The Role of Sell Frictions for Inventories and Business Cycles

Wouter J. Den Haan
London School of Economics, CEPR and CFM

Tiancheng Sun
University of Edinburgh

June 3, 2024

Abstract

Although investment in inventories significantly impacts GDP fluctuations, inventories are often omitted from business-cycle models due to their complex cyclical behavior. We incorporate finished-goods inventories into a New-Keynesian framework by introducing a tractable microfounded "sell friction." Our approach simplifies existing approaches by avoiding product-specific idiosyncratic shocks while capturing the essence of the popular stockout avoidance motive. Specifically, firms strategically accumulate inventories by bringing more products to the market than they anticipate selling, thereby boosting expected sales. Our setup automatically generates key stylized facts such as the countercyclical nature of the inventory-sales ratio and the greater volatility of output compared to sales under business cycles driven by monetary-policy (demand) shocks. A novel aspect of our analysis is the recognition of an inventory good as an asset and that cyclical fluctuations of its value play a key role following supply shocks. Specifically, the value of an inventory good is robustly countercyclical in our model when the productivity-growth process mirrors the observed positive autocorrelation. This ensures that the model also robustly replicates stylized inventory facts in response to productivity (supply) shocks, which has been a challenge in the literature. Using inventory and sales data to discipline the model, we find that productivity shocks account for a large fraction of GDP fluctuations, ranging from 62.5% to 94%. Furthermore, the goods-market friction yields non-trivial effects on the magnitude of aggregate fluctuations, underscoring the importance of incorporating inventories into macroeconomic models.

Keywords: Search friction, goods and service sector, TFP and monetary-policy shocks.

JEL Classification: E32, G31.

Acknowledgement: We would like to thank Greg Kaplan, Michael McMahon, Pascal Michaillat, Ben Moll, Valerie Ramey, and Petr Sedláček for help comments.

1 Introduction

In the third quarter of 2023, real US GDP increased by 4.90% of which more than a quarter, namely 1.27 percentage points, consisted of investment in private inventories. This was not an unusual quarter.[1] As documented in section 2, inventory investment is not only quantitatively important, it also displays *systematic* cyclical behavior. This is an old observation. In fact, both the quantitative and the cyclical relevance of inventory investment was acknowledged in the literature quite a while ago.[2]

During the last couple decades, several theoretical frameworks have been proposed.[3] Nevertheless, inventories are still rarely modeled in modern business-cycle analysis. An apparent reason is that the behavior of inventories, production, and sales is challenging and difficult to capture with standard frameworks. Thus, the objective of this paper is to develop a microfounded framework that can capture key inventory, production, and sales data facts for *both* demand and supply shocks *and* is simple enough to incorporate into state-of-the-art business-cycle models. We focus on finished-goods inventories in the manufacturing, wholesale, and retail sector which cover on average 61.6% of total inventories. This type of inventories is responsible for 87.0% of the volatility of investment in non-farm inventories.[4]

What are those challenging inventory facts?[5] One might think that inventories build up during recessions as firms face difficulties in selling their goods. In fact, the investment in inventories as well as the inventory level are strongly procyclical. But this could still be quite easily explained with a scale effect, that is, inventory levels would scale up and down with aggregate activity. There is more to it, however, because the reason inventories are procyclical is that output is more volatile than sales.[6] It seems quite plausible that adjusting production levels is costly, but a model with such costs would predict that output is *less* volatile than sales and inventories would do the

[1]In the first quarter of 2023, real GDP increased by 2.20%, while the role of investment in private inventories was equal to *minus* 2.22 percentage points, that is, without the drop in inventories, the increase in GDP would have been twice as large. These numbers are from the March 28 2024 release of the Bureau of Economic Analysis.

[2]See, for example, Blinder and Hotz-Eakin (1986).

[3]Exemplary papers are Eichenbaum (1989), Ramey (1991), Bils and Kahn (2000), Coen-Pirani (2004), Khan and Thomas (2007), and Kryvtsov and Midrigan (2013).

[4]See section 2 for additional information.

[5]See section 2 for a detailed discussion, but also Ramey and West (1999) for an earlier discussion.

[6]We will document this and other key empirical facts using an updated US data set in section 2. See Kahn (1987), Ramey (1989), Blinder and Maccini (1991), Kahn (1992), Wen (2005), Wen (2008) and Kryvtsov and Midrigan (2013) for earlier discussions on this intriguing empirical fact. Wen (2005) documents that this fact does not hold when a band-pass filter is used that extracts that part of the data associated with a cycle between two and three quarters. For our updated sample with US data, we find, however, that the standard deviation of the *growth* rate of production in the goods sector is 17% higher than the standard deviation of the *growth* rate of the associated sales series. The first-difference filter, i.e., the growth rate, is not the same as a band-pass filter, but it also emphasizes high frequencies.

adjusting, not output levels.[7] This is not observed in the data. It is true, however, that firms are less efficient during recessions in that they hold more goods in inventory per unit of sales, that is, the inventory-sales ratio is countercyclical.[8]

In this paper, we develop a new framework to model inventories that can replicate these key inventory facts in response to both demand and supply shocks. Furthermore, it can be incorporated into a New-Keynesian (NK) business-cycle model because of its simplicity. In terms of the relationship to the literature, there are two aspects worth mentioning. First, we capture an existing reason for why firms hold inventories with a simpler structure than what is used in the literature. That reason is the "stockout-avoidance motive."[9] The idea is that firms face idiosyncratic demand shocks for their products *and* they have to set the price and production level before this idiosyncratic shock is known. One can think of this uncertainty as a matching friction; the larger the standard deviation of the idiosyncratic shock, the bigger the friction. This motivated us to adopt a standard matching friction like the one used in the macro-labor search literature. This approach is much simpler because it can be implemented using a representative firm and avoids the complexity that heterogeneity adds to the analysis in terms of calibration and numerical solutions. The implications of our model are similar to the version with heterogeneity and an explicit stockout-avoidance mechanism: in response to a positive demand shock, there is a reduction in markups which induces firms to be more efficient, that is, they hold less inventories relative to sales.[10] Our paper also differs from the literature in that variations in markups arise endogenously as a consequence of sticky prices because we incorporate this inventory-holding motive into a general-equilibrium New-Keynesian model.[11]

The second aspect of our approach worth mentioning is that – to the best of our knowledge – we are the first to stress the importance to think about inventories as an asset when studying its business-cycle properties. Whereas many papers in the inventory literature have a constant discount factor, we show that cyclical variations in the marginal rate of substitution, i.e., the pricing kernel, are key to ensure that the model can also replicate observed key inventory, production, and sales facts in response to productivity shocks.

More precisely, we do the following. A key step is to introduce a matching friction such that a good produced is no longer sold with unit probability. That is, sales are no longer equal to production, but depend on the (search) effort put in by buyers and the total amount of goods brought to the market which is equal to newly produced

[7] See Eichenbaum (1989), Ramey (1989), Blinder and Maccini (1991), and Wen (2005).

[8] This empirical fact has received a lot of attention in the literature. See, for example, Bils and Kahn (2000), Coen-Pirani (2004), Wen (2008), and Kryvtsov and Midrigan (2013). The literature points out that this negative correlation between the inventory/sales ratio and aggregate activity could be due to inventories responding slower than sales to changes in economic conditions.

[9] See Kahn (1992), Wen (2005), Wen (2008), and Kryvtsov and Midrigan (2013).

[10] Below we explain why production is more volatile than sales.

[11] In Kryvtsov and Midrigan (2013), markups are endogenous for the same reason, but they have a general-equilibrium model in which monetary policy follows an exogenous money-supply rule instead of the usual interest-rate-setting Taylor rule.

goods plus the beginning-of-period inventory stock. We assume that both goods and services are affected by such a sell friction. The service sector differs from the goods sector because no inventories are accumulated when firms sell less than what they could sell.[12] Firms in both sectors strategically supply more products than they anticipate selling to boost expected sales, which in the goods sector leads to optimal inventory accumulation.

The inventory-sales ratio is a key variable in the inventory literature and it plays a key role in our model as well. Given our theoretical analysis, a more convenient empirical measure is the customer-finding rate or fraction sold. This is a simple monotone inverse transformation of the inventory-sales ratio.[13] Thus, whereas the inventory-sales ratio is countercyclical, the customer-finding rate is procycyclical. The customer-finding rate is the same as the sell fraction, that is, the ratio of goods sold relative to available goods for sale, that is, newly produced goods plus the inventory stock. As in the standard New-Keynesian framework, firms face monopolistic competition. In our framework, this means that firms can choose the price and production level independently; both affect the demand for their product and, thus, the customer-finding rate. In contrast, to the New-Keynesian framework, sales are no longer equal to production and the difference between the two leads to inventory accumulation. Our generalization of the NK demand function implies that the New-Keynesian Phillips Curve not only includes current inflation, expected inflation and marginal costs, but also the customer-finding rate (i.e., the inventory-sales ratio) as well as an asset price, namely the value of inventory goods.[14]

Our simple goods-market friction naturally predicts observed facts related to the behavior of inventories, production, and sales. To understand this, suppose that the customer-finding rate is constant. Sales will then be less volatile than output, since the level of inventories is a stock and only increases gradually (and not at all on impact). Of course, if the customer-finding rate would increase a lot when output increases, then sales would be more volatile than output. Thus, parameters must be such that the model-predicted volatility of the customer-finding rate resembles the volatility observed in the data, that is, it should be procyclical, but not too volatile.

We will show that our model can predict a procyclical (countercyclical) customer-finding rate (inventory-sales ratio) in response to both a monetary-policy (demand) shock as well as a productivity (supply) shock.[15] It is not surprising that the model can do so in response to a demand-type shock as this will induce buyers to adjust

[12] The potential level of sales is simply what a standard production function implies given the productivity level as well as the amount of capital and number of workers in place.

[13] See equation 3.

[14] This could affect empirical properties of the NK Phillips curve, since asset prices are potentially quite volatile. Also, it means that the real interest rate – through its effect on the marginal rate of substitution – has a direct impact on the Phillips curve.

[15] As discussed in section 3.5, the model of Kryvtsov and Midrigan (2013) can also replicate key inventory facts following a monetary-policy shock, but cannot do so in response to a productivity shock for the usual case with sticky prices.

search effort levels which has a direct effect on the customer-finding rate. What about productivity shocks? In general equilibrium, an increase in productivity does not only lead to an increase in supply, but also to income which in turn leads to an increase in demand and search effort. By itself, this effect will lead – at best – to an acycyclical customer-finding rate. But there is another element in our model and that is that the value of an inventory good is countercyclical. This causes the customer-finding rate to be procyclical, because firms will set prices and production levels such to economize on inventories relative to sales levels. To understand this, it is important to realize that an inventory good is a durable asset and a key determinant of its value is the marginal rate of substitution.[16] During an expansion, consumption is expected to increase which means that the marginal rate of substitution drops. That is, economic agents would prefer to save less and the value of assets like inventory goods drops. This will induce firms to set the price and output level such that the customer-finding rate increases and inventories increase by less than they would have done if the customer-finding rate would have remained constant.

To ensure that our relatively simple model robustly predicts a countercyclical marginal rate of substitution, it is important that the process for TFP is – like its empirical counterpart – a non-stationary process with a positive serial correlation in the growth rate.[17]

We use the model to illustrate how inventory facts provide identifying information for the role of monetary policy and TFP shocks for GDP fluctuations. Specifically, using some key estimated inventory moments – and taking into account sampling variation – we find that TFP shocks are responsible for at least 62.5% and at most 94% of GDP fluctuations. Furthermore, the presence of the goods-market friction yields non-trivial effects on aggregate fluctuations, underscoring the importance of incorporating inventories into macroeconomic models.

The remainder of this paper is organized as follows. Section 2 describes our goods-market efficiency measure used, i.e., the customer-finding rate, its relationship to the inventory-sales ratio, and describes key aspects of its observed cyclical behavior. Section 3 describes the model with a goods sector only and discusses key properties of our framework. Section 4 extends the model to include a service sector. There are no inventories in the service sector, but firms in this sector also face the possibility that sales are less than what could be provided given available resources. Think of empty

[16]The value of an inventory good also depends on the price at which it can be sold at a future date and with what probability, but the marginal rate of substitution turns out to be key.

[17]There are ways to get the desired hump-shaped Impulse Response Function (IRF) for consumption when TFP is a stationary process, for example, with habits. Even in our simple model with a standard utility function without habits, it is possible to get a hump-shaped IRF for consumption with a simple stationary TFP process. But we prefer to work with the slightly more complicated non-stationary process, because it is and more realistic and ensures that the model is robustly consistent with key observed inventory facts. Bansal and Yaron (2004) point out that long-run properties of the model's driving process are important for asset prices. A novel insight about our paper is that the way TFP is modeled not only affects an asset price, the value of an inventory good, but also key business-cycle variables.

restaurant tables. The last section concludes.

2 Empirical

In this section, we document some key stylized facts regarding the (cyclical) behavior of inventories, production, and sales. We also discuss the role of inventory investment for fluctuations in GDP and aggregate expenditure components.

Inventory components considered. We use quarterly US data for the period starting in the first quarter of 1967 and ending in the last quarter of 2019.[18] Inventories consist of materials and supplies, work-in-progress, and finished goods. Our theoretical analysis analyzes the third component and we abstract from the first two inventory categories in our empirical analysis as well. Our inventory series include finished goods in the manufacturing, the wholesale, and the retail sector. The idea of our goods-market friction is that produced goods do not instantaneously and frictionlessly end up in the hands of ultimate buyers, i.e., consumers and investors. This indicates that we have to include *all* finished goods no matter where they are located.

This inventory aggregate of finished goods covers on average 61.6% of total non-farm inventories of which 22% is located in the manufacturing sector, 41% in the wholesale sector, and 37% in the retail sector. Regarding variability, we find that manufacturing finished-goods inventories explain 8.8% of the business-cycle fluctuations in non-farm inventories, wholesale 30.1%, and retail 23.3%, so together 62.2%. For the change in inventories, these numbers are 30.7%, 29.6%, and 26.7% for the three components separately and 87.0% for the three components together.[19] Thus, the three types of finished-goods inventories considered form a large part of the stock of non-farm inventories and capture a big part of the fluctuations in non-farm inventories.

Customer-finding rate and inventory-sales ratio. In a New-Keynesian business-cycle model without inventories, the price set by firms determines demand, which is equal to the amount of goods sold, which in turn is assumed to be exactly equal to production. That is, the probability of finding a buyer for each good produced is equal to one. In general, however, what the firm sells and could sell are not the same. In our theoretical model, this occurs because of a goods-market friction and a key variable is the firm's customer-finding rate or fraction sold. This rate is denoted by $f^f_{g,t}$ for the goods sector and by $f^f_{s,t}$ for the service sector. In the goods sector, the difference between sales and production affects the stock of inventory goods.

[18] Data are from the Bureau of Economic Analysis and details are given in appendix A.1. We start in 1967Q1 because that is the first quarter for which a series for finished-goods inventories of the manufacturing sector is available. We end the sample in 2019Q4 to exclude the unusually large and irregular fluctuations observed during the pandemic.

[19] Appendix A.1 discusses details on how these and other statistics reported in this section are calculated.

The customer-finding rate can be measured using the observed fraction of available goods sold. That is,[20]

$$f^f_{g,t} = \frac{s_{g,t}}{y_{g,t} + (1-\delta_x)x_{t-1}}, \tag{1}$$

where $y_{g,t}$ denotes newly produced goods, $s_{g,t}$ firm sales, and $(1-\delta_x)x_{t-1}$ the amount of last period's accumulated inventories that did not depreciate.

The amount produced, $y_{g,t}$, equals sales plus the investment in inventories, that is,

$$y_{g,t} = s_{g,t} + x_t - (1-\delta_x)x_{t-1}. \tag{2}$$

Combining the last two equations gives

$$f^f_{g,t} = \frac{s_{g,t}}{s_{g,t} + x_t} = \frac{1}{1 + {x_t}/{s_{g,t}}}. \tag{3}$$

That is, the customer-finding rate is a simple transformation of the inventory-sales ratio, a popular statistic in the inventories literature. It is important to use *final* sales in our empirical measure of the customer-finding rate, because our theoretical model does not have separate manufacturing, wholesale, and retail firms, each with their own inventories. We use the "final sales for goods and structures" series. We are interested in the mean, the volatility, and the cyclicality of the customer-finding rate.

Consistent production series. Production is known to be more volatile than sales, which means that inventories *increase* during expansions and *decrease* during recessions.[21] This is a robust empirical finding that has been challenging for the inventory literature. If it is costly to adjust production levels, then one would expect that output is *less* volatile than sales. The contribution of our paper is to show that the observed relative volatility is a natural prediction in a model in which the friction is related to *selling* goods instead of producing them.

We would like to construct a production series that corresponds to the inventory and sales data for the sector producing goods and structures from Table 5.8.5 of the National Income and Product Accounts published by the Bureau of Economic Analysis. There are two possibilities. One approach is to make an assumption about the depreciation rate and construct a production series using equation (2). A natural choice is to use 0.4% which is consistent with average investment in inventories as a fraction of GDP.[22]

[20]The superscript f indicates that the finding rate is viewed from the point of view of firms. In our model, there is also a finding rate from the buyers' point of view, which will be denoted by $f^b_{g,t}$.

[21]See Ramey and West (1999) for a discussion and early empirical evidence. As pointed out by the authors, the identity in equation (2) implies that production *must* be more volatile than sales if the correlation between investment in inventories, $x_t - (1-\delta_x)x_{t-1}$, and sales, $s_{g,t}$, is positive as observed in the data.

[22]Kryvtsov and Midrigan (2013) use a value equal to zero, but this value is not that different from 0.4% and the statistics calculated are not sensitive to such minor changes in the assumed depreciation

The alternative is to use Gross Domestic Product generated by the sectors producing goods and structures as reported in Table 1.2.6. For both output series we find that the volatility levels of their cyclical components are higher than the one for sales. And this difference is the biggest for the GDP series.[23] Using the constructed series has the disadvantage that it requires an assumption about the depreciation rate. But we prefer to report data characteristics using the constructed one because it means that all series are based on the same set of firms.[24]

Inventory, production, and sales properties. In table 1, the column labeled DATA documents key observations related to inventories, production, and sales. The top panel reports statistics related to the customer-finding rate. The average fraction sold is equal to 0.506 which corresponds to an average inventory-sales ratio just below 1.[25] That is, quarterly sales, quarterly newly produced goods, and the stock of inventories are roughly equal to each other.

We use HP-filtered data to evaluate business-cycle fluctuations. The customer-finding rate is strongly procyclical and that is true when GDP is used and when our constructed production measure is used.[26] To put fluctuations of the customer-finding rate of this sector in proper context, we consider the output measure of this sector itself and not GDP.

The procyclicality is also illustrated in figure 1 which plots the business-cycle components of the customer-finding rate and the output series for the goods and structures sector. The figure documents that the fraction sold dropped by several percentage points during the 1974, the 1982, and the 2008 recession. If goods cannot be carried over as inventory, then a drop in the fraction sold from 0.51 to 0.49 would correspond to a 4% price drop which is obviously nontrivial. When the good can be stored as inventory and sold in subsequent periods, then that would still incur storage costs and depreciation, and possibly a loss in value

The customer-finding rate is *negatively* correlated with beginning-of-period inventory stocks. This is not one of the key moments considered in the inventory literature. It is intriguing, however, that the correlation coefficients of the customer-finding rate with newly produced goods and lagged inventory have different signs. We will show

rate.

[23]Although, the correlation of the cyclical components of these two series is equal to 0.98, the GDP series is quite a bit more volatile.

[24]As discussed in detail below, there is tight link between the volatility of the customer-finding rate and the relative volatility of output to sales. Our measure for the customer-finding rate is a simple transformation of the inventory-sales ratio. If the empirical production data used are not consistent with the sales and inventory data, then it will not be possible for a model to match both statistics because model-generated data for inventories, production, and sales data are, of course, consistent with each other.

[25]The BEA and the media report the inventory-sales ratio based on monthly sales, which is three times bigger than the one used here based on quarterly sales.

