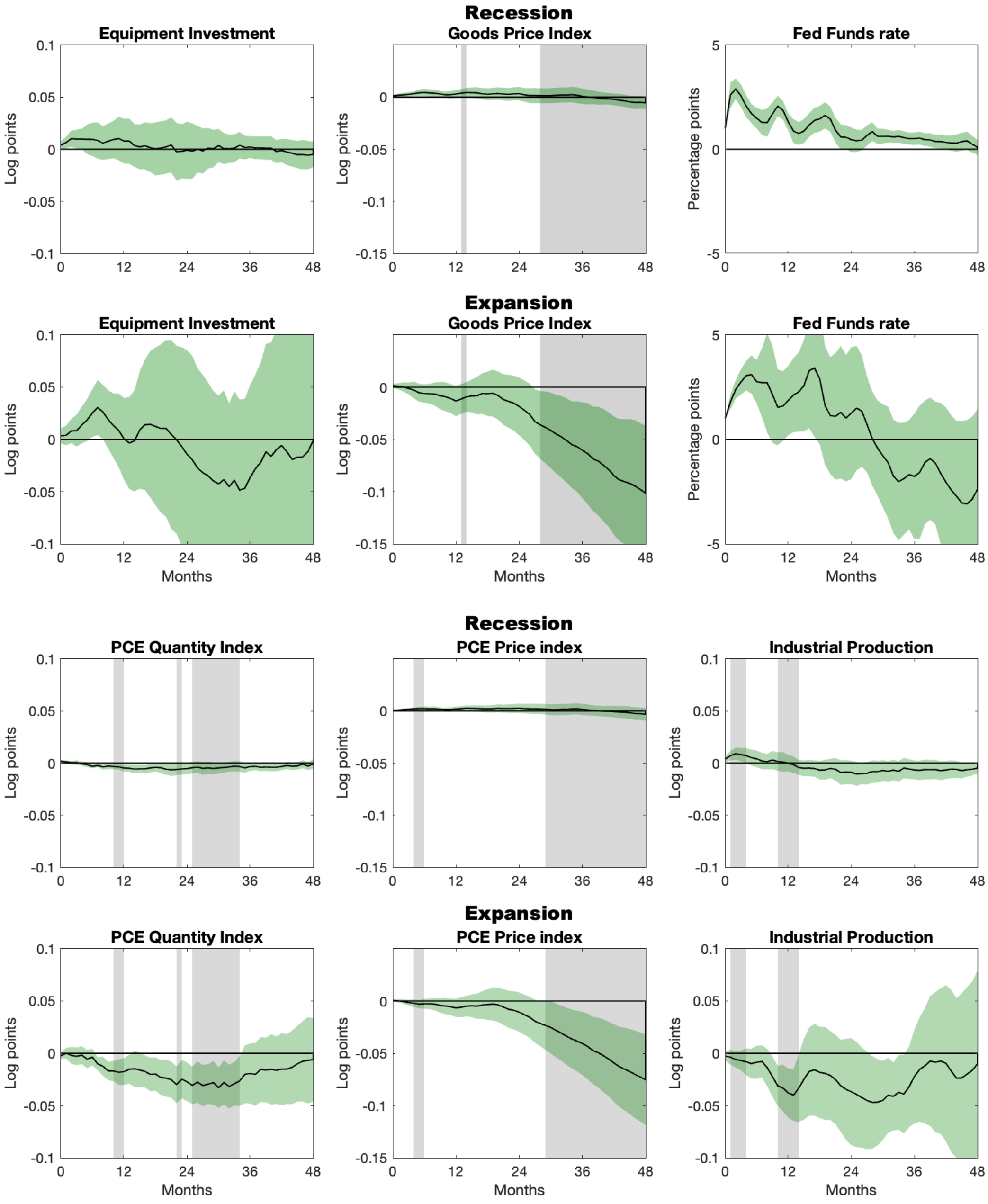
Notes: Estimated impulse responses to a contractionary one percent monetary policy shock in the logit smooth transition local projection model with state variable being the detrended measure of capacity buffers. Row one depicts the response in the regime where firms have large capacity buffers, and row two depicts the response in the regime where firms have small capacity buffers. Black lines indicate the point estimates of the impulse response while green areas show the 95 percent confidence intervals calculated from the Driscoll-Kray standard errors.

Figure 30: Impulse Responses pre and post 1990



Notes: Estimated impulse responses to a contractionary one percent monetary policy shock in the logit smooth transition local projection model with state variable being the detrended measure of capacity buffers. Row one depicts the response in the regime where firms have large capacity buffers, and row two depicts the response in the regime where firms have small capacity buffers. Black lines indicate the point estimates of the impulse response while green areas show the 95 percent confidence intervals calculated from the Driscoll-Kray standard errors.

Figure 31: Recession-Dependent Impulse Responses



Notes: Estimated impulse responses to a contractionary one percent monetary policy shock in the logit smooth transition local projection model with state variable being a 21 month trailing moving average of growth in consumption expenditures. Row one depicts the response in the recessionary regime, and row two depicts the response in the expansionary regime. Black lines indicate the point estimates of the impulse response while green areas show the 95 percent confidence intervals calculated from the Driscoll-Kray standard errors.