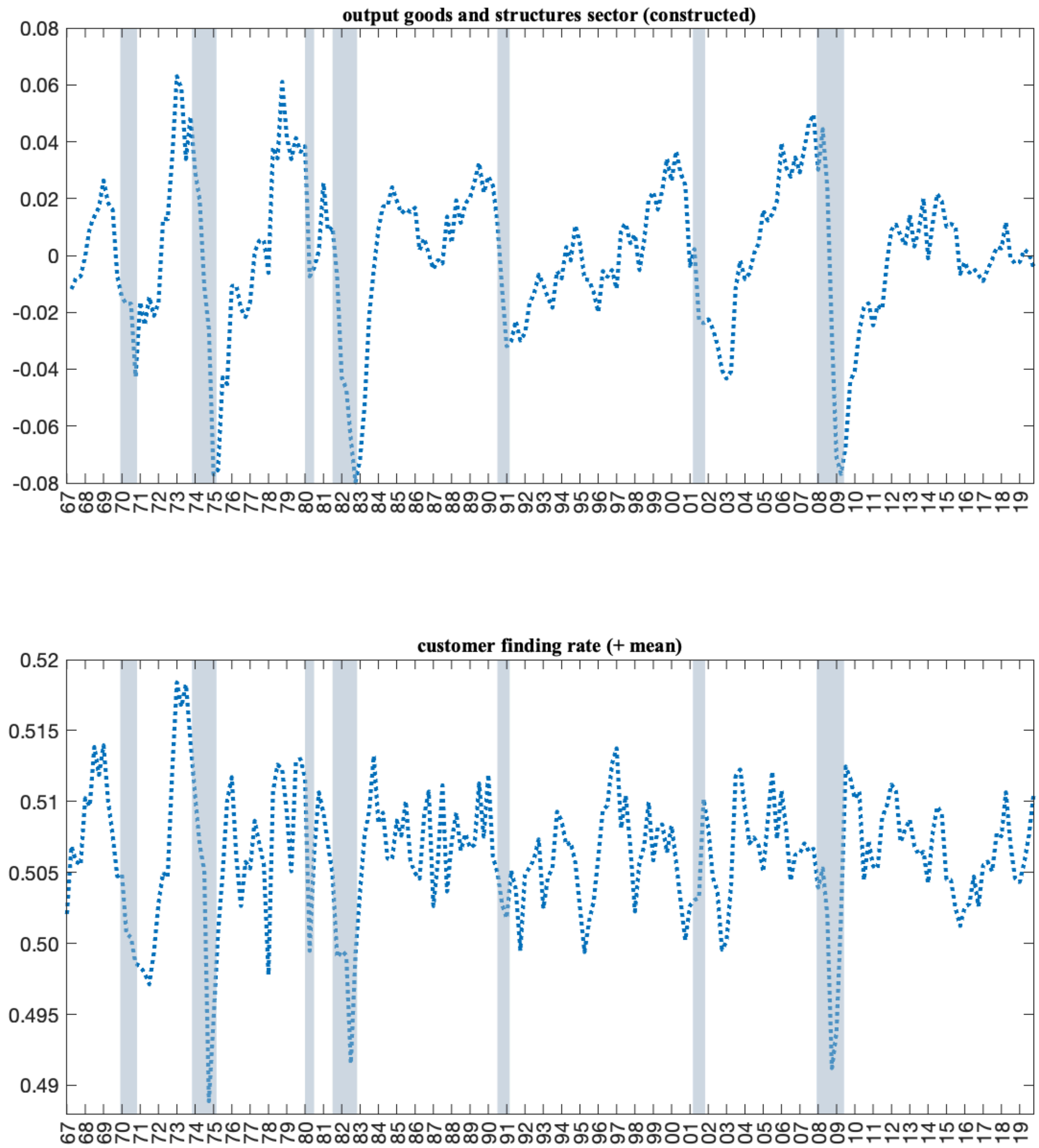
[26]It is also true when the GDP series for the goods and structures sector is used. All these productivity measures are highly correlated and imply similar correlation coefficients.

Table 1: Inventory stylized facts and model predictions

	DATA	MODEL $\nu_g, \nu_s, \Gamma_y \in$ calibrated range			MODEL $\nu_g, \nu_s, \Gamma_y, \sigma_R, \sigma_A$ estimated		
		TFP&R	TFP	R	TFP&R	TFP	R
customer-finding rate statistics							
$\mathbb{E}[f_g^f]$	0.506	=	=	=	=	=	=
	(0.002)						
$\frac{\sigma_{f_g^f}}{\sigma_{y_g}}$	0.170	0.099	0.066	0.164	0.147	0.073	0.218
	(0.157)	(0.009)	(0.005)	(0.005)	(0.013)	(0.005)	(0.008)
$\rho(f_g^f, y_g)$	0.514	0.594	0.444	0.935	0.683	0.559	0.901
	(0.109)	(0.067)	(0.023)	(0.005)	(0.061)	(0.029)	(0.009)
$\rho(f_g^f, x_{g,-1})$	-0.223	-0.018	-0.222	0.645	0.069	-0.063	0.392
	(0.105)	(0.115)	(0.078)	(0.049)	(0.123)	(0.096)	(0.047)
inventory, production, and sales statistics							
$\mathbb{E}[\frac{x}{s_g}]$	0.976	=	=	=	=	=	=
	(0.077)						
$\frac{\sigma_{y_g}}{\sigma_{s_g}}$	1.124	1.175	1.212	1.081	1.109	1.192	1.006
	(0.026)	(0.030)	(0.034)	(0.020)	(0.027)	(0.033)	(0.019)
$\frac{\sigma_x}{\sigma_{s_g}}$	0.835	0.812	0.936	0.313	0.667	0.877	0.257
	(0.054)	(0.041)	(0.011)	(0.009)	(0.054)	(0.013)	(0.013)
$\frac{\sigma_{x/s_g}}{\sigma_{s_g}}$	0.749	0.453	0.313	0.692	0.637	0.342	0.857
	(0.045)	(0.044)	(0.033)	(0.008)	(0.054)	(0.032)	(0.017)
$\rho(\frac{x}{s_g}, s_g)$	-0.583	-0.599	-0.355	-0.991	-0.752	-0.509	-0.971
	(0.113)	(0.069)	(0.033)	(0.001)	(0.054)	(0.037)	(0.002)
$\rho(s_g, y_g)$	0.941	0.594	0.444	0.935	0.683	0.559	0.901
	(0.033)	(0.067)	(0.023)	(0.005)	(0.061)	(0.029)	(0.009)
$\rho(x, y_g)$	0.630	0.839	0.867	0.977	0.770	0.869	0.687
	(0.095)	(0.026)	(0.019)	(0.001)	(0.045)	(0.019)	(0.044)

Notes. Inventory series are based on finished goods in the manufacturing, wholesale, and retail sector. Sales are final sales in the sector producing goods and structures. The customer-finding rate, f_g^f, is calculated using equation (3). Also, x denotes inventories, s_g sales, and y_g output of the goods and structures sector. The DATA column reports standard errors in parentheses; these are calculated using the VARHAC procedure of Den Haan and Levin (1997) which corrects for serial correlation and heteroskedasticity. The columns for model-generated statistics report the means across 10,000 replications of length 212 (same length as the data set) as well as – in brackets – the standard deviation across replications. The column labeled "TFP&R" uses a mix for the two innovation standard deviations as discussed in the main text. In the other columns only one type of shock is driving fluctuations. The estimated parameter values are as follows: $\nu_g = 0.3469$, $\nu_s = 0.6713$, $\Gamma_y = 0.012$, $\sigma_R/\sigma_A = 0.8974$. The representative combination for the calibrated range consists of $\nu_g = \nu_s = 0.565$ and $\Gamma_y = 0.03$. In this case, the value of σ_R/σ_A is set equal to 0.5921 which ensures that the means of both $\rho(f_g^f, x_{g,-1})$ and $\sigma_{y_g}/\sigma_{s_g}$ across replications are inside the empirical 95% confidence intervals of their empirical counterpart. Throughout this paper, we extract business-cycle components using the HP filter with a smoothing coefficient of 1,600.

Figure 1: Cyclical behavior of the customer-finding rate



Notes. These panels plot the HP-filtered values of (the log of) goods-sector production and the customer-finding rate, i.e., the fraction of available goods sold, calculated according to equation (3) using as the measure for inventories finished goods in the manufacturing, wholesale, and retail sector and final sales in the sector producing goods and structures.

that this statistic has important identifying information for the relevance of monetary policy and TFP shocks. The bottom panel of table 1 focuses on traditional inventory statistics. It shows that the inventory-sales ratio is countercyclical which must be true given that the customer-finding rate is procyclical. Inventories of finished goods are procyclical, that is, recessions are *not* periods when sellers are stuck with increased stocks of unsold goods. A related – but even more intriguing – observation is the well-known fact that production is *more* volatile than sales. For our sample, output is 12% more volatile.

The cyclicality of the customer-finding rate and the volatility of output relative to the volatility of sales are quantitatively related to each other. Suppose that originally the situation is as follows: output is equal to 1, the (undepreciated) inventory stock is equal to 1, and sales are equal to 1. This means that the fraction sold, i.e., the customer-finding rate, is equal to 0.5, which is almost identical to the observed average. If output increases by 1% and the customer-finding rate remains constant, then sales increase by only 0.5% so output is twice as volatile as sales. The reason why sales are less volatile is that inventories are less volatile than output and on impact do not change at all. The cyclicality reported in table 1 indicates that the customer-finding rate would increase from 0.5 to 0.5017 when output increases with 1%, which means that sales increases with $0.842\%(= 100 \times (0.5017 \times (1.01+1) - 1)/1)$. This implies that on impact the change in output is equal to 1.19 times the change in sales. This is more than the overall observed relative standard deviation. But as inventories increase, the percentage increase in sales will get closer to the percentage increase in output.[27] If the increase in the customer-finding rate would be 0.249 basis points then sales would be as volatile as output on impact.

The purpose of this numerical example is twofold. First, it relates the empirical volatility of the customer-finding rate to the relative volatility of output to sales, although only on impact. Second and more importantly, it makes clear how sensitive the relative volatility of output to sales is to small changes in the customer-finding rate. It drops from twice as volatile to only 1.19 times as volatile when the customer-finding rate increases with 0.17 basis points instead of remaining constant. To be consistent with a procyclical customer-finding rate, the model should predict that the customer-finding rate goes up in an expansion, but if it increases too much, then sales will be more volatile than output which is counterfactual.

Customer-finding rate in the service sector. Firms providing services are also likely to face a sell friction. For example, a restaurant will fill more tables when demand increases. Also, it can expect to fill more tables – but a smaller fraction – if it increases the number of tables. In theory, a customer-finding rate can be constructed for services. Taking into account that there are no inventories we would get that $f^f_{s,t} = s_{s,t}/y_{s,t}$. In practice, however, there is a problem since $y_{s,t}$ would *not* be actual production, but

[27] Also, the relative standard deviations reported in the table are based on HP-filtered data so not perfectly comparable.

the amount of services that *could* be supplied during the period given the number of workers hired and capital installed.[28] However, some information on the customer-finding rate in the service sector for the Euro area and the European Union may be obtained from a relatively new survey of the European Commission. This survey asks firms providing services the following question: *"If the demand addressed to your firm expanded, could you increase your volume of activity with your present resources? Yes - No. If so, by how much? $\cdots$%."*

The answers are used to construct a capacity utilization measure.[29] Figure 2 displays the demeaned raw data for the log of the index and the log of Euro-Area GDP. The figure indicates that the utilization index moved together with economic activity.

One should be careful in concluding that this figure indicates that the customer-finding rate in the service sector is procyclical. First, the survey question does not make explicit what is meant with "resources." For example, a hair salon owner may interpret it as the number of booths in their salon. That is, resources are interpreted as capital as is usually the case in capacity utilization measures. But for our analysis, "resources" should also include variable inputs such as labor because those are key in determining potential output in the subsequent period. Another caveat is that the series are only available since 2011. In terms of business-cycles, this means that the Eurozone debt crisis and the pandemic are included, two economic downturns for which demand factors are believed to have been important. So it is not clear whether this measure will also be procyclical during other types of recessions. By contrast, the countercyclical behavior of the inventory-sales ratio and, thus, the procyclicality of the customer-finding rate in the goods sector is a well documented robust finding.

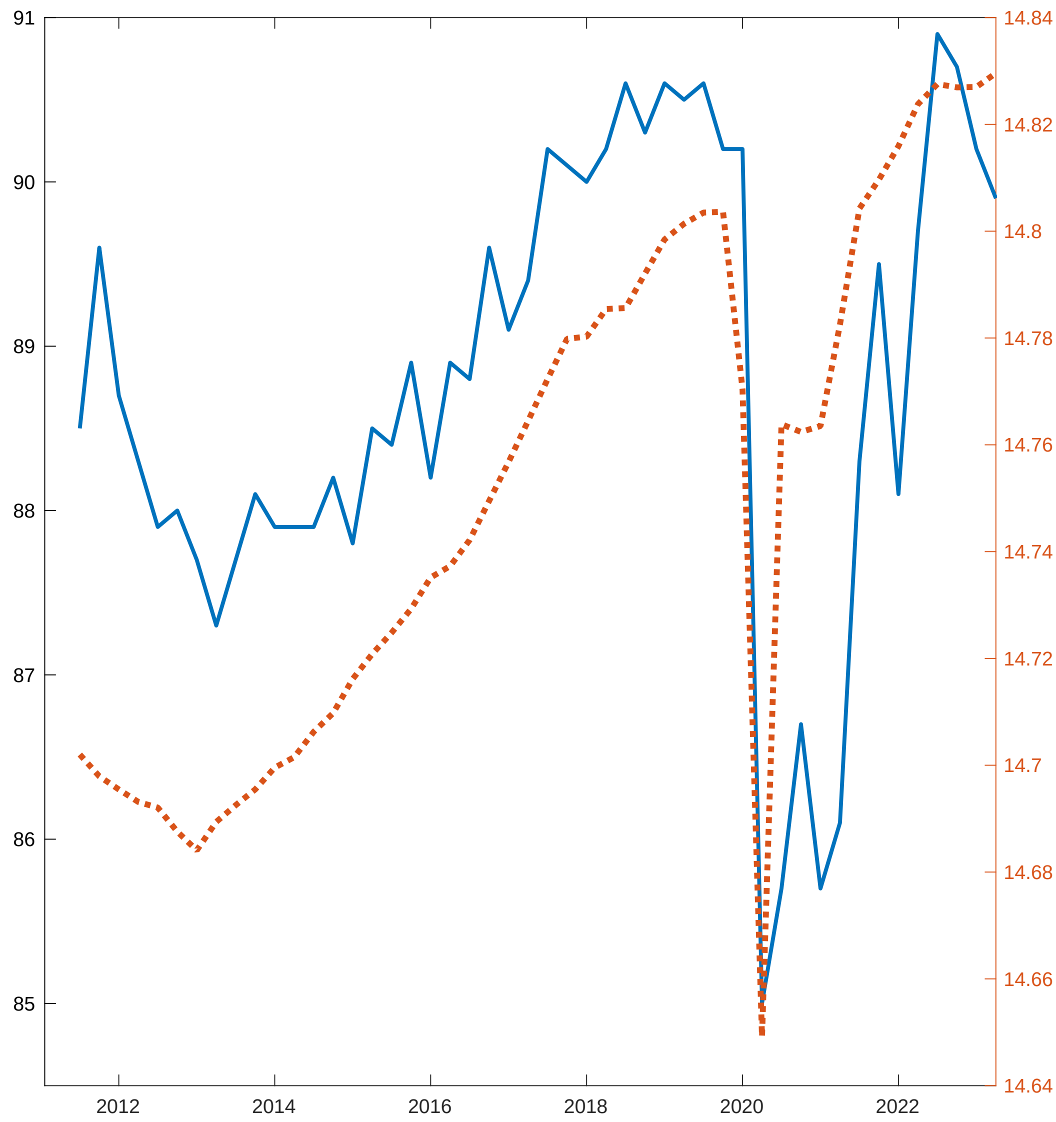
Business-cycle statistics including investment in inventories. Investment in inventories is on average a small component of GDP. For our sample, the change in private inventories (CIPI) is on average equal to 0.4% of GDP and 2.7% of total investment. But these statistics are completely misleading in terms of revealing the quantitatively important role of inventory investment for business-cycle fluctuations. Table 2 documents the role of consumption, investment excluding inventory investment, and inventory investment for fluctuations in GDP, total investment, and the sum of the two expenditure components usually modeled explicitly in business-cycle models which are consumption and total investment. We look both at HP-filtered levels and first-differences.[30] When the aggregate considered is equal to the sum of the components

[28]As shown in equation (3), we can construct a measure for the customer-finding rate for the goods sector using the observed inventory-sales ratio. But that is obviously not an option for the service sector.

[29]Specifically, the capacity utilization rate is defined in percentage as 100/(1+percentage increase/100).

[30]Since CIPI is at times negative, we cannot take logs to get scale free statistics. Therefore, we first scale all variables with the trend component of GDP, calculated as the exponent of the HP-trend of the log of GDP. And then we HP-filter this scaled variable.

Figure 2: Euro-Area capacity utilization in the service sector (-) and real GDP (:)



Notes. This figure plots the service-sector capacity-utilization index constructed by the European Commission (scale on left axis) and the log of real GDP for the Euro Area (scale on right axis).

considered, then the variance decomposition gives numbers that add up to 1.[31]

We find that CIPI is responsible for 44%, 21%, and 21% of the fluctuations of total investment, total investment plus consumption, and GDP, respectively. When we look at first differences, then these numbers are 79%, 38%, and 36%. These results indicate that not modeling inventory investment means missing an important part of fluctuations in key economic aggregates.

The stylized facts reported in this section also indicate that inventory investment is not just some uncorrelated additive noise. Thus, not incorporating inventories in business-cycle models is likely to mean that the model is misspecified in some dimensions. To conclude, table 3 reports some standard business-cycle statistics. This includes some statistics regarding the consumption of goods and services. Specifically, consumption of goods is more volatile than the consumption of services which is important for how we structure our full model with both a goods and a service sector.

Table 2: Variance decomposition

	c	i-CIPI	CIPI	sum
Business-cycle frequencies				
i	-	0.565	0.435	≡1
	-	(0.060)	(0.063)	
$c+i$	0.395	0.397	0.208	≡1
	(0.064)	(0.069)	(0.040)	
GDP	0.419	0.414	0.212	1.045
	(0.055)	(0.090)	(0.044)	
first differences				
Δi	-	0.215	0.785	≡1
	-	(0.030)	(0.082)	
$\Delta c + \Delta i$	0.367	0.258	0.375	≡1
	(0.053)	(0.046)	(0.053)	
ΔGDP	0.293	0.206	0.357	0.855
	(0.044)	(0.037)	(0.048)	

Notes. This table reports the role of the component listed in the top row for fluctuations in the aggregates listed in the left column at business-cycle frequencies. Here, c denotes total consumption, i total investment, and CIPI the change in (the level of) private inventories. Since CIPI can be negative, we cannot take logs to get scale-free statistics. Therefore, we scale variables with the trend component of GDP, calculated as the exponent of the HP-trend of the log of GDP. Next, we HP-filter this scaled variable. Standard errors are reported in partentheses and these are calculated using the VARHAC procedure of Den Haan and Levin (1997) which corrects for serial correlation and heteroskedasticity.

[31] When looking at the level of GDP, we find that the three components considered explain more than the total variance which means that the components left out, i.e., government expenditures and net export, actually help to reduce fluctuations in GDP because of a negative covariance.

Table 3: Business-cycle statistics

$\sigma_c/\sigma_{\text{GDP}}$	0.809
	(0.023)
$\sigma_i/\sigma_{\text{GDP}}$	4.576
	(0.321)
$\rho(\text{CIPI}, \text{GDP})$	0.613
	(0.077)
$\rho(\Delta x, \text{GDP})$	0.446
	(0.111)
$\mathbb{E}[c_g/c]$	0.337
	(0.023)
$\mathbb{E}[c_s/c]$	0.663
	(0.023)
$\sigma_{c_g}/\sigma_{\text{GDP}}$	1.608
	(0.090)
$\sigma_{c_s}/\sigma_{\text{GDP}}$	0.507
	(0.088)

Notes. This table documents the usual business-cycle statistics. Here, c denotes total consumption, i total investment, c_g consumption of goods, and c_s consumption of services. Since CIPI and Δx can be negative, the statistics based on these series are calculated as explained in the notes of table 2. Because of data availability, the numbers in the bottom half are for the sample from 2002Q1 to 2019Q4 whereas the numbers in the top half start in 1967Q1 like the other statistics calculated in this section. Standard errors are reported in partentheses and these are calculated using the VARHAC procedure of Den Haan and Levin (1997) which corrects for serial correlation and heteroskedasticity. The business-cycle components have been extracted using the HP filter.

3 Model with just a goods sector

In this section, we describe an economy in which firms only sell goods, i.e., not services. This allows us to present the key mechanisms related to the goods-market friction and inventory accumulation in a transparent manner and helps in understanding model properties. To facilitate this, we economize on notation and do not use a g subscript to indicate that firms are producing goods. In section 4, we add a service sector in which firms also face a search friction in finding customers.

The economy consists of a set of homogeneous households, a set of firms selling goods in a monopolistic market, and a central bank. Except for the goods-market friction which results in the accumulation of inventories, the model adopts standard New-Keynesian (NK) features. We do not have a separate wholesale and retain sector. But the idea is that our goods-market friction describes the friction in getting goods in the hands of buyers after production which in reality also involves moving through the wholesale and retail sector.[32]

3.1 Households

There is a continuum of households indexed by $h \in [0, 1]$. Households earn income by supplying labor and capital as well as through firm ownership and bond holdings. Income is used to buy consumption and investment goods as well as bonds. There is a continuum of goods indexed by $i \in [0, 1]$. A key feature of our model is that acquiring goods does not only require payment, but also some effort.[33]

Household labor supply. As in Erceg et al. (2000), we assume that households provide differentiated labor services to allow for sticky wages. The firm's labor input, n_t, depends on a CES aggregation of these differentiated labor services, that is,

$$n_t = \left(\int_{h=0}^{1} n_{h,t}^{\frac{\varepsilon_n - 1}{\varepsilon_n}} dh \right)^{\frac{\varepsilon_n}{\varepsilon_n - 1}}, \tag{4}$$

where $n_{h,t}$ is the amount of labor supplied by household h and $\varepsilon_n > 1$ is the elasticity of substitution between labor.

[32] In section 2, we discussed that 61% of total non-farm inventories consists of this type of inventories and this type is quantitatively important for several expenditure aggregates. We don't try to model the other type of inventories which consists of raw materials and intermediate goods. This type of inventory is typically motivated by a fixed cost in ordering as in Khan and Thomas (2007). Alternatively, this can be captured by assuming that inventories are an input in the production function as in Kydland and Prescott (1982) and Ramey (1989).

[33] Eurostat reports that the time spend on shopping and acquiring services ranges from 17 minutes per day in Romania to 35 minutes per day in Germany. See https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20181123-1.

The labor demand curve faced by household h is then given by

$$n_{h,t} = \left(\frac{W_{h,t}}{W_t}\right)^{-\varepsilon_n} n_t, \tag{5}$$

where $W_{h,t}$ is the nominal wage charged by household h and W_t is the aggregate nominal wage defined as

$$W_t \equiv \left(\int_{h=0}^{1} W_{h,t}^{1-\varepsilon_n} dh\right)^{\frac{1}{1-\varepsilon_n}}. \tag{6}$$

Adjusting the nominal wage incurs a utility cost given by[34]

$$\frac{1}{2}\eta_W \left(\frac{W_{h,t}}{W_{h,t-1}} - 1\right)^2 n_t, \tag{7}$$

where $\eta_W > 0$ measures the degree of nominal wage stickiness.

Acquiring consumption and investment goods. As in the standard New-Keynesian model there is a continuum of goods indexed by $i \in [0,1]$ and a CES aggregator is used to determine the amount of goods available to the household. That is,

$$c_{h,t} + i_{h,t} \leq s_{h,t} \equiv \left(\int_{i=0}^{1} s_{i,h,t}^{\frac{\varepsilon-1}{\varepsilon}} di\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{8}$$

where $s_{i,h,t}$ denotes the amount of good i sold by the producer of good i to household h, ε the elasticity of substitution, $s_{h,t}$ the amount of the aggregated good bought by household h, which can be allocated to consumption, $c_{h,t}$, and investment, $i_{h,t}$.

A key difference relative to the standard New-Keynesian framework is that households not only have to pay for the goods bought, but also incur an acquisition or collection cost to get the goods in their possession. Specifically, the amount of good i acquired, $s_{i,h,t}$, has to satisfy the following constraint:

$$s_{i,h,t} = f_{i,t}^b e_{i,h,t}. \tag{9}$$

where $e_{i,h,t}$ is the effort put in by household h to acquire good i. The value of $1/f_{i,t}^b$ indicates how much effort is required to buy one unit of good i and households take this as given.[35] The superscript b indicates that the friction is viewed from the buyer's

[34]We assume that all adjustment costs are utility costs. This has the advantage that the total amount of goods produced is still equal to the usual expenditure components, here consumption and investment.

[35]In equilibrium, $f_{i,t}^b$ is determined by the amount of goods supplied in market i and the *aggregate* amount of effort that households put in to acquire good i, $\int_{h=0}^{1} e_{i,h,t} dh$, which is not affected by the choice of an *individual* household. There is no randomness. If the household puts in $1/f_{i,t}^b$ units of

point of view. We refer to $e_{i,h,t}$ as "effort" to highlight that our goods-market friction is modeled in the same way as a matching friction. But we assume that this acquisition cost is in terms of goods that are being lost during the acquisition process.[36]

Household problem. The household problem is given by

$$\max_{\left\{\begin{array}{c} c_{h,t}, k_{h,t}, i_{h,t}, b_{h,t}, n_{h,t}, \\ W_{h,t}, e_{h,t}, e_{i,h,t}, s_{h,t}, s_{i,h,t} \end{array}\right\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\begin{array}{c} \xi_c \frac{(c_{h,t}-(\xi_e e_{h,t}-\bar{\xi}_e))^{1-\gamma}-1}{1-\gamma} - \xi_n n_{h,t} \\ -\frac{1}{2}\eta_W \left(\frac{W_{h,t}}{W_{h,t-1}} - 1\right)^2 n_t \end{array} \right]$$

subject to

$$\int_{i=0}^{1} \frac{P_{i,t}}{P_t} s_{i,h,t} di + \frac{b_{h,t}}{P_t} \leq \frac{W_{h,t}}{P_t} n_{h,t} + r_{k,t} k_{h,t-1} + d_{h,t} + \frac{1+R_{t-1}}{P_t} b_{h,t-1}, \tag{10a}$$

$$c_{h,t} + i_{h,t} \leq s_{h,t} = \left(\int_{i=0}^{1} s_{i,h,t}^{\frac{\varepsilon-1}{\varepsilon}} di\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{10b}$$

$$k_{h,t} = (1-\delta_k) k_{h,t-1} + i_{h,t} \left(1 - \frac{\eta_i}{2}\left(\frac{i_{h,t}}{i_{h,t-1}} - 1\right)^2\right), \tag{10c}$$

$$s_{i,h,t} = f_{i,t}^b e_{i,h,t}, \tag{10d}$$

$$e_{h,t} = \int_{i=0}^{1} e_{i,h,t} di, \tag{10e}$$

$$n_{h,t} = \left(\frac{W_{h,t}}{W_t}\right)^{-\varepsilon_n} n_t. \tag{10f}$$

Here, $P_{i,t}$ denotes the price charged by firm i, P_t the aggregate price index, $r_{k,t}$ the real rental rate, $k_{h,t}$ the end-of-period-t capital stock of household h, δ_k its depreciation rate, $e_{h,t}$ the total amount of effort put in by household h, and $d_{h,t}$ is the amount of firm profits received by household h. The constant term $\bar{\xi}_e$ is used as a normalization to set the effort term, $\xi_e e_{h,t} - \bar{\xi}_e$, equal to zero in the steady state. Alternatively, one can interpret $\bar{\xi}_e$ as home production and $\xi_e e_{h,t}$ as the effort cost.

effort and pays $p_{i,t}$, then it will receive 1 unit of the good with certainty.

[36] This means that effort is a perfect substitute for consumption goods. As shown in section 3.4, this formulation allows us to obtain sharp *analytical* insights for model properties. But it also helps to ensure that search effort is sufficiently procyclical. Without this assumption, the increase in consumption during an expansion would lower the marginal utility of consumption and this *nonlinearity* would dampen the upward effect on search effort. In our formulation, the fall in the marginal utility of consumption also lowers the cost of increasing effort (i.e., loosing goods in the process of purchases). In the full model, this acquisition cost can be in terms of both services and goods.

As is common in the literature, we follow Hansen (1985) and Rogerson (1988) and assume that the disutility of working is linear in hours worked.

FOCs. In deriving the FOCs, we substitute out $s_{h,t}$, $e_{i,h,t}$, $e_{h,t}$, and $n_{h,t}$. The Lagrange multipliers associated with the budget constraint (10a) is denoted by $\lambda_{h,t}$, the one associated with the purchases-allocation constraint (10b) by $\psi_t \lambda_{h,t}$, and the one associated with the capital accumulation equation (10c) by $\lambda_{k,h,t}\lambda_{h,t}$. That is, we express the Lagrange multipliers of these two constraints as multiples of the Lagrange multiplier of the budget constraint. The FOCs for $c_{h,t}$, $b_{h,t}$, $k_{h,t}$, $i_{h,t}$, $s_{i,h,t}$, and $W_{h,t}$ are given by

$$\psi_{h,t}\lambda_{h,t} = \xi_c \left(c_{h,t} - \xi_e(e_{h,t} - \overline{\xi}_e)\right)^{-\gamma}, \tag{11a}$$

$$\lambda_{h,t}\frac{1}{P_t} = \beta \mathbb{E}_t \left(\lambda_{h,t+1}\frac{1+R_t}{P_{t+1}}\right), \tag{11b}$$

$$\lambda_{h,t}\lambda_{k,h,t} = \beta \mathbb{E}_t \left(\lambda_{h,t+1} r_{k,t+1} + \lambda_{h,t+1}\lambda_{k,h,t+1}\left(1-\delta_k\right)\right), \tag{11c}$$

$$\begin{aligned}\psi_{h,t} &= \lambda_{k,h,t}\left(1 - \frac{\eta_i}{2}\left(\frac{i_{h,t}}{i_{h,t-1}} - 1\right)^2 - \eta_i \frac{i_{h,t}}{i_{h,t-1}}\left(\frac{i_{h,t}}{i_{h,t-1}} - 1\right)\right) \\ &\quad + \beta \mathbb{E}_t \left(\frac{\lambda_{h,t+1}}{\lambda_{h,t}}\lambda_{k,h,t+1}\eta_i \left(\frac{i_{h,t+1}}{i_{h,t}}\right)^2 \left(\frac{i_{h,t+1}}{i_{h,t}} - 1\right)\right),\end{aligned} \tag{11d}$$

$$\psi_{h,t}\lambda_{h,t}\left(\frac{s_{h,t}}{s_{i,h,t}}\right)^{\frac{1}{\varepsilon}} f^b_{i,t} = \xi_c \left(c_{h,t} - \xi_e(e_{h,t} - \overline{\xi}_e)\right)^{-\gamma}\xi_e + \lambda_{h,t}\frac{P_{i,t}}{P_t}f^b_{i,t}, \tag{11e}$$

$$\begin{aligned}&\left(\varepsilon_n \xi_n \left(\frac{W_{h,t}}{W_t}\right)^{-\varepsilon_n - 1} + \lambda_{h,t} w_t \left(1 - \varepsilon_n\right)\left(\frac{W_{h,t}}{W_t}\right)^{-\varepsilon_n}\right) \\ &\quad + \beta \mathbb{E}_t \left(\eta_W \frac{n_{t+1}}{n_t}\left(\frac{W_{h,t+1}}{W_{h,t}} - 1\right)\frac{W_t W_{h,t+1}}{W^2_{h,t}}\right) \\ &\quad = \eta_W \left(\frac{W_{h,t}}{W_{h,t-1}} - 1\right)\frac{W_t}{W_{h,t-1}},\end{aligned} \tag{11f}$$

with $w_t = W_t/P_t$. In a symmetric equilibrium, all households make the same choice. Thus, $\lambda_{h,t} = \lambda_t$, $n_{h,t} = n_t$, $W_{h,t} = W_t$, and $\psi_{h,t} = \psi_t$. This means that the last equation simplifies to

$$\begin{aligned}\left(\varepsilon_n \xi_n n_t + \lambda_t w_t \left(1 - \varepsilon_n\right)\right) + \beta \mathbb{E}_t \left(\eta_W \frac{n_{t+1}}{n_t}\left(\frac{W_{t+1}}{W_t} - 1\right)\frac{W_t W_{t+1}}{W^2_t}\right) \\ = \eta_W \left(\frac{W_t}{W_{t-1}} - 1\right)\frac{W_t}{W_{t-1}}.\end{aligned} \tag{12}$$

Combining equations (11a) and (11e) and using that $\psi_{h,t} = \psi_t$ gives

$$\frac{\xi_e}{f_{i,t}^b} + \frac{P_{i,t}}{P_t}\frac{1}{\psi_t} = 1. \tag{13}$$

Aggregate price index. The aggregate price index, P_t, is defined to satisfy[37]

$$P_t = \left(\int_{i=0}^{\infty} \left(P_{i,t} + \frac{P_t \xi_e}{f_{i,t}^b} \right)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}. \tag{14}$$

If search is not costly, then $\xi_e = 0$ and we get the usual aggregate price index, $P_t = \left(\int_0^1 (P_{i,t})^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}$. Combining our expression for P_t with equation (13) implies that $\psi_t = 1$ in each period.[38] In a symmetric equilibrium we get

$$P_t = P_{i,t} + \frac{P_t \xi_e}{f_{i,t}^b}, \tag{15}$$

where $P_{i,t}$ is the same for all i *but* less than P_t because the search cost drives a wedge between the two in our environment.[39]

Good-i demand equation. As in the standard New-Keynesian framework, goods markets are characterized by monopolistic competition. Thus, we need to derive the aggregate demand for good i. From equations (11a) and (11e), we obtain the demand for good i by household h, that is,

$$s_{i,h,t} = \left(\frac{\xi_e}{f_{i,t}^b} + \frac{P_{i,t}}{P_t} \right)^{-\varepsilon} s_{h,t}. \tag{16}$$

---

[37] In our setup, the household "aggregates" the separate goods into a bundle. If instead there is a final-goods producer who sells goods in a competitive market, then this expression for P_t is the one consistent with zero profits under the CES aggregator. This expression is also equal to the marginal cost for the household of purchasing an extra unit of $c_{h,t}$ (or $i_{h,t}$) when effort is chosen optimally and this cost is expressed in nominal units. The definition of this aggregate price index does not matter when prices are flexible. It does matter when prices are sticky, because it is used to construct the inflation measure in the Taylor rule, equation (31).

[38] If $\psi_t = 1$, then the Lagrange multiplier of the budget constraint is equal to the Lagrange multiplier of the constraint that $c_{h,t} + i_{h,t} \leq s_{h,t}$. That is, acquiring an additional unit of consumption or investment is as costly as the impact of increasing $s_{h,t}$ with one unit on the budget constraint. So another motivation of the price index is to turn the reasoning around and start with the condition that $\psi_t = 1$, which then results in our price index.

[39] In Kryvtsov and Midrigan (2013), there is also a wedge between P_t and $P_{i,t}$. In their framework, the reason is that intermediate-goods producers have to choose production before they know demand for their product which means that the final-goods producers may be constrained in their demand for some intermediate goods. Not being able to choose optimal quantities means that the final-goods firm has to charge a premium.

All households are identical and will make the same choices in a symmetric equilibrium. Demand for good i is then given by

$$s_{i,t} = \int_{h=0}^{\infty} s_{i,h,t} dh = \left(\frac{\xi_e}{f_{i,t}^b} + \frac{P_{i,t}}{P_t} \right)^{-\varepsilon} s_t. \tag{17}$$

Thus, the demand for good i (relative to aggregate demand s_t) depends not only on the relative price of good i, but also on the search cost to acquire the good, $\xi_e/f_{i,t}^b$.

3.2 Firms

There is a unit mass of firms that produce differentiated goods, indexed by $i \in [0,1]$. As in the standard New-Keynesian model, they have monopolistic power and face a demand function that is decreasing in the price chosen. In a monopoly problem without inventories, this price determines sales which in turn is exactly equal to production. In our setup, the firm has two instruments to affect demand, namely the price *and* the amount of goods it brings to the market. The latter is equal to newly produced output plus inventories carried over. Increased supply lowers the search cost for households which in turn increases demand for the firm's goods.[40]

Selling process and the goods-market friction. We assume that the total amount of goods sold, $s_{i,t}$, is given by

$$s_{i,t} = \mu e_{i,t}^{1-\nu} (y_{i,t} + (1-\delta_x) x_{i,t-1})^{\nu}, \text{ with } 0 < \nu < 1, \tag{18}$$

where $y_{i,t}$ denotes newly produced goods, $x_{i,t-1}$ the amount of goods not sold in the previous period and carried over into this period as inventory, and δ_x captures both the depreciation rate of inventories and a maintenance cost of holding inventories.[41] The level of sales increases if the firm brings more goods to the market, however, the *fraction* sold $s_{i,t}/(y_{i,t} + (1-\delta_x)x_{i,t-1})$, is strictly decreasing in the amount of goods supplied for a given household effort level. By contrast, when consumers put in more effort then total sales as well as the fraction sold will be higher with supply kept constant. Since the firm has a monopoly, it understands that it can affect $e_{i,t}$ and thus the fraction sold with its two instruments, $P_{i,t}$ and $y_{i,t}$. This is different from random search in which

[40] In appendix C, we illustrate this in a simple one-period version of our model and show how optimal household behavior is taken into account in setting these two firm instruments.

[41] Since undepreciated inventory goods are perfect substitutes for newly produced goods, it doesn't matter for model properties whether a positive δ_x captures maintenance costs or depreciation *except* for the definition of GDP; whereas maintenance costs lower GDP, depreciation does not. So to keep the model simple, we introduce only one parameter when describing the model. When calibrating the full model, we introduce a separate maintenance cost parameter, η_x, to ensure the correct results for GDP.

success of a match is taken as given.[42]

There are different ways to motivate this approach to model the selling process. Clearly, goods will only be sold if buyers put in some effort to obtain them. This effort can consist, for example, of acquiring information to figure out what to buy or shipping costs. Similarly, producers have to make goods available to be able to sell them. But there may be bottle necks in getting goods to buyers, so producing one more good does not necessarily mean selling one more good.

Even though there is just one producer in each market, the function may also be interpreted as a matching function. Specifically, consider a monopolistic firm that is a national supplier who sets the same price in different regions and/or sub-periods. But not all goods are sold, because there is uncertainty how many consumers will show up in each region or in the different sub-periods. When the firm provides more goods to the overall market, then expected sales would increase, but the fraction sold would decrease.[43] Alternatively, it may be the case that good i is not homogeneous and although the producer sets one price, there are different versions of this good (for example, a different color or a different flavor). In this case, the sell friction captures a search friction and the function in equation (18) can again be interpreted as a matching function. It is key that there is only one supplier in the market for good i, however, since we want to maintain the standard monopolistic-competition assumption of the New-Keynesian model. Consequently, there cannot be competitive search. That is, the choices of firm i affect the fraction it sells and the firm understands this.

Buyers' effort effectiveness, $f^b_{i,t}$, and the customer-finding rate by the firm, $f^f_{i,t}$, are given by

$$f^b_{i,t} = \mu(\theta_{i,t})^{-\nu} = \mu\left(\frac{e_{i,t}}{y_{i,t} + (1-\delta_x)x_{i,t-1}}\right)^{-\nu}, \tag{19a}$$

$$f^f_{i,t} = \mu(\theta_{i,t})^{1-\nu} = \mu\left(\frac{e_{i,t}}{y_{i,t} + (1-\delta_x)x_{i,t-1}}\right)^{1-\nu}, \tag{19b}$$

where $\theta_{i,t}$ represents tightness in this market.[44]

---

[42]Our setup also differs from a directed-search environment in which there are multiple sellers in the market for the same good, but they create sub-markets by setting different prices associated with different matching probabilities. In our model, there is only one supplier in the market for good i.

[43]For example, suppose that there are two regions (or two sub-periods), $j \in \{1, 2\}$, and there are two potential customers, $h \in \{1, 2\}$. The probability that customer h shows up in market j is equal to $1/2$. Suppose the firm produces two goods. Expected sales are highest when one good is provided to each sub market. Specifically, expected sales would be equal to 1.5 and the fraction sold equal to $3/4$. If the set of potential customers (effort) remains the same, but the firm would increase production to 3 units, then expected sales would increase to 1.75, but the fraction sold would fall to $7/12$.

[44]As usual, tightness is considered from the demand side. That is, a high $\theta_{i,t}$ means that buyers have to put in more effort to acquire the same amount goods, but firms will sell more for a given level of goods brought to the market.

A microfounded demand equation with a role for supply. Using the expression for the customer-finding rate, we can now write the firm's demand equation as

$$s_{i,t} \leq \left(\frac{\xi_e}{\mu \left(\frac{e_{i,t}}{y_{i,t}+(1-\delta_x)x_{i,t-1}} \right)^{-\nu}} + \frac{P_{i,t}}{P_t} \right)^{-\varepsilon} s_t. \tag{20}$$

The idea that the firm can affect the demand by its supply, i.e., $y_{i,t} + (1-\delta_x)x_{i,t-1}$, as well as the price is not new in the inventory literature. It is also present in Bils and Kahn (2000) and Coen-Pirani (2004). The difference is that they simply add an ad hoc supply component to a standard demand equation, whereas our demand equation is the outcome of a model with a goods-market friction.[45] This does not only gives us a microfounded functional form, but also makes clear that households' effort choice should be an input of this function.

Inventories. The law of motion for the end-of-period-t inventory stock, $x_{i,t}$, is given by

$$x_{i,t} \;=\; (1 - f^f(\theta_{i,t}))\,(y_{i,t} + (1-\delta_x)x_{i,t-1})\,. \tag{22}$$

Cost minimization. The production technology is given by

$$y_{i,t} = (A_t n_{i,t})^{\alpha}\, k_{i,t}^{1-\alpha}, \tag{23}$$

where $n_{i,t}$ is the amount of labor hired by firm i, $k_{i,t}$ is the capital stock held by firm i and A_t is a TFP stochastic disturbance. We assume that A_t is an $I(1)$ process with the following law of motion:

$$\ln\left(\frac{A_t}{A_{t-1}}\right) = \rho_A \left(\frac{A_{t-1}}{A_{t-2}}\right) + \varepsilon_{A,t}, \tag{24}$$

where $\varepsilon_{A,t}$ is a Normally-distributed innovation with zero mean and standard deviation σ_A. Rotemberg (2003) and Lindé (2009) argue that innovations take time to fully diffuse before reaching maximum impact which would require that $\rho_A > 0$. By contrast, when A_t is a stationary AR(1), then the maximum impact occurs instantaneously. The news

[45] The demand function in Coen-Pirani (2004) is given by

$$s_{i,t} = \tilde{\gamma}_t (y_{i,t} + (1-\delta_x)x_{i,t-1})^{\phi} \left({}^{P_{i,t}}/_{P_t}\right)^{\varepsilon}, \tag{21}$$

where $\tilde{\gamma}_t$ is an *exogenous* random variable that shifts demand. Thus, it serves the same function as our effort term, $e_{i,t}$, but our effort term is endogenous. Bils and Kahn (2000) add the same ad hoc supply-related term to a traditional demand function, $d_t\left({}^{p_{i,t}}/_{p_t}\right)$, but for their purpose it is not necessary to specify the functional form of $d_t(\cdot)$.

literature also indicates that there are TFP shocks that are associated with (further) expected growth.[46] Whether A_t is stationary or not does not matter for standard business-cycle properties.[47] However, whether a productivity shock is associated with positive or expected productivity *growth* may matter for some inventory properties which is a key result of this paper.

The cost of producing $y_{i,t}$ is given by

$$\begin{aligned} \min_{n_{i,t},k_{i,t}} w_t n_{i,t} \quad &+ \quad r_{k,t}k_{i,t} \\ &\text{s.t.} \\ y_{i,t} \quad &\leq \quad (A_t n_{i,t})^\alpha k_{i,t}^{1-\alpha}. \end{aligned} \tag{25}$$

Solving the above cost minimization problem, we have that

$$\frac{w_t n_{i,t}}{r_{k,t}k_{i,t}} = \frac{\alpha}{1-\alpha}, \tag{26}$$

and the cost of production is a linear function of output given by $\left(\frac{w_t}{\alpha}\right)^\alpha \left(\frac{r_{k,t}}{1-\alpha}\right)^{1-\alpha} \frac{y_{i,t}}{A_t}$.

In the standard monopolist firm problem, the firm chooses the price level and understands it affects the quantity sold by doing so. The relationship between the two variables is determined by the buyers' demand function. In our environment, there is a difference between quantity sold, $s_{i,t}$, and quantity supplied, $y_{i,t} + (1-\delta_x)x_{i,t-1}$. Moreover, the firm now has two instruments, namely the price and the amount of newly produced goods, $y_{i,t}$. Having two instruments, it can also control two outcomes, namely quantity sold and buyers' effort, that is, tightness.

Firm $i \in [0,1]$ solves the following optimization problem:[48]

$$\max_{\{P_{i,t},y_{i,t},x_{i,t},s_{i,t},\theta_{i,t}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{\lambda_0} \left(\begin{array}{c} \frac{P_{i,t}}{P_t} s_{i,t} - \left(\frac{w_t}{\alpha}\right)^\alpha \left(\frac{r_{k,t}}{1-\alpha}\right)^{1-\alpha} \frac{y_{i,t}}{A_t} \\ -\frac{\eta_P}{2}\left(\frac{P_{i,t}}{P_{i,t-1}} - 1\right)^2 s_t \end{array} \right)$$

s.t.

$$s_{i,t} \leq \left(\frac{\xi_e}{f^b(\theta_{i,t})} + \frac{P_{i,t}}{P_t}\right)^{-\varepsilon} s_t, \tag{27a}$$

$$s_{i,t} \leq f^f(\theta_{i,t})(y_{i,t} + (1-\delta_x)x_{i,t-1}), \tag{27b}$$

$$x_{i,t} \leq (1 - f^f(\theta_{i,t}))\left(y_{i,t} + (1-\delta_x)x_{i,t-1}\right). \tag{27c}$$

The Lagrange multipliers of the demand constraint, the sales constraint, and the

[46]See Beaudry and Portier (2006).

[47]See Christiano and Eichenbaum (1990).

[48]In appendix C, we describe a very simple partial equilibrium model for the market of good i to explain the additional degree of freedom that firms have in this type of environment.

inventories accumulation constraint are denoted by $\lambda^f_{i,d,t}$, $\lambda^f_{i,s,t}$, and $\lambda^f_{i,x,t}$, respectively.[49] The first-order conditions are given by the three constraints (which will be binding) and

$$MC_t = \left(\frac{w_t}{A_t\alpha}\right)^{\alpha}\left(\frac{r_{k,t}}{1-\alpha}\right)^{1-\alpha} = \frac{w_t n_t}{\alpha y_t}, \tag{28a}$$

$$MC_t = f^f(\theta_{i,t})\lambda^f_{i,s,t} + (1 - f^f(\theta_{i,t}))\lambda^f_{i,x,t}, \tag{28b}$$

$$\lambda^f_{i,s,t} = \frac{P_{i,t}}{P_t} - \lambda^f_{i,d,t}, \tag{28c}$$

$$\lambda_{i,x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\begin{array}{c} f^f(\theta_{i,t+1})\lambda^f_{i,s,t+1} \\ +(1-f^f(\theta_{i,t+1}))\lambda^f_{i,x,t+1}\end{array}\right)\right], \tag{28d}$$

$$\begin{aligned}&\left(\frac{\partial f^f(\theta_{i,t})}{\partial\theta_{i,t}}\right)(y_{i,t} + (1-\delta_x)x_{i,t-1})\,(\lambda^f_{i,s,t} - \lambda^f_{i,x,t}) \\ &= -\lambda^f_{i,d,t}\varepsilon\left(\frac{\xi_e s_t}{f^b(\theta_{i,t})^2}\right)\left(\frac{\partial f^b(\theta_{i,t})}{\partial\theta_{i,t}}\right)\left(\frac{\xi_e}{f^b(\theta_{i,t})} + \frac{P_{i,t}}{P_t}\right)^{-\varepsilon-1},\end{aligned} \tag{28e}$$

$$\begin{aligned}&s_{i,t} - \varepsilon\lambda^f_{i,d,t}\left(\frac{\xi_e}{f^b(\theta_{i,t})} + \frac{P_{i,t}}{P_t}\right)^{-\varepsilon-1} s_t = \eta_P\left(\frac{P_{i,t}}{P_{i,t-1}} - 1\right)\left(\frac{P_t}{P_{i,t-1}}\right)s_t \\ &+ \beta\eta_P\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\frac{P_{i,t+1}}{P_{i,t}} - 1\right)\left(\frac{P_{i,t+1}P_t}{P^2_{i,t}}\right)s_{t+1}\right],\end{aligned} \tag{28f}$$

where MC_t denotes marginal costs. Equation (28b) states that the marginal cost of producing one additional unit is equal to the expected benefit which is either in the form of selling an extra unit this period or leaving the period with an extra unit of inventories. Equation (28c) is the first-order condition for sales and it makes clear that the marginal benefit of relaxing the sales constraint, $\lambda^f_{i,s,t}$, is equal to the revenue, $P_{i,t}/P_t$, minus the cost of having to satisfying the household demand equation, $\lambda^f_{i,d,t}$. Equation (28d) specifies that the value of leaving the period with an inventory good is equal to the discounted expected value of bringing it to the market next period which could mean either a sale or again ending up in the inventory stock. Equation (28e) presents the tradeoff when changing tightness. If the firm operates at a higher level of tightness (e.g., by producing less), then this means that the fraction sold increases and the value of doing so depends on the differential benefit between selling a good now, $\lambda^f_{i,s,t}$, or keeping it as inventory, $\lambda^f_{i,x,t}$. On the other hand, a higher tightness means that the effort cost for the household increases which would mean a tightening of the firm's demand constraint. Finally, equation (28f) is the first-order condition related to $P_{i,t}$. From this equation we get a modified New-Keynesian Phillips Curve, which we will discuss next.

[49] In equilibrium, the sales constraint (27b) is identical to equation (10d) since $f^f_{i,t} = f^b_{i,t}e_{i,t}/(y_{i,t} + (1-\delta_x)x_{i,t-1})$. When facing this constraint, however, the household takes $f^b_{i,t}$ as given whereas the firm knows it affects $f^f_{i,t}$.

New-Keynesian (NK) Phillips Curve for our model with inventories. In the remainder of this section, we focus on the symmetric equilibrium.[50] If we combine equations (15), (28b), (28c), and (28f),then we get the following expression:

$$1-\varepsilon\lambda_{d,t}^{f} = \eta_P \frac{P_t}{P_{i,t}}\left(\begin{array}{c}\left(\frac{P_{i,t}}{P_{i,t-1}}-1\right)\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \\ -\beta\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\frac{P_{i,t+1}}{P_{i,t}}-1\right)\left(\frac{P_{i,t+1}}{P_{i,t}}\right)\frac{s_{t+1}}{s_t}\right]\end{array}\right), \tag{29}$$

where $\lambda_{d,t}^{f}$ is equal to

$$\lambda_{d,t}^{f} = \frac{P_{i,t}}{P_t} + \frac{1-f_t^f}{f_t^f}\lambda_{x,t}^{f} - \frac{MC_t}{f_t^f}. \tag{30}$$

If there is no goods-market friction, then $P_{i,t} = P_t$ and $\lambda_{d,t}^{f} = P_{i,t}/P_t - MC_t$, which means that we get the standard NK Phillips curve. With our goods-market friction, however, both the revenue term, $P_{i,t}/P_t$, and the cost term, MC_t, are modified. Specifically, three additional terms enter the NK Phillips curve. The first is the customer-finding rate, f_t^f. The second is the gap between $P_{i,t}$ and P_t. The third is the value of carrying a good into the next period as inventory, $\lambda_{x,t}^{f}$, which is an NPV and depends on market discount rates. Consequently, the properties of this alternative NK Phillips curve are affected by more factors than the usual NK Phillips curve.

The interpretation of the left-hand side of equation (29) is the following. Recall that this is the FOC associated with changing $P_{i,t}$. Increasing the price comes with a direct increase in revenue equal to 1 per unit of sales.[51] But the increase in the price comes with a reduction in sales and the magnitude of this effect depends on the value of ε. How this reduction in sales affects profits is first of all determined by the usual price term, but this is equal to $P_{i,t}/P_t \leq 1$ here, whereas this ratio would be equal to 1 in the standard NK model. To understand the additional two terms suppose that $f_t^f = 1/4$. A one-unit reduction in sales means that output can be lowered by $1/f_t^f = 4$ units so costs drop by $4 \times MC_t$ and not $1 \times MC_t$. This reduction of output with four units causes not only sales to be one unit less, but also a lower end-of-period inventory level of three ($= (1-f_t^f)/f_t^f$) units.

Rupert and Sustek (2019) document that the New-Keynesian model robustly predicts an increase in inflation and aggregate activity following an expansionary monetary policy shock. By contrast, the real interest rate could increase or decrease. That is, there is no real-interest-rate channel. The relationship between inflation and real activity is pinned down by the Phillips curve and the real interest rate does not show up in the traditional Phillips curve. However, the real interest rate does play a role in our New-Keynesian Phillips Curve, since it is the inverse of the marginal rate of substitu-

[50] In the symmetric equilibrium, the price of the intermediate good, $P_{i,t}$, is the same for all firms, but *not* equal to P_t because of search costs.

[51] Note that we have divided the FOC by s_t.

tion which has a direct effect on $\lambda^f_{x,t}$. Specifically, a drop in the real interest rate would increase $\lambda^f_{x,t}$ and thus reduce the markup, just like an increase in inflationary pressure does.

3.3 Monetary policy

The central bank follows a standard Taylor rule:

$$R_t = -\ln\beta(1-\Gamma_{\text{lag}}) + \Gamma_{\text{lag}}R_{t-1} + \Gamma_\pi \frac{P_t}{P_{t-1}} + \Gamma_y \left(\frac{Y_t}{\widetilde{Y}_t} - 1\right) + \varepsilon_{R,t}, \tag{31}$$

where Y_t stands for GDP in the sticky-price economy, $\widetilde{Y}_t$ for GDP in the economy with flexible prices and wages, and $\varepsilon_{R,t}$ is a monetary-policy innovation, which we assume has a mean-zero Normal distribution with a standard deviation equal to σ_R.

3.4 What determines the customer-finding rate?

The set of first-order conditions for the firm is quite large and seems complex. Fortunately, key inventory behavior such as the customer-finding rate (i.e., the inverse of the inventory-sales ratio) is determined quite intuitively by two forward looking variables. To get to that point, we first rewrite the first-order conditions into a sub-system of five equations. We also provide an interpretation of the equations. Some readers may find the algebra and the discussion a bit tedious. If that is the case, then please skip to the two propositions, since those results are quite intuitive.

Consider the following sub-set of equilibrium conditions:

$$1 = \frac{\xi_e}{f^b(\theta_t)} + \frac{P_{i,t}}{P_t} \tag{32a}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = f^f(\theta_t)\left(\frac{P_{i,t}}{P_t} - \lambda^f_{d,t} - \lambda^f_{x,t}\right) \tag{32b}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = \varepsilon\lambda^f_{d,t}\frac{\nu}{1-\nu}\xi_e\theta_t \tag{32c}$$

$$\lambda^f_{x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\begin{array}{c} f^f(\theta_{t+1})\lambda^f_{s,t+1} \\ +(1-f^f(\theta_{t+1}))\lambda^f_{x,t+1}\end{array}\right)\right], \tag{32d}$$

$$1-\varepsilon\lambda^f_{d,t} = \eta_P\frac{P_t}{P_{i,t}}\left(\begin{array}{c}\left(\frac{P_{i,t}}{P_{i,t-1}}-1\right)\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \\ -\beta\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\frac{P_{i,t+1}}{P_{i,t}}-1\right)\left(\frac{P_{i,t+1}}{P_{i,t}}\right)\frac{s_{t+1}}{s_t}\right]\end{array}\right), \tag{32e}$$

Recall that tightness, θ_t, is defined as effort over the supply of goods. So an increase in tightness leads to an increase in the customer-finding rate, $f^f(\theta_t)$, and a decrease in shopping efficiency for the buyer, $f^b(\theta_t)$. Equation (32a) is the firm's demand constraint which specifies that the firm can charge a higher price if it reduces search cost for the consumer, that is, decreases tightness by increasing supply.

From the first-order condition for the output level, equation (28b), we get that

$$MC_t - \lambda^f_{x,t} = f^f(\theta_t)(\lambda^f_{s,t} - \lambda^f_{x,t}) \tag{33}$$

The interpretation is the following. If a firm produces an extra unit of output then it costs MC_t to produce, but since it is guaranteed the value of an inventory good, $\lambda^f_{x,t}$, one can think of the *net* cost of producting as $MC_t - \lambda^f_{x,t}$. This net cost has to equal the expected net benefit which is equal to the fraction sold times the value of a sale, $\lambda^f_{s,t}$, relative to the value of an unsold good, $\lambda^f_{x,t}$. From equation (28c), we know that $\lambda^f_{s,t}$ is equal to the price minus the cost of having to satisfy the demand constraint, $P_{i,t}/P_t - \lambda^f_{d,t}$. Using this in equation (33) gives equation (32b).

Equation (32c) is a rewritten version of the firm's first-order condition for θ_t where we have also used equation (33). The right-hand side represents the cost of increasing tightness as it puts pressure on the demand constraint. The left-hand side specifies the net benefits of a sale, $\lambda^f_{s,t} - \lambda^f_{x,t}$, which according to equation (33) is related to the gap between marginal costs and the value of an inventory good. Equations (32d) and (32e) are identical to equations (28d) and (28f), respectively.

Propositions.[52] Our model is a dynamic model and determining the impact of shocks on θ_t requires a numerical solution taking into account expectations of future developments. However, this sub-system makes clear that in terms of determining the customer-finding rate all the dynamics are captured by two forward-looking variables, namely $\lambda^f_{x,t}$ and $\lambda^f_{d,t}$. That is, given values for $\lambda^f_{x,t}$ and $\lambda^f_{d,t}$, equations (32a), (32b), and (32c) determine θ_t, MC_t, and $P_{i,t}/P_t$.[53] And the behavior of tightness controls how the behavior of the key inventory variables, that is, the relationship between production, sales, and the accumulation of inventories.

Proposition 1 $\frac{\partial f^f(\theta_t)}{\partial \lambda^f_{x,t}} < 0$. *That is, an increase in the value of carrying an unsold good into the future as inventory is associated with a reduction (increase) in the customer-finding rate (inventory-sales ratio).*

Proposition 2 $\frac{\partial f^f(\theta_t)}{\partial \lambda^f_{d,t}} < 0$. *That is, an increase in inflationary pressure (relative to expected future inflation) is associated with an increase (decrease) in the customer-finding rate (inventory-sales ratio).*[54]

[52]Proofs are provided in appendix B.

[53]This convenient property depends crucially on consumption and effort being perfect substitutes. Without this assumption, the analysis would be complicated as the marginal rate of substitution between consumption and effort would no longer be constant and enter as an additional endogenous variable in this system.

[54]This proposition is only relevant when $\eta_P > 0$, that is, when prices are sticky, because $\lambda^f_{d,t}$ is a constant when $\eta_P = 0$.

Is a higher value of tightness a good thing? Before providing some intuition, it might be useful to consider whether a high customer-finding rate is "a good thing." Similarly, is having a low inventory-sales ratio attractive because it means that the same level of sales can be sustained with a lower level of inventories. In understanding the discussion of model predictions below, it is important to realize that an increase in goods-sector efficiency is not necessarily an indication that firms are doing well. An increase in the customer-finding rate may be a protective measure in response to a negative shock. Specifically, a firm might lower the supply of available goods relative to buyers' effort levels in response to some negative shocks. This would imply an *increase* (decrease) in the customer-finding rate (inventory-sales ratio). And although increased tightness is an optimal response to dampen the negative impact of the shock, the firm is still worse off.

With this in mind, let's discuss the reasons behind the two propositions.

Why does $\lambda^f_{x,t}$ matter for the customer-finding rate? A drop in $\lambda^f_{x,t}$ means that having an unsold good at the end of the period becomes less valuable. This provides an incentive for the firm to lower the inventory-sales ratio or – in the language of this paper – increase the customer-finding rate, i.e., increase tightness. The firm could do this by producing less or by inducing higher effort with a reduction in prices. By using equation (33) in equation (32d) we get that

$$\lambda^f_{x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\frac{\lambda_{t+1}}{\lambda_t}MC_{t+1}\right]. \tag{34}$$

This equation makes clear that an unsold good is an asset with future payoffs. Here the payoff is not having to produce the good in the future. And key in determining its value are changes in the marginal rate of substitution, just as is the case for other assets.[55] Consequently, in the presence of expected growth, the marginal rate of substitution will decrease, $\lambda^f_{x,t}$ will fall and firms will be less keen to hold inventories and the customer-finding rate increases. This also makes sense when we think of accumulating inventories as a form of saving which should fall when the future looks brighter than the present.

Why does $\lambda^f_{d,t}$ matter for the customer-finding rate? A decrease in $\lambda^f_{d,t}$ represent an increase in inflationary pressure (relative to expected future inflation), but $\lambda^f_{d,t}$ only fluctuates when it is costly to adjust prices. When $P_t/P_{t-1}-1$ is high (relative to expected future inflation), then firms are in the upward-sloping part of the quadratic adjustment-cost function and are held back by increasing their prices further. This rigidity causes positive nominal demand shocks to stimulate real activity in standard New-Keynesian models. That is, the inability to increase prices means that sales in

[55]In our numerical work, we find that changes in the expected value of λ_{t+1}/λ_t are much more important than expected changes in MC_{t+1}. Moreover, MC_t and $\lambda^f_{x,t}$ are positively correlated. Appendix B shows that this must be true locally around the steady state, but we find it to be true numerically in our simulations as well.

real terms must increase which in turn increases output. Such an expansion is welfare improving for the economy as a whole because it lowers firms' monopoly markups. In our model, firms can control sales not only by the prices they set, but also by output levels which affect tightness and customers' search costs. When restricted to reduce sales by increasing prices, firms increase tightness by restricting the increase in output somewhat which would increase search costs for customers which in turn would dampen the increase in demand, just as an increase in the price would. Although the flexibility to affect tightness restricts the increase in output, it still increase when $\lambda^f_{d,t}$ falls. That is, the additional flexibility does not undo the expansionary effect of nominal demand shocks on output in the presence of sticky prices.[56] There is another angle to describe this model prediction intuitively. A positive nominal demand shock implies a persistent reduction in markups. In response to lower profitability, firms reduce the inventory-sales ratio, that is, increase the customer-finding rate (sell fraction).

3.5 Comparison with the literature

The previous section described two channels through which aggregate shocks affect the behavior of inventories. The first is through a valuation effect of inventories where it is important to realize that an inventory good is a durable asset. The second is through changes in the markup which changes the desirability of "excess" production, i.e., inventory accumulation.

We think that the first channel is new.[57] As shown in the next section, it is this channel that makes it possible for our general-equilibrium model to match key inventory, production, and sales facts in response to supply shocks.

The "markup channel" is not new and captures a mechanism similar to what is referred to as the "stockout-avoidance motive" in the literature.[58] The usual setup assumes that the demand for a firm's good is subject to idiosyncratic shocks *and* that distributors have to set the price and production level before that shock is known. If a specific good turns out to have a positive preference shock, then the price set is too low to clear the market and a rationing rule is imposed. On the other hand, the price would be too high following a negative shock and not all available goods will be sold, that is, the distributor accumulates inventories.

Now suppose that prices are somewhat sticky *and* aggregate shocks are known before the firm sets its price and production level. In response to a positive demand shock, the real markup would fall because prices are sticky. This implies a reduction

[56]In our subsystem, it can be shown analytically that for small changes around the steady-state marginal costs increase together with tightness when $\lambda^f_{d,t}$ falls, which in turn implies that output must increase. See appendix B. Numerical results indicate that this is a robust finding for larger changes as well.

[57]In fact, the marginal rate of substitution is typically assumed to be constant in the macro-inventory literature.

[58]Examples of such a framework can be found in Kahn (1987), Wen (2008), and Kryvtsov and Midrigan (2013).

in the value of inventory goods which means that the relative cost of oversupply to a stockout increases. The firm will therefore lower the supply relative to expected sales, which implies a reduction in the inventory-sales ratio, or – in our terminology – an increase in the customer-finding rate.[59]

At first sight, this setup with firm-specific idiosyncratic shocks looks quite different than ours. It has a distribution of preference shocks, a difference in timing regarding when good-specific preference shocks and aggregate shocks are known, and no role for buyers' effort choices. By contrast, we have a representative firm and only aggregate shocks. However, one could interpret the idiosyncratic preference shocks as a matching friction like the one we adopt. Specifically, the larger the variance of the idiosyncratic-shock distribution, the more likely that the distributor will face a stockout or inventory accumulation. And just as the matching friction creates a gap between the price of the individual firm and the aggregate price level, there is a gap between the price set by the distributor and the price set by the final-goods producer.[60] In terms of the calibration, the stockout approach needs information on the cross-sectional distribution of the idiosyncratic preference shocks. We have to take a stand on the matching function and how the effort choice affects the household. Our simpler representative-firm approach may be more suitable to be incorporated in larger models. But the idiosyncratic-stockout approach allows one to study how different aggregate shocks affect firms with different idiosyncratic-shock realizations differently.

Kryvtsov and Midrigan (2013). Despite the importance of inventories for business-cycle fluctuations, there are relatively few papers that develop general-equilibrium business-cycle models that incorporate a role for inventories. A notable exception is Kryvtsov and Midrigan (2013) (KM) which incorporates the stockout-avoidance setup. But in contrast to the literature, changes in markups are endogenous and occur because of sticky prices, as is the case in our model.[61] They show that their model is consistent with key inventory facts in response to monetary-policy shocks. But they also point out that their model is only consistent with productivity shocks when prices are flexible. But our model's predictions are consistent with key inventory, production, and sales facts in response to productivity shocks as well, both when prices are flexible

[59]In this type of framework, there is a negative correlation between sales and inventory accumulation at the *firm* level even though – as shown in Kryvtsov and Midrigan (2013) – it is positive at the aggregate level. The latter is consistent with the data. Unfortunately, we don't know what the sign of the correlation is at the firm level and it may very well have a different sign than the one at the aggregate level. The difficulty of determining the sign at the firm level is that one would need data for the volume, not the value of inventories. But it seems not implausible that an *individual* firm facing a sudden temporary drop in *firm-specific* demand will see its sales drop and inventories increase.

[60]Note that this is true even though there is a symmetric equilibrium in which all individual goods sell at the same price.

[61]By contrast, Coen-Pirani (2004) considers exogenous changes in the markup and Bils and Kahn (2000) consider an environment in which firms face an exogenous random price and only choose their production levels.

and when they are not.[62]

There are two reasons for this. The first is that the KM model monetary policy with an exogenous monetary-supply rule and we adopt a standard NK approach with a standard Taylor rule. The Taylor rule ensures that the central bank responds to inflationary pressure. In basic versions of our model, this ensures divine coincidence, that is, model outcomes for real variables when prices are sticky are the same as the corresponding outcomes when prices are flexible. But when we add the usual additional features such as sticky wages and investment-adjustment costs, then our model no longer satisfies divine coincidence. The reason that our model *robustly* predicts that the customer-finding rate is procyclical is that the value of an unsold good is consistently countercyclical, because we have an empirically realistic representation for the law of motion of TFP. That is, (saving through) investment accumulation is less attractive during an expansion which means that – relative to the increase in effort – firms bring less goods to the market which means that the customer-finding rate increases. This key new mechanism will be discussed in more detail in the next section.

3.6 Model predictions

A business-cycle model with inventories would have to be able to generate the following main inventory, production, sales properties: (1) the customer-finding rate is procyclical, (2) output is more volatile than sales, (3) the inventory stock is procyclical, and (4) investment in inventories is procyclical. If the first prediction is satisfied, then the model also correctly predicts that the inventory-sales ratio is countercyclical. The model will replicate empirical facts two and three if the first one is satisfied *but* the customer-finding rate, f_t^f, is not too volatile. If fluctuations in f_t^f are too large, i.e., more than what is observed in the data, then sales will increase by more than output and inventory fluctuations will be countercyclical. Both properties are counterfactual. Changes in the inventory-stock level are temporary following a demand shock. This means that *investment* in inventory must flip sign. However, we find that investment in inventories is procyclical as well since the initial response in the inventory-stock level dominates the subsequent gradual response back to its pre-shock value.

We will show that the model can generate a procyclical customer-finding rate in response to both a TFP and a monetary-policy shock. That the customer-finding rate is procyclical when there is a positive demand shock is not that surprising. However, one might expect that an increase in the supply of goods would increase by more than

[62] Two other papers also consider inventory behavior in response to TFP shocks. Bils and Kahn (2000) develop an ingenious (but nontrivial) mechanism that affects the markup in a partial equilibrium environment in which firms take the price as given. Results are driven by changes in the markup. We have a general-equilibrium framework and TFP shocks would leave the markup unaffected when prices are fully flexible when the markup is appropriately defined as in equation (30). McMahon (2011) introduces a delay between the production of a good and its sale although the firm can shorten the delay at a cost. The necessary delay ensures that (i) output is more volatile than sales following an increase in production and (ii) the investment stock is procyclical. However, the inventory-sales ratio increases when TFP increases, at least around the steady state.

buyers' effort when TFP increases, which would mean that the customer-finding rate is decreasing during a TFP-driven expansion.[63] This turns out to be not true: model predictions are also consistent with the observed procyclicality (countercyclicality) of the customer-finding rate (inventory-sales ratio) when business-cycle fluctuations are driven by TFP shocks. Since these results are a bit surprising, we discuss the results for the TFP shock first and for a wider range of parameterizations.[64] In this section, we will focus mainly on the qualitative properties of the model and the role that different model elements play.[65] When we discuss the full model with services, then

[63]In appendix C, we consider a one-period version of our model in which the value of an unsold good at the end of the period, λ_x^f, is a fixed parameter. In this version of the model, the customer-finding rate remains constant following a productivity increase when prices are flexible. The reason is that the productivity improvement also increases resources for households which leads to higher demand and an increase in search effort. However, when firms are restricted in lowering prices, then demand is limited and the customer-finding rate will fall. This would generate a negative comovement between the customer-finding rate and aggregate activity, which is the opposite of what is observed in the data. Thus, variations in the value of $\lambda_{x,t}^f$ are essential to match observed inventory facts.

[64]Kryvtsov and Midrigan (2013) mainly focus on monetary-policy shocks, but show that the inventory-sales ratio is *pro*cyclical in response to TFP shocks in the presence of sticky prices which turns out to be not true in our model.

[65]In section 4, we motivate an extensive calibration strategy for the analysis of the full model. We use the same parameter values here except for the parameter ν which is a key parameter as it controls the curvature of the matching function. The idea is to set ν for the goods-only model so that model properties of the goods-only resemble the properties of the calibrated full model. There is a crucial difference in the two models in that the goods-sector in the full model is responsible for a larger share of investment (as it is in the data) making it more volatile than the goods-sector in the goods-only economy. This is why a different value for ν is necessary in the goods-only version of our model. Specifically, we set ν equal to 0.866 which ensures that the change of the goods-sector customer-finding rate on impact relative to the change in production in response to a monetary-policy shock is equal to the one for the complete model, evaluated at the benchmark parameter values for the full model. That is, it isn't too volatile. As the customer-finding rate is less volatile following TFP shocks, this value for ν will ensure empirically plausible results for a TFP shock as well. This value for ν is substantially higher than the calibrated values obtained for the full model. To satisfy the very diligent reader we will provide an explanation. Sub-system (32) makes clear that tightness (and thus the customer-finding rate) only depends on the value of an unsold good, $\lambda_{x,t}^f$, and inflationary pressure, $\lambda_{d,t}^f$. In the complete model, there is a similar sub-system which determines tightness in the goods sector as a function of $\lambda_{x,t}^f$, inflationary pressure in the goods sector, $\widetilde{\lambda}_{g,t}^f$, and inflationary pressure in the service sector, $\widetilde{\lambda}_{s,t}^f$. The subsystems of the two models contain other variables, but not production levels. Suppose we use the same parameters for both the goods-only and the complete model. It turns out that the responses of the value of an unsold good is similar in the two economies. This would imply similar responses for tightness and the customer-finding rate. However, production in the goods sector is more volatile in the complete model, since it is concentrated on producing investment goods. But what is key for inventory accumulation is the change in the customer-finding rate *relative* to the change in production. For completeness, one more somewhat less important issue. It is true that the response of the inflationary-pressure term, $\lambda_{d,t}^f$, is stronger for goods than for services, which makes sense given that investment relies more on goods and investment is more volatile than consumption. This does imply a somewhat higher response in the customer-finding rate. However, it isn't proportional to the higher production response. One reason is that the goods sector benefits from paying the same wages and rental rate as the less expanding service sector during an expansion. This is basically the opposite

we discuss quantitative properties in more detail, which will include a comparison of model moments of HP-filtered data and their empirical counterpart. Here we restrict ourselves to a discussion of Impulse Response Functions (IRFs).[66]

3.6.1 Responses to a TFP shock

In this section, we highlight the properties of the model for different parameterizations. At each step we add a feature that is typically included in New-Keynesian models and discuss how this affects model predictions for the behavior of inventory, production, and sales. A key parameter for the TFP process is the autoregressive parameter for productivity growth, ρ_A, which we set equal to 0.35, at which value the model matches the observed serial correlation of TFP growth adjusted for capacity utilization.[67]

Flexible prices. With flexible prices, i.e., when $\eta_P = 0$, any inflationary or deflationary pressure has no effect on the firm's demand constraint. From the system of equations (32), we know that tightness, θ_t, and the customer-finding rate, $f^f(\theta_t)$, would remain constant if the value of an unsold good, $\lambda^f_{x,t}$, would remain constant as well. Why? With flexible prices, there are no reasons for the firms to change either $P_{i,t}/P_t$ or tightness, θ_t. The firm would simply scale up production.[68] And consistent with the demand equation, buyers simply scale up effort with their increased income. The value of $\lambda^f_{x,t}$, however, would not be constant. It falls following a positive productivity shock, because consumption is expected to increase which lowers the marginal rate of substitution, which in turn lowers the value of bringing goods into the future.[69] Consequently, the customer-finding rate (inventory-sales ratio) is procyclical (countercyclical) as observed in the data.[70]

---

of what is displayed in figure 8, since an increase in inflationary pressure in the goods sector increases marginal costs as a function of $P_{g,t}/P_t$.

[66]All model properties are based on a first-order perturbation approximation.

[67]See section 4.6 and in particular footnote 94 for details. In appendix E.1, we discuss model predictions when TFP is instead assumed to be a stationary process.

[68]What will happen with P_t depends on the rest of the model and in particular on whether monetary policy responds to de/inflationary pressure.

[69]As shown in equation (34), the value of $\lambda^f_{x,t}$ is equal to the expected discounted value of future marginal costs. In the standard NK model with flexible prices, marginal costs are a constant fraction of the price level where the gap is determined by the elasticity of substitution of the different goods. As indicated in subsystem (32), the determination of marginal costs is a bit more complicated and marginal costs are affected by changes in $\lambda^f_{x,t}$. But MC_t falls when $\lambda^f_{x,t}$ falls (keeping $\lambda^f_{d,t}$ constant). This is shown analytically in appendix B for small shocks around the steady state and found numerically for large shocks. Thus, a reduction in the discount factor leads to a fall in $\lambda^f_{x,t}$, which leads to a fall in marginal costs, which in turn leads to a further fall in $\lambda^f_{x,t}$.

[70]Key is that the IRF of consumption is hump-shaped. Ramey (2016) shows that estimated consumption IRFs display such a hump-shaped pattern for several empirical specifications. Our simple model can generate hump-shaped consumption IRFs for *some* parameter values when A_t is stationary, but does so robustly for our benchmark $I(1)$ specification. An alternative way to generate a

Sticky prices. Figure 3 displays the results when prices are sticky, there are no investment adjustment costs, wages are not sticky, and the central bank does not respond to the output gap, i.e., $\Gamma_y = 0$. The results are almost the same as when prices are fully flexible. The reason is that our model approximately satisfies divine coincidence for these parameter values. That is, the central bank sets monetary policy according to a standard Taylor rule and accommodates a positive TFP shock and both the inflation rate and the output gap are basically unchanged. Divine coincidence does not hold exactly and there is a small increase in inflation equal to a few basis points. This would mean that $\lambda^f_{d,t}$ falls slightly which increases the customer-finding rate a bit further. In terms of the output gap, the deviation from divine coincidence is so small that it isn't visible in this panel.[71]

Model predictions are consistent with key inventory and business-cycle facts. Regarding the inventory facts, the customer-finding rate is procyclical, inventories are procyclical, and output is more volatile than sales. Since the stock of inventories is monotonically increasing, investment in inventories is procyclical as well. These results are driven by the fall in $\lambda^f_{x,t}$ which in turn is driven by the expected increase in consumption. The (small) increase in inflation pushes tightness in the same direction.

On impact, the customer-finding rate increases with 7.67 basis points. At that point, output has increased by 1.08%. Using the 0.144 ratio from table 1, this output increase would imply an increase in the customer-finding rate of 16.3 basis points. So the model somewhat underpredicts the relative response of the customer-finding rate. But the 0.144 ratio is based on HP-filtered data whereas the IRFs display raw reponses.[72]

There is one prediction that is not satisfactory and that is that investment actually drops on impact.[73] Investment-adjustment costs and sticky wages will push up the initial investment response. How these standard model features affect the properties we are interested in will be discussed next.

Adding investment-adjustment costs. With investment adjustment costs, i.e., $\eta_i > 0$, the model generates a positive investment response on impact.[74] The results are shown in figure 4. The response of the customer-finding rate is now stronger which

hump-shaped consumption IRF is to introduce habits. See Fuhrer (2000).

[71]For comparability, we keep the scale of the vertical axis the same in the different experiments.

[72]And comparing the relative magnitudes of the raw numbers would be even less appropriate at longer horizons because then a larger part of the output response would be part of the trend. But the numbers indicate that model predictions are not unrealistic quantitatively. Section 4.7 provides a more detailed discussion of quantitative results for the calibrated full model.

[73]And this drop would be bigger if the HP-filtered residual of investment would be considered.

[74]Consumers like to smooth consumption. Given the expected further increase in TFP, the optimal response is to lower investment initially. This would ensure that the consumption response is close to its long-run permanent increase on impact. The initial drop in investment, then requires steep increases in investment in subsequent periods to ensure that the capital stock adjusts appropriately to the permanent increase in TFP. This time path for consumption would be costly to implement, however, in the presence of investment-adjustment costs.

is consistent with the sharper drop in $\lambda^f_{x,t}$ which in turn can be explained by the more gradual increase in consumption which implies a lower discount factor during the transition.

Adding sticky wages. One might think that the presence of sticky wages plays a key role for firms' inventory choices. That is, shouldn't firms produce more and accumulate inventories when productivity is high and wage increases are restricted because of wage adjustment costs? But neither wages nor the sticky-wage parameter, η_W, appear in our subsystem that determines the tightness and the customer-finding rate. So why is there no effect of sticky wages on inventory accumulation unless there is an indirect effect through $\lambda^f_{x,t}$ and/or $\lambda^f_{d,t}$? In the NK model, the key variable is the level of marginal costs relative to the price level. In the textbook NK model, this markup is constant when prices are fully flexible and in response to a TFP shock also when prices are sticky but the model satisfies divine coincidence.[75] The situation is a bit more complicated in our setup, but still follows the logic of the NK environment. If wages adjust slowly to increased productivity levels, then firms would adjust the scale of their operations upward and they would do so up to the point where marginal costs are again appropriate given the values of a sold good, $P_{i,t}/P_t$, and an unsold good, $\lambda^f_{x,t}$. In that situation, overall activity is higher, but the optimal level of tightness will only be different if $\lambda^f_{x,t}$ or $\lambda^f_{d,t}$ take on different values. Consequently, inventory accumulation will also only be different if wage stickiness affects the behavior of $\lambda^f_{x,t}$ or $\lambda^f_{d,t}$.

Consistent with the standard NK model, the initial response of the aggregate economy to a TFP shock is indeed stronger in the presence of sticky wages. Moreover, the output gap is now substantially positive. Consequently, $\lambda^f_{x,t}$ and $\lambda^f_{d,t}$ will behave differently which in turn will affect inventory behavior.

Figure 4 displays the results. Specifically, adding sticky wages to the model reduces the magnitude of the increase in the customer-finding rate. This is beneficial because it ensures that the output response is quite a bit stronger than the sales response as is also observed in the data. There are two reasons why the customer-finding rate increase is dampened. Inflationary pressure is reduced with sticky wages. Moreover, with sticky wages there is initially a sharper increase in real activity which implies that consumption increases by more on impact, but then grows at a slower pace. This means that the marginal rate of substitution and, thus, the value of an unsold good drop by less. Both the smaller increase in inflation and the smaller drop in $\lambda^f_{x,t}$ imply a smaller increase in tightness as was shown with our system of equations (32).

With sticky wages, the deviation from divine coincidence increases. Inflationary pressure is still small; except for a 13 basis points increase on impact it is less than three basis points during the transition. However, there is now a nontrivial output gap which starts out at roughly 1.9% of flexible-price output on impact. Since $\Gamma_y = 0$ for

[75] This constant markup is a function of ε only.

this parameterization, this positive output gap has no direct consequences for monetary policy.

Adding a monetary policy response to a positive output gap. Under divine coincidence, there would be no inflationary pressure because the deflationary pressure due to increased supply is offset by monetary stimulus. In the last example, there is some inflationary pressure. That is, the accommodation of the central bank is too strong. We can control this by generalizing the Taylor rule and letting the central bank raise the nominal interest rate in response to a positive output gap, that is $\Gamma_y > 0$. The results are shown in figure 6. The customer-finding rate responds now less sharply due to the central bank providing less accommodation to the TFP-driven expansion. Specifically, it increases by 16.6 on impact instead of 17.9 basis points. There is still some inflationary pressure, for this value of Γ_y. When we increase Γ_y to 0.10 then there is more deflationary than inflationary pressure following the shock.[76] The customer-finding rate is still clearly procyclical with a peak response of 14.2 basis points instead of 16.6.

3.6.2 Responses to a monetary-policy shock

A monetary-policy shock affects the economy like a demand shock when prices are sticky, which in turn leads to an increase in buyers' effort relative to the amount of goods that firms bring to the market. Consequently, the customer-finding rate increases. From the firms' perspective, an increase in the customer-finding rate raises revenues just as an increase in the price does. Since the results are less surprising for a monetary-policy shock, we only present the results for the last parameterization which includes all features typically present in New-Keynesian models.

The results are shown in figure 7. The size of the shock is chosen to generate a 25 basis point reduction in the (annual) nominal interest rate. For these parameter values, the initial output response is substantially stronger than the sales response, but after a while the responses become quite similar. One can adjust ν, that is, the curvature of the function to affect the relative volatility and choose it such that inventory facts replicate unconditional data properties (taking into account sampling uncertainty) for both types of shocks qualitatively. That calibration is carried out for the full model.

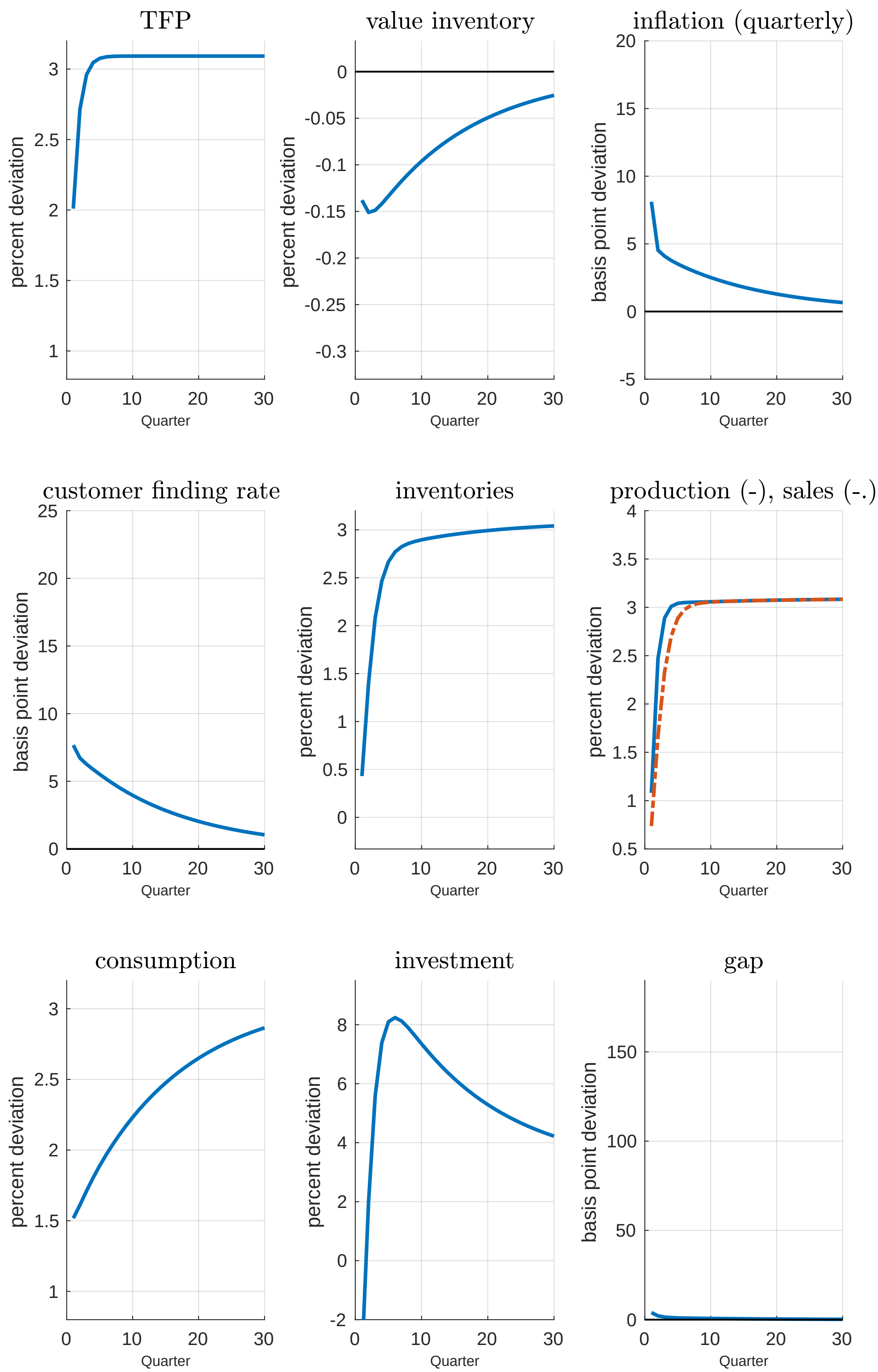
[76] It is not possible to get a zero inflation response in each period by adjusting just one parameter. At $\Gamma_y = 0.1$, there is initially deflationary pressure with a peak response of minus 6.5 basis points. This indicates that the central bank's accommodation is reduced by too much, i.e., Γ_y is perhaps too high. But then this is followed by some minor inflationary pressure of at most 2.21 basis points. So it seems reasonable to conclude that the central bank "roughly" keeps inflation at target with this value of Γ_y.

3.7 TFP versus monetary-policy shocks

At the beginning of this section, we highlighted four empirical findings that we would like our model to generate. And we have shown that the model can do this for both TFP and monetary-policy shocks. Nevertheless, there are some differences between the responses to the two shocks that we want to point out. First, the shape of the IRF for inventories is different. following a TFP shock, the inventory stock increases and then continues to increase which implies that both inventories and the investment in inventories will be clearly procyclical.[77] By contrast, following a monetary-policy shock the IRF for inventories displays a large increase on impact after which the response gradually decreases. Investment in inventories is still procyclical because the sharp initial increase dominates the subsequent gradual decreases. Second, the dynamic pattern of the customer-finding rate *relative* to the one for output is quite different for the two shocks. Following the monetary-policy shock, the shape of the IRF of the customer-finding rate resembles the IRF of output. By contrast, the response of the customer-finding rate following a TFP shock is relatively short lived with most of the movement happening in the first couple periods, whereas the output response is – by construction – long lived. In section 4.7.2, we will show that this difference is helpful in establishing the relative importance of supply versus demand shocks when we challenge the model to explain an additional empirical finding.

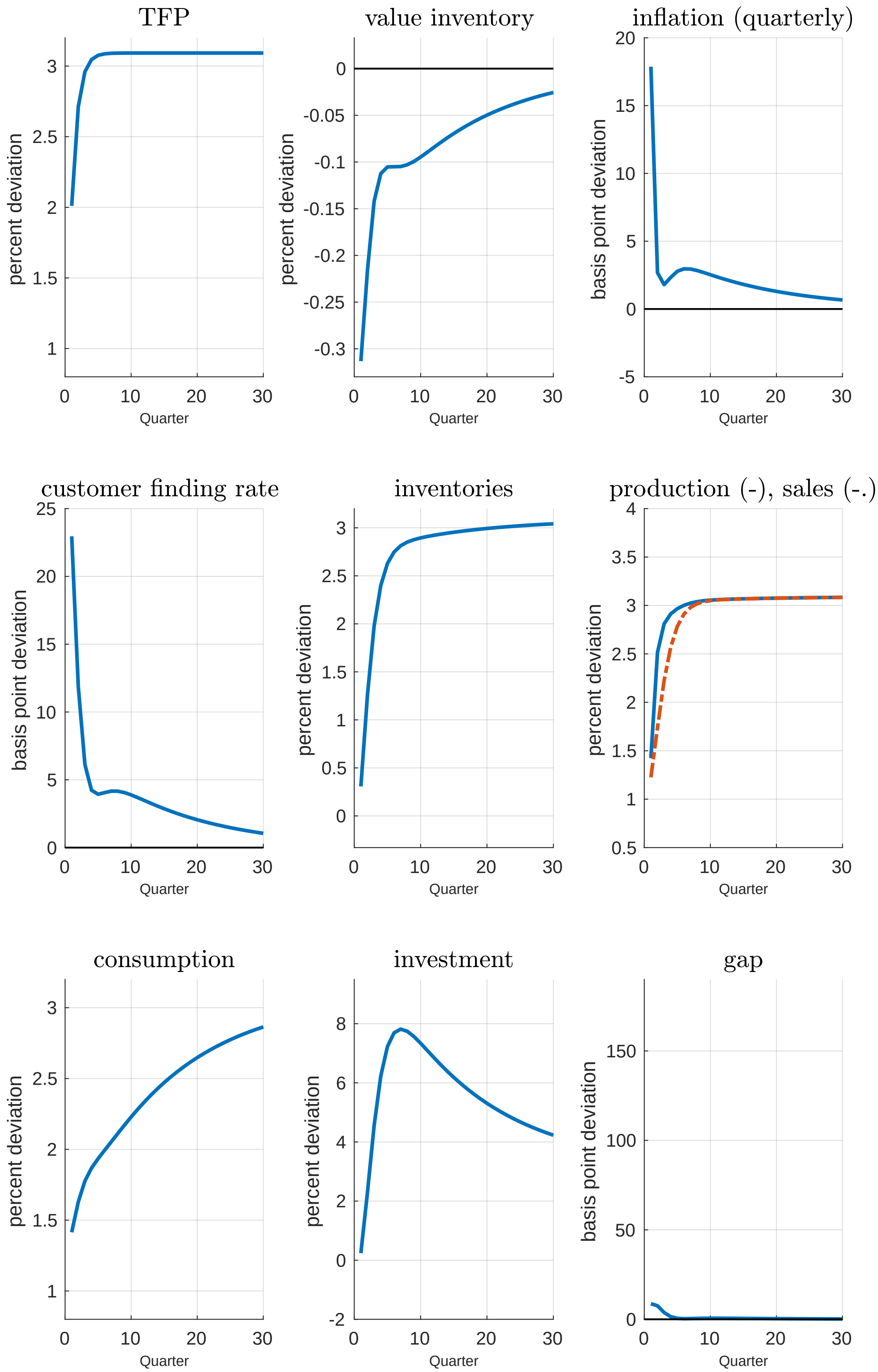
[77] Even if TFP is assumed to be stationary with the usual autoregressive coefficient equal to 0.95, then the inventory stock displays a hump before it starts its decline towards it pre-shock level.

Figure 3: TFP shock; $\eta_P > 0, \eta_i = 0, \eta_W = 0, \Gamma_y = 0$



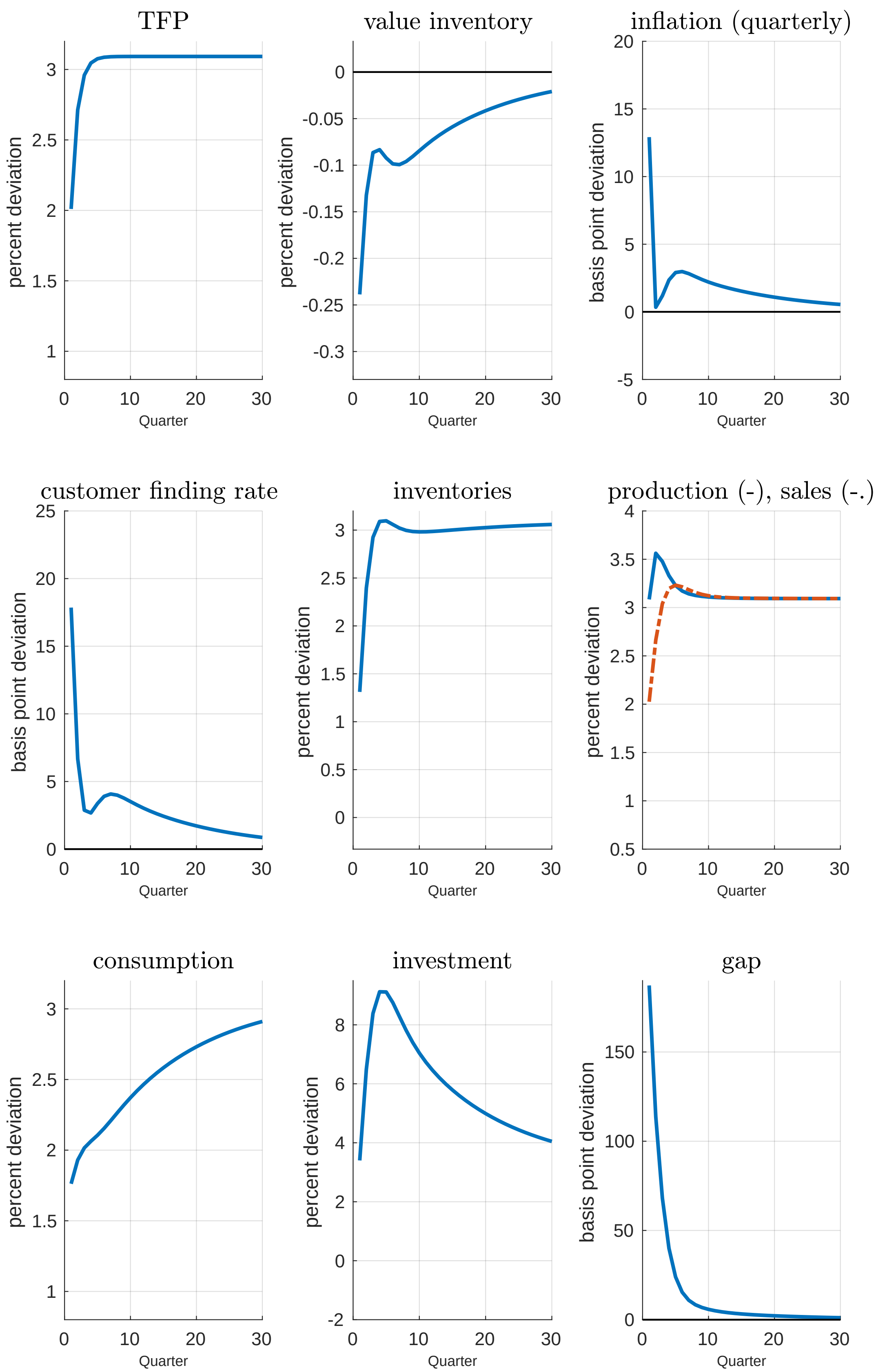
Notes. Impact of a TFP shock with sticky prices ($\eta_P = 10$), but no sticky wages nor investment adjustment costs. The value of Γ_y is of very little importance here, since the output gap is approximately zero.

Figure 4: TFP shock: plus investment adjustment costs; $\eta_P > 0, \eta_i > 0, \eta_W = 0, \Gamma_y = 0$



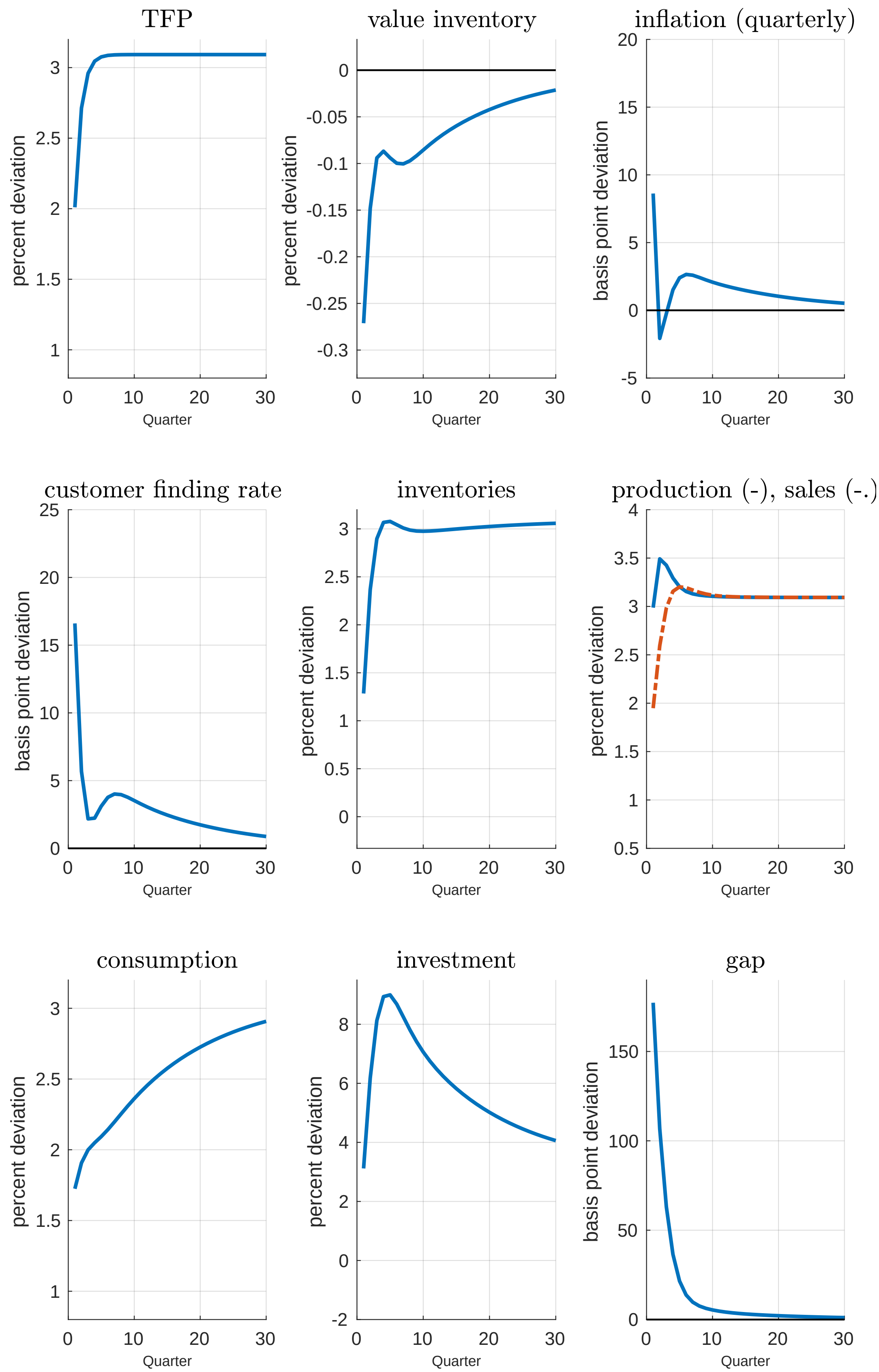
Notes. Impact of a TFP shock with sticky prices ($\eta_P = 10$) and investment adjustment costs ($\eta_i = 0.1$). No sticky wages. The value of Γ_y is of very little importance here, since the output gap is approximately zero.

Figure 5: TFP shock: plus sticky wages; $\eta_P > 0, \eta_i > 0, \eta_W > 0, \Gamma_y = 0$



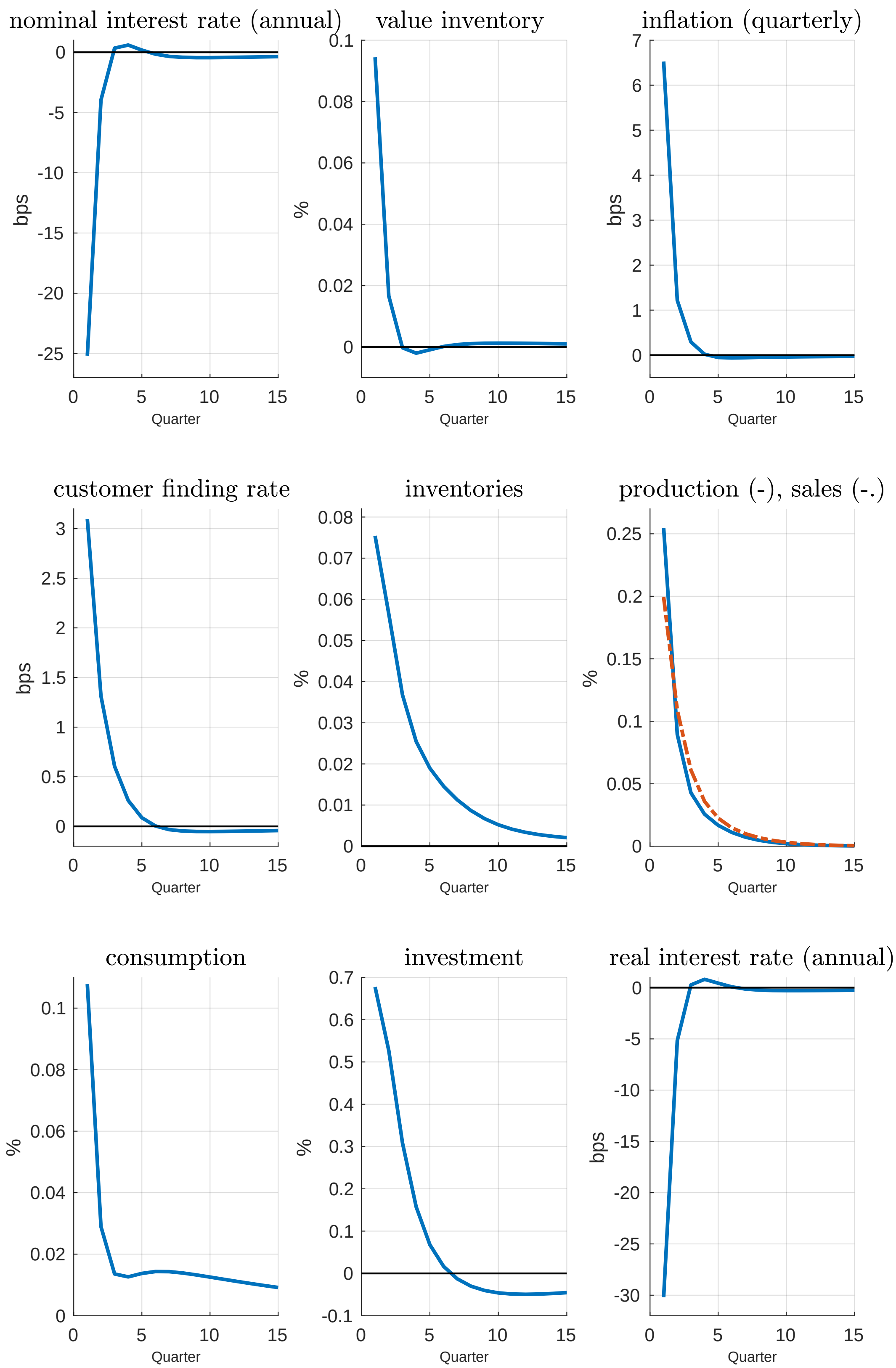
Notes. Impact of a TFP shock with sticky prices ($\eta_P = 10$), sticky wages ($\eta_W = 10$), and investment adjustment costs ($\eta_i = 0.1$.)

Figure 6: TFP shock: plus output gap response; $\eta_P > 0, \eta_i > 0, \eta_W > 0, \Gamma_y > 0$



Notes. Impact of a TFP shock with sticky prices ($\eta_P = 10$), sticky wages ($\eta_W = 10$), and investment adjustment costs ($\eta_i = 0.1$). Taylor rule includes positive response to output gap ($\Gamma_y = 0.1$).

Figure 7: monetary-policy shock; $\eta_P > 0, \eta_i > 0, \eta_W > 0, \Gamma_y > 0$



Notes. Impact of a monetary-policy shock with sticky prices ($\eta_P = 10$), sticky wages ($\eta_W = 10$), and investment adjustment costs ($\eta_i = 0.1$). Taylor rule includes positive response to output gap ($\Gamma_y = 0.1$).

4 Model with goods and services

This section starts by motivating an expanded version of our model that includes a service sector. Next, it develops such a model and discusses its properties.

4.1 Why is it important to include services?

Given the quantitative importance of inventory investment for business-cycle fluctuations, it is important to have a business-cycle model with inventories that fits empirical business-cycle facts for inventories. The model developed in the previous section has only one type of production sector, namely one that that produces goods which end up in inventories if they are not sold. Of course, there are many business-cycle models in which there is only one type of good. But that is not quite satisfactory here. Matching inventory facts means matching the relative volatility of production and sales, the procyclicality of (investment in) inventories, and the countercyclicality of the inventory-sales ratio, i.e., the procyclicality of the customer-finding rate (fraction sold). But these are empirical facts about the goods-producing sector. It might perhaps be possible to consider it as a model for the aggregate economy if the depreciation rate of unsold goods is considered an average of the 100% depreciation rate of "unsold" services and the one for goods. But the empirical facts presented in section 2 regarding the joint behavior of inventories, production, and sales are for the goods sector, not for a combination of the goods and service sector.

Whereas there are no inventories in the service sector, providers of services are also likely to face frictions in finding buyers as documented in figure 2. It is easy to adapt our model to the service sector. All that is needed is to set the depreciation rate of unsold services to 100%. There would be one important notational change. For a firm that produces services, the variable $y_{i,t}$ would no longer be *actual* output, but the amount of services that the firm *potentially* could supply given the amount of labor and capital it has in place. And the customer-finding rate of the service sector would then be equal to actual sales relative to this potential level of sales. Similar to the sell friction in the goods market, an increase in the level of potential sales would increase expected sales, but reduce the fraction sold.

It is important to understand how sell frictions affect cyclical activity in the service sector given the importance of the service sector in modern economies. It is also important to understand how sell frictions in the two sectors interact. In the previous section, we learned that cyclical fluctuations in $\lambda^f_{x,t}$, i.e., the value of an unsold good is key in driving the cyclicality of the customer-finding rate in the goods sector. In the service sector, $\lambda^f_{x,t}$ would be equal to zero. Does this mean that the cyclical behavior of the customer-finding rate in the service sector is quite different? Are there spillovers between the two sectors? That is, does an increase in the customer-finding rate in the goods-sector have implications for the service sector? Those are the key questions that we shed light on in this section.

We want to answer these questions in a transparent insightful way. That is, using a simple approach to incorporate both goods and services. The simplicity is achieved by imposing certain restrictions on preferences for different types of consumption and on how different types of investment increase the capital stock. These assumptions will restrict the *relative* quantitative responses of sub-components such as the responsiveness of consumer goods relative to consumer services. But they avoid the complexity of a more flexible setup.

4.2 Key assumption and its implications

To ensure that adding a service sector is transparent, we assume that goods and services enter the utility function and the capital accumulation equation in a Leontief manner. The advantage of this assumption is that the implied demand functions for goods and services remain relatively simple and are as close as possible to the one from the economy with only goods. Nevertheless, there are some differences and in particular, the relative price of goods versus services matters.

Individual goods and individual services are aggregated using the usual Dixit-Stiglitz aggregator to a "goods" aggregate and a "services" aggregate. A goods-producing firm and a services-providing firm face a demand function with the same components, namely, its own relative price and its own customer-finding rate. Key in our framework is that the firm can affect demand and the customer-finding rate not only by the price it sets but also by its "capacity." For a goods-producing firm this is simply actual production plus the inventory stock. For a firm producing services, this is the maximum amount of services it could provide given the amount of capital it has in place and the number of workers it has hired. Thus, a firm providing services faces the usual production function, but its functional value describes *potential* sales. The Leontief assumption is directly responsible for the relatively simple functional form of the firms' demand functions. The choices for the two firms can be quite different, however, because unsold goods end up in inventories for goods-producing firms and "not-utilized" capacity has zero value for those producing services. In particular, we will show that the customer-finding rates of the two sectors could move in the same, but also in opposite directions.

We assume that there is just one homogeneous consumption good, but we interpret it as a mixture of durables and non-durables. Under the assumption that the benefit flow from this stock is linearly related to the stock, we can include the *stock* of this consumption good in the Leontief structure. The calibrated depreciation rate will take into account that this good is mixture of both durable and non-durable goods. This complication does not affect the key properties that we are interested in like the behavior of customer-finding rates in the two sectors or inventory facts. But it does allow us to make *expenditures* on consumer goods more volatile than purchases of consumer services, as is observed in the data.[78]

[78] The Leontief structure imposes a relationship between the stock of goods and services, so services are as volatile as the *stock* of durables.

As in the goods-only model, we assume that the effort cost is a perfect substitute with consumption, but the cost of searching for goods and services could be in the form of either goods or services, or both.

The following set of equations captures this setup:

$$c_t \leq \min\left\{\frac{\bar{c}_{g,t}}{\omega_{g,c}}, \frac{c_{s,t} - \Upsilon_s(\xi_e e_t - \overline{\xi}_e)}{\omega_{s,c}}\right\}, \tag{35a}$$

$$i_t \leq \min\left\{\frac{i_{g,t}}{\omega_{g,i}}, \frac{i_{s,t}}{\omega_{s,i}}\right\}, \tag{35b}$$

$$\bar{c}_{g,t} = (1 - \delta_c)\,\bar{c}_{g,t-1} + c_{g,t} - \Upsilon_g(\xi_e e_t - \overline{\xi}_e). \tag{35c}$$

Here, c_t denotes aggregate consumption, $\bar{c}_{g,t}$ the stock of consumption goods, $c_{g,t}$ the *purchases* of consumption goods, $c_{s,t}$ the purchases of consumption services, i_t denotes aggregate investment, $i_{s,t}$ investment goods, $i_{s,t}$ investment services.[79]

The weights $\omega_{g,c}$, $\omega_{s,c}$, $\omega_{g,i}$, and $\omega_{s,i}$, are the usual Leontief weights, but $\omega_{g,c}$ also takes into account that the stock of consumption goods, $\bar{c}_{g,t}$, delivers a benefit flow. When $\Upsilon_g > 0$, then search effort is associated with a cost in terms of goods. That is, the increase in the stock $\bar{c}_{g,t}$ is equal to consumption goods expenditures net of this search cost. Similarly, when $\Upsilon_s > 0$ then search effort is associated with a cost in terms of services. The model allows for both coefficients to be positive.

4.3 Household problem

The household problem is given by the following optimization problem:[80]

$$\max_{\left\{\begin{array}{c} c_t, \bar{c}_{g,t}, c_{g,t}, c_{s,t}, s_{i,g,t}, s_{i,s,t}, \\ b_t, n_t, i_t, i_{g,t}, i_{s,t}, k_t, e_t \end{array}\right\}_{t=0}^{\infty}} \quad \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left\{\xi_c \frac{c_t^{1-\gamma} - 1}{1-\gamma} - \xi_n n_t\right\}$$

subject to

[79] We assume that intellectual-property products are produced in the service sector.

[80] To economize on notation, we drop the h subscript which was used above to indicate that this problem is for an *individual* household. The important thing to remember is the following. In this household problem, e_t is the total amount of effort the household exerts and under control of the household. By contrast, the effort variables affecting $f_{i,b,t}$ in different markets are the average effort levels across households and are taken as given by an individual household. In equilibrium, these will all be the same, but that cannot be imposed when deriving first-order conditions. Also, we assume again that wages are sticky and set as in section 3 but leave that out of the discussion to highlight better what is new in this section.

$$\int_{i=0}^{1} P_{i,g,t} s_{i,g,t} di + \int_{i=0}^{1} P_{i,s,t} s_{i,s,t} di + b_t$$
$$\le W_t n_t + R_{k,t} k_{t-1} + d_t + (1 + R_{t-1}) b_{t-1}, \tag{36a}$$
$$c_t \le \min \left\{ \frac{\bar{c}_{g,t}}{\omega_{g,c}}, \frac{c_{s,t} - \Upsilon_s(\xi_e e_t - \overline{\xi}_e)}{\omega_{s,c}} \right\}, \tag{36b}$$
$$i_t \le \min \left\{ \frac{i_{g,t}}{\omega_{g,i}}, \frac{i_{s,t}}{\omega_{s,i}} \right\}, \tag{36c}$$
$$\bar{c}_{g,t} = (1 - \delta_c)\, \bar{c}_{g,t-1} + c_{g,t} - \Upsilon_g(\xi_e e_t - \overline{\xi}_e), \tag{36d}$$
$$c_{g,t} + i_{g,t} = \left(\int_{i=0}^{1} s_{i,g,t}^{\frac{\varepsilon_g - 1}{\varepsilon_g}} di \right)^{\frac{\varepsilon_g}{\varepsilon_g - 1}}, \tag{36e}$$
$$c_{s,t} + i_{s,t} = \left(\int_{i=0}^{1} s_{i,s,t}^{\frac{\varepsilon_s - 1}{\varepsilon_s}} di \right)^{\frac{\varepsilon_s}{\varepsilon_s - 1}}, \tag{36f}$$
$$k_t = (1 - \delta_k)\, k_{t-1} + i_t \left(1 - \frac{\eta_i}{2} \left(\frac{i_t}{i_{t-1}} - 1 \right)^2 \right), \tag{36g}$$
$$e_t = \int_{i=0}^{1} \left(\frac{s_{i,g,t}}{f_{i,g,t}^b} + \frac{s_{i,s,t}}{f_{i,s,t}^b} \right) di. \tag{36h}$$

Here, k_t denotes the end-of-period-t capital stock, $s_{i,g,t}$ purchases of type i goods, $s_{i,s,t}$ purchases of type i services, e_t total "effort" (which really is a loss in consumption goods and/or services), b_t end-of-period-t bond holdings, R_t the risk-free nominal interest rate on investing in bonds in period t, W_t the nominal wage rate, $R_{k,t}$ the nominal rental rate of capital, $P_{i,g,t}$ the price of goods of type i, $P_{i,s,t}$ the price of services of type i, d_t firm profits, $1/f_{i,g,t}^b$ the effort required to obtain one unit of good i, and $1/f_{i,s,t}^b$ the effort required to obtain one unit of type-i services.

The Leontief structure implies that optimal choices are such that

$$\bar{c}_{g,t} = \omega_{g,c} c_t, \tag{37a}$$
$$c_{g,t} = \omega_{g,c} \left(c_t - c_{t-1} (1 - \delta_c) \right) + \Upsilon_g(\xi_e e_t - \overline{\xi}_e), \tag{37b}$$
$$c_{s,t} = \omega_{s,c} c_t + \Upsilon_s(\xi_e e_t - \overline{\xi}_e), \tag{37c}$$
$$i_{g,t} = \omega_{g,i} i_t, \tag{37d}$$
$$i_{s,t} = \omega_{s,i} i_t. \tag{37e}$$

The ω coefficients satisfy $\omega_{g,c} = 1 - \omega_{s,c}$ and $\omega_{g,i} = 1 - \omega_{s,i}$. The Υ coefficients satisfy $\Upsilon_g = 1 - \Upsilon_s$.

Demand functions. From the household first-order conditions, we can derive the following demand functions:

$$s_{i,g,t} = \left(\left(\Upsilon_g + \Upsilon_s \frac{P_{s,t}}{P_{g,t}} \right) \frac{\xi_e}{f^b\left(\theta_{i,g,t}\right)} + \frac{P_{i,g,t}}{P_{g,t}} \right)^{-\varepsilon_g} s_{g,t} \tag{38}$$

$$s_{i,s,t} = \left(\left(\Upsilon_g \frac{P_{g,t}}{P_{s,t}} + \Upsilon_s \right) \frac{\xi_e}{f^b\left(\theta_{i,s,t}\right)} + \frac{P_{i,s,t}}{P_{s,t}} \right)^{-\varepsilon_s} s_{s,t} \tag{39}$$

Because of the Leontief structure, these demand functions are not that much more complicated than the one of the model without services. As before, the effort term in the demand function takes into account search efficiency, $f^b(\theta_{i,\cdot,t})$, and the cost of effort. What is new is that the latter can be in the form of goods or services or both. For example, if searching for goods requires some services, i.e., $\Upsilon_s > 0$, then the demand for goods also depends on the aggregate price of services. The functional forms of $f^b(\theta_{i,\cdot,t})$ and $f^f(\theta_{i,\cdot,t})$ are identical to the ones given in equation (19), but we allow for sector-specific scaling coefficients, μ_g and μ_s, as well as sector-specific curvature parameters, ν_g and ν_s.

Price Indices. The aggregate price indices for goods and services are given by[81]

$$P_{g,t} = \left(\int_0^1 \left(\frac{(\Upsilon_g P_{g,t} + \Upsilon_s P_{s,t})\xi_e}{f^b\left(\theta_{i,g,t}\right)} + P_{i,g,t} \right)^{1-\varepsilon_g} di \right)^{\frac{1}{1-\varepsilon_g}}, \text{ and} \tag{40}$$

$$P_{s,t} = \left(\int_0^1 \left(\frac{(\Upsilon_g P_{g,t} + \Upsilon_s P_{s,t})\xi_e}{f^b\left(\theta_{i,s,t}\right)} + P_{i,s,t} \right)^{1-\varepsilon_s} di \right)^{\frac{1}{1-\varepsilon_s}}. \tag{41}$$

The aggregate price for consumption goods is given by

$$P_t = \omega_{g,c} P_{g,t} + \omega_{s,c} P_{s,t}. \tag{42}$$

This price index will be used to define the inflation and the real interest rate.

4.4 Model block that determines customer-finding rates

There is a sub-set of equilibrium conditions that pins down key firm-level variables related to prices, tightness (and thus the customer-finding rate), and marginal costs. Specifically, it determines $\frac{P_{g,t}}{P_t}$, $\frac{P_{i,g,t}}{P_t}$, $\theta_{g,t}$, $\frac{P_{s,t}}{P_t}$, $\frac{P_{i,s,t}}{P_t}$, $\theta_{s,t}$, $MC_{g,t}$ and $MC_{s,t}$ given three variables that are related to expected future outcomes.[82] This subsystem resembles

[81] As pointed out in footnote 37, these are the values that would ensure zero profits if there was a producer that would combine the differentiated goods into an aggregate and they are also equal to the marginal cost of goods and services from the household's perspective.

[82] The presence of search costs implies that $P_{j,t}$, i.e., the aggregate price index for sector j, is not equal to but bigger than $P_{i,j,t}$, even in the symmetric equilibrium. In the equivalent setup with a

equation (32) which specifies the sub-system for the goods-only economy. The sub-system for the full model with services is given by[83]

$$\frac{P_{g,t}}{P_t} = \frac{\tilde{\psi}_t \xi_e}{f_g^b(\theta_{g,t})} + \frac{P_{i,g,t}}{P_t}, \tag{43a}$$

$$MC_{g,t} - \lambda^f_{x,t} = \left(\frac{P_{i,g,t}}{P_t} - \lambda^f_{x,t} - \frac{P_{g,t}}{P_{g,t}} \widetilde{\lambda}^f_{g,t} \right) f_g^f(\theta_{g,t}), \tag{43b}$$

$$MC_{g,t} - \lambda^f_{x,t} = \varepsilon_g \widetilde{\lambda}^f_{g,t} \xi_e \tilde{\psi}_t \frac{\nu_g}{1-\nu_g} \theta_{g,t}, \tag{43c}$$

$$\frac{P_{s,t}}{P_t} = \frac{\tilde{\psi}_t \xi_e}{f_s^b(\theta_{s,t})} + \frac{P_{i,s,t}}{P_t}, \tag{43d}$$

$$MC_{s,t} = \left(\frac{P_{i,s,t}}{P_t} - \frac{P_{s,t}}{P_{s,t}} \widetilde{\lambda}^f_{s,t} \right) f_s^f(\theta_{s,t}), \tag{43e}$$

$$MC_{s,t} = \varepsilon_s \widetilde{\lambda}^f_{s,t} \xi_e \tilde{\psi}_t \frac{\nu_s}{1-\nu_s} \theta_{s,t}, \tag{43f}$$

$$1 = \omega_{g,c} \frac{P_{g,t}}{P_t} + \omega_{s,c} \frac{P_{s,t}}{P_t}, \tag{43g}$$

$$MC_{g,t} = \frac{A_{s,t}}{A_{g,t}} MC_{s,t}, \tag{43h}$$

$$\widetilde{\psi}_t = \Upsilon_g \frac{P_{g,t}}{P_t} + \Upsilon_s \frac{P_{s,t}}{P_t}, \tag{43i}$$

where $\widetilde{\lambda}^f_{g,t}$, $\widetilde{\lambda}^f_{s,t}$, and $\lambda^f_{x,t}$ are given by

$$1 - \varepsilon_g \widetilde{\lambda}^f_{g,t} = \left(\eta_{P,g} \frac{P_t}{P_{i,g,t}} \right) \left(\begin{array}{c} \left(\frac{P_{i,g,t}}{P_{i,g,t-1}} - 1 \right) \frac{P_{i,g,t}}{P_{i,g,t-1}} \\ -\beta \mathbb{E}_t \left[\frac{\lambda_{t+1}}{\lambda_t} \left(\frac{P_{i,g,t+1}}{P_{i,g,t}} - 1 \right) \frac{P_{i,g,t+1}}{P_{i,g,t}} \frac{s_{g,t+1}}{s_{g,t}} \right] \end{array} \right), \tag{44a}$$

$$1 - \varepsilon_s \widetilde{\lambda}^f_{s,t} = \left(\eta_{P,s} \frac{P_t}{P_{i,s,t}} \right) \left(\begin{array}{c} \left(\frac{P_{i,s,t}}{P_{i,s,t-1}} - 1 \right) \frac{P_{i,s,t}}{P_{i,s,t-1}} \\ -\beta \mathbb{E}_t \left[\frac{\lambda_{t+1}}{\lambda_t} \left(\frac{P_{i,s,t+1}}{P_{i,s,t}} - 1 \right) \frac{P_{i,s,t+1}}{P_{i,s,t}} \frac{s_{s,t+1}}{s_{s,t}} \right] \end{array} \right), \tag{44b}$$

$$\lambda^f_{x,t} = \beta (1-\delta_x)(1-\eta_x) \mathbb{E}_t \left[\frac{\lambda_{t+1}}{\lambda_t} MC_{g,t+1} \right]. \tag{44c}$$

final-goods producer, $P_{j,t}$, would be the price of the final composite good that the consumer pays. To obtain good i, the final-goods producer has to pay $P_{i,j,t}$ *and* the search costs. In our setting, the consumer itself does the searching of the different goods, but the definition of the aggregate price indices incorporate search costs in the same way as with a final-goods producer.

[83]The expressions for marginal costs are identical to the one in equation (28a. These would be different across the two sectors when productivity in the goods sector, $A_{g,t}$, differs from productivity in the service sector, $A_{s,t}$.

We added an equation to introduce an auxiliary variable, $\widetilde{\psi}_t$, to make the system more understandable. But this is just a weighted function of ${}^{P_{g,t}}/_{P_t}$. and ${}^{P_{s,t}}/_{P_t}$, where the Υ_g and Υ_s coefficients indicate the relative importance of goods and services in search costs. Also, we make explicit that holding inventories implies a maintenance cost which is captured here with the parameter, η_x. This comes on top of depreciation.[84]

Although the Leontief structure helps in simplifying the equations, the system is larger than the one for the goods-only economy and has additional terms. The reader who is mainly interested in the intuition may skip to the next section in which we provide an intuitive discussion. But for those interested in the details, we will give a description of the system.

Equations (43a) and (43d) are rewritten versions of the demand equations for a type-i good and type-i service. It differs from the demand equation in equation (32a) in two ways. First, the demand for good i in sector j also depends on the relative aggregate demand for sector j which is captured by the relative aggregate price of sector j, $P_{j,t}/P_t$. Second, the cost of searching is no longer just equal to ξ_e, but depends on the relative importance of goods and services in obtaining purchases, measured by $\widetilde{\psi}_t$. Equations (43b) and (43e) are the equivalent of equation (32b). As explained above, it equates the marginal cost of producing an extra unit with the marginal benefits taking into account that (i) not all that could be sold is sold, i.e., $f_j^f(\theta_{j,t}) < 1$, (ii) unsold goods have value, i.e., $\lambda_{x,t}^f > 0$, and (iii) the demand function the firm faces acts as a constraint, i.e., $\widetilde{\lambda}_{j,t}^f > 0, j \in \{g, s\}$. These equations are basically the same as equation ((32b) except that the sector's relative price is included. The equivalent versions of Equation (32c), i.e., the firm's first-order condition for tightness, are equations (43c) and (43f). The latter two take into account that search costs depend on both goods and services and, thus, on their relative prices. Equation (43g) simply states that the weighted sum of the two relative sector prices have to add up to 1.[85] Equation (43i) defines the auxiliary variable, $\widetilde{\psi}_t$, which indicates that search costs depend on the relative price of goods and the relative price of services.

Having two sectors, we also have two Phillips Curves and they are given in equations (44a) and (44b).[86] Equation (44c) gives an expression for the value of bringing an inventory good into the next period and corresponds to equation (32d).

4.5 The interaction between the two sectors

In response to positive demand shocks, the customer-finding rate increases in both sectors which works through changes in $\widetilde{\lambda}_{g,t}^f$ and $\widetilde{\lambda}_{s,t}^f$, exactly as in the goods-only

[84]But recall from footnote 41, that the distinction between δ_x and η_x only matters for GDP accounting. For all other model properties only the value of $(1 - \delta_x)(1 - \eta_x)$ matters.

[85]If both sides of the equation are multiplied by P_t, then it simply says that the aggregate price index, P_t, is a weighted average of the price indices of the two sectors.

[86]The question arises whether it makes sense to adopt a Phillips Curve using aggregate inflation when – as is the case here – there are sector-sprecific Phillips Curves and sectoral relative prices behave differently.

model. Since this discussion is (again) quite intuitive, we postpone further discussion until section 4.7 in which we discuss quantitative model properties.

In section 3, we learned that the countercyclical behavior of $\lambda^f_{x,t}$ is the reason for a procyclical customer-finding rate in the goods sector in response to TFP shocks. Does this mean that the customer-finding rate for services is acyclical in the extended model, because services that are not sold have no value? That would be true in an economy with only services. But it is not necessarily true here, because there are interactions between the two sectors.

To study the interaction between the two sectors, we focus on the case where prices are fully flexible. Under flexible prices, we have that $\widetilde{\lambda}_{g,t} = \widetilde{\lambda}_{s,t} = 0$ which simplifies the sub-system of the previous section considerably.[87]

First case: $\Upsilon_g = \omega_{c,g}$ and $\Upsilon_s = \omega_{c,s}$. That is, we assume that the role of goods and services for search costs are the same as their utility contributions which equals their expenditure shares. The big advantage of this assumption is that $\widetilde{\psi}_t$ is a constant (and equal to 1). Consequently, there are only two interactions between the equations that determine the outcomes for the goods sector and the equations that do this for the service sector. First, a change in the relative price for goods necessarily implies a change in the relative price for services in the opposite direction, as indicated by equation (43g). Second, marginal costs satisfy the following relationship

$$MC_{g,t} = MC_{s,t}\frac{A_{s,t}}{A_{g,t}}. \tag{45}$$

The reason is that firms in both sectors minimize costs and face the same wage rate and rental cost of capital. In our benchmark calibration, the steady state levels of $A_{g,t}$ and $A_{s,t}$ are not equal, but this ratio of marginal cost levels would remain constant because ${}^{A_{g,t}}/_{A_{g,t-1}} = {}^{A_{s,t}}/_{A_{s,t-1}}$. For the first two cases considered here, the discussion would be exactly the same when marginal costs in the two sectors are always equal.

Note that TFP does not show up in the subsystem if ${}^{A_{g,t}}/_{A_{s,t}}$ is a constant. Moreover, $\widetilde{\lambda}^f_{g,t}$ and $\widetilde{\lambda}^f_{s,t}$ remain unaffected when we look at the case with flexible prices. This means that customer-finding rates, marginal-cost levels, and relative prices are only affected if the value of holding a good in inventory, $\lambda^f_{x,t}$, changes. If $\lambda^f_{x,t}$ remains unchanged, then an increase in TFP would keep marginal costs unchanged, because the reduction due to the increase in TFP is offset by the increase in output. And the customer-finding rates would remain unchanged because the increase in the supply of available goods would be accompanied by an increase in effort. Section 3 made clear, however, that the value of $\lambda^f_{x,t}$ falls in response to a TFP shock, because during goods times the marginal rate of substitution falls, which has a negative effect on the value of assets.

[87] The model with services also satisfies approximate divine coincidence unless additional frictions like wage stickiness are added. The reason is that the central bank goes against the deflationary pressure induced by a productivity increase with a monetary expansion. With fully-flexible prices, $\widetilde{\lambda}^f_{g,t}$ and $\widetilde{\lambda}^f_{s,t} = 0$ are always *exactly* equal to zero, which allows us to derive analytical results.

We will now discuss what the subsystem tells us about model outcomes when $\lambda^f_{x,t}$ falls and ${}^{A_{g,t}}/_{A_{s,t}}$, $\widetilde{\lambda}^f_{g,t}$ and $\widetilde{\lambda}^f_{s,t}$ remain constant. Using equations (43a), (43b), and (43c) we can solve for the relative price for good i, ${}^{P_{i,g,t}}/_{P_t}$, tightness in the goods sector, $\theta_{g,t}$, and marginal costs in the good sector, $MC_{g,t}$, as a function of the relative price of goods, ${}^{P_{g,t}}/_{P_t}$.[88] The same can be done as a function of ${}^{P_{s,t}}/_{P_t}$ for the service sector. An increase in ${}^{P_{g,t}}/_{P_t}$ means that the demand curve for goods has shifted out as goods become more attractive relative to services. In response, firms in the goods sector produce more. The latter increases marginal costs. Thus, the goods-sector marginal-costs curve is an upward sloping function of ${}^{P_{g,t}}/_{P_t}$. Moving upward along this curve is associated with an increase in the customer-finding rate, because the supply of available goods increases by less than demand. In exactly the same way, we can plot service-sector marginal costs as an upward sloping function of ${}^{P_{s,t}}/_{P_t}$ or as a downward sloping function of ${}^{P_{g,t}}/_{P_t}$, since ${}^{P_{s,t}}/_{P_t} = (1 - \omega_{g,c}{}^{P_{g,t}}/_{P_t})/\omega_{c,s}$.

These two functions are plotted in figure 8, where we have scaled the marginal cost function for the service sector with ${}^{A_{g,t}}/_{A_{s,t}}$ consistent with equation (45). The solution of the subsystem is given by the intersection of the two curves at which point equation (45) is satisfied.

Now suppose that there is an increase in TFP. As discussed above, production and effort would scale up together with TFP and nothing would change in the subsystem *if* $\lambda^f_{x,t}$ would remain the same.[89] But we learned in section 3 that $\lambda^f_{x,t}$ falls when productivity increases. The reduction in the value of unsold goods will dampen the increase in goods-sector production and increases tightness and the customer-finding rate. In the figure, this is represented by the downward shift of the $MC_{g,t}$ curve. The marginal-costs curve for the service sector is unchanged, since there cannot be a change in the zero value of unsold services. At the old level of $P_{g,t}/P_t$, marginal costs in the goods sector are lower than those in the service sector adjusted for the (constant) value of ${}^{A_{g,t}}/_{A_{s,t}}$. Consequently, $P_{g,t}/P_t$ has to increase. That is, the dampened response of production in the goods-sector will lead to an increase in its relative price which necessarily means a decrease in the relative price of services. Thus, the downward shift of the goods-sector $\mathrm{MC}_{g,t}$ curve is followed by a movement along the new curve raising $P_{g,t}/P_t$. This implies an outward shift in the demand curve for firms producing goods and a *further* increase in tightness in the goods sector. For the service sector, there is a movement along the old unchanged marginal-costs curve and we know that the lower relative price of services implies an inward shift of firms' demand curves and a lower customer-finding rate. This would indicate that the customer-finding rates in the two sectors would move in opposite directions in response to TFP disturbances: procyclical in the goods sector and countercyclical in the service sector.

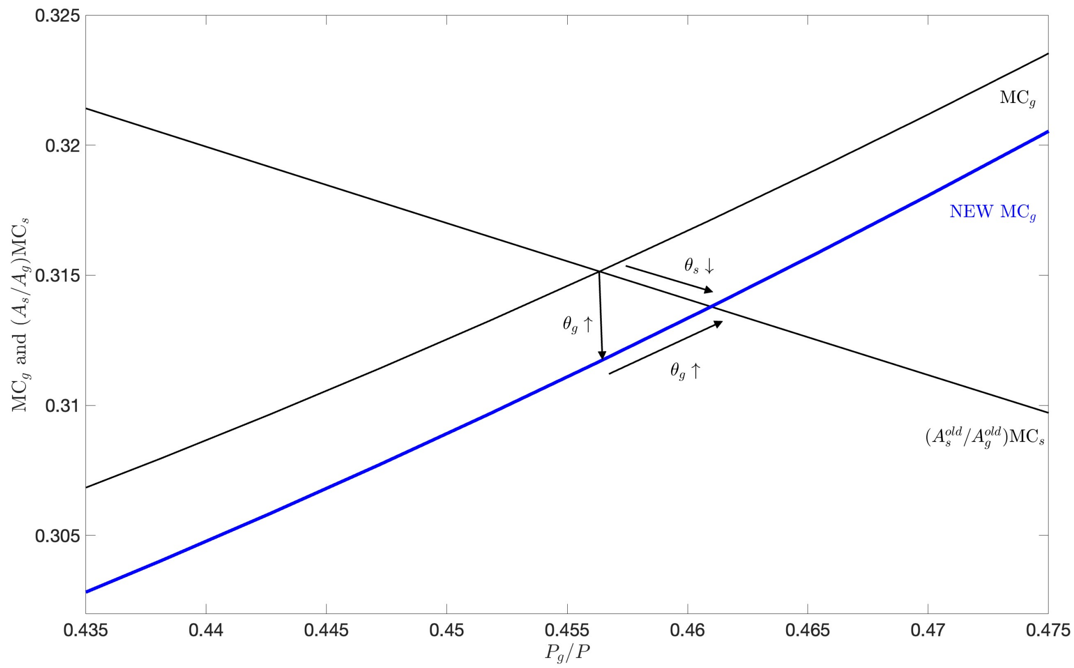
The figure does not reveal how large the reductions in the customer-finding rate are for the service sector. In section 4.7, we show that they are small for the case considered here where ${}^{A_{g,t}}/_{A_{s,t}}$ remains constant. Moreover, the prediction of a countercyclical

[88]Recall that ${}^{P_{i,g,t}}/_{P_t}$ differs from ${}^{P_{g,t}}/_{P_t}$ in that it takes into account search costs.

[89]Recall that $\widetilde{\lambda}^f_{g,t}$ and $\widetilde{\lambda}^f_{s,t}$ are not affected because $\eta_{P,g} = \eta_{P,s} = 0$.

response of the customer-finding rate in the service sector is not robust. One reason is that the full model does not have flexible prices. More interesting is the analysis of the third case which shows that a procyclical response of the customer-finding rate in the service sector is possible if productivity in the service sector does not response one for one with productivity in the goods sector.

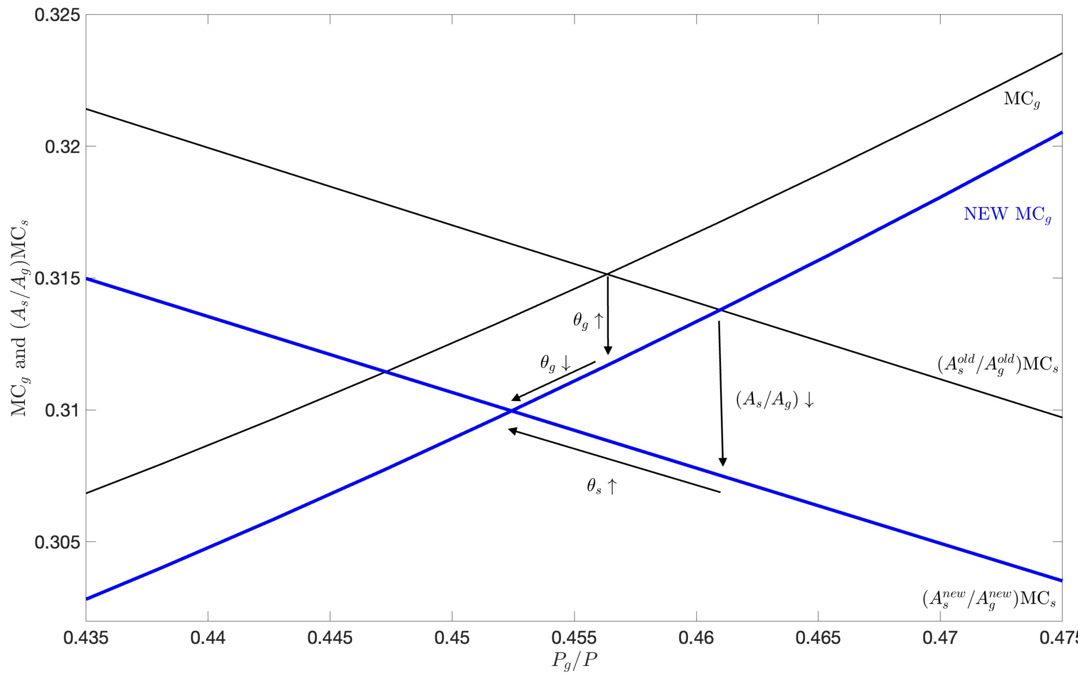
Figure 8: impact of an increase in TFP and associated fall in $\lambda^f_{x,t}$ on both sectors; $\Delta\left(\frac{A_{g,t}}{A_{s,t}}\right)=0$



Notes. As made clear in section 3, an increase in TFP leads to a decrease in the value of an unsold good, $\lambda^f_{x,t}$. Sub-system (43) implies that the drop in $\lambda^f_{x,t}$ implies an increase in tightness, $\theta_{g,t}$, and a downward shift of the marginal cost curve of the goods sector. This figure corresponds to the benchmark case when $A_{g,t}/A_{s,t}$ remains constant following an aggregate TFP shock.

Second case: $\Upsilon_g = 0$ **and** $\Upsilon_s = 1$**.** This means that search requires the use of only services and not goods. It is still the case that the sectoral marginal-cost level implied by the sub-system is increasing with the relative price of the sector. That is, the qualitative features of figure 8 remain unchanged. However, there is *no* change in the customer-finding rate for the service sector, $\theta_{s,t}$ when – after the drop in $\lambda^f_{x,t}$ – there is a movement *along* the marginal-costs curve for the service sector as $P_{g,t}/P_t$ increases. The reason is that the reduction in $P_{s,t}/P_t$ now lowers $\widetilde{\psi}_t$, i.e., search costs relative to the aggregate price index. This boost in demand goes against the reduction in demand because of the reduction in $P_{s,t}/P_t$. Consequently, the solution to the firm problem is to let $P_{i,s,t}/P_t$ decline at the same rate as $P_{s,t}/P_t$ which means that $\widetilde{\psi}_t$ would also drop by the same percentage, which would leave tightness and the customer-finding rate in the service-sector unchanged. This result indicates that countercyclicality can turn into acylicality, but only at this *corner* choice for Υ_g and Υ_s. However, we will now discuss a reason why the customer-finding rate in the service sector could very well be *pro*cyclical.

Figure 9: impact of an increase in TFP and associated fall in $\lambda^f_{x,t}$ on both sectors; $\Delta\left(\frac{A_{g,t}}{A_{s,t}}\right) > 0$



Notes. As made clear in section 3, an increase in TFP leads to a decrease in the value of an unsold good, $\lambda^f_{x,t}$. Sub-system (43) implies that the drop in $\lambda^f_{x,t}$ implies an increase in tightness, $\theta_{g,t}$, and a downward shift of the marginal-costs curve of the goods sector. The figure considers the case when productivity in the goods sector is affected more heavily by the aggregate TFP shock than productivity in the service sector.

Third case: $\Delta\left(\frac{A_g}{A_s}\right) > 0$ **.** The finding that the two customer-finding rates move in opposite directions following a TFP shock can be easily overturned when TFP in the goods sector is more responsive than TFP in the service sector. The associated decrease in $\lambda^f_{x,t}$ leads again to a drop in the marginal-costs curve for the goods sector. Since there is now also a change in the relative productivity levels, there must be an additional shift as indicated by equation (45). And since we plot MC_g and $MC_s{}^{A_s}/_{A_g}$ this means a downward shift of $MC_s{}^{A_s}/_{A_g}$. If the drop in ${}^{A_s}/_{A_g}$ is big enough, then $P_{g,t}/P_t$ actually drops instead of increases, as indicated in figure 9. This means that $P_{s,t}/P_t$ increases and the movement along the $MC_s{}^{A_s}/_{A_g}$ curve now implies that tightness and the customer-finding rate in the service sector increase. The movement along the MC_g curve implies a reduction in tightness in the goods-sector, but we find that this is dominated by the direct effect. Consequently, customer-finding rates increase in both sectors. A quantitative illustration is given in appendix E.1.

4.6 Parameter calibration and estimation

Parameter values of our benchmark calibration are given in table 4. The top panel contains the parameters for which we use values that are common in the literature.[90] The second panel contains parameter values that are pinned down by empirical observations using standard calibration arguments.

The second column of the table list the relevant empirical observation. The third column indicates whether this empirical observation is the *only* piece of information used to pin down the parameter value (indicated with "O") or whether it is the *main* piece of information, but is pinned down in a system of equations (indicated with "M"). The Leontieff weights, $\omega_{g,c}$, $\omega_{g,s}$, $\omega_{g,i}$, and $\omega_{g,i}$, are pinned down by observed average ratios of goods relative to service purchases, where we use investment in intangibles as our measure of investment in services. In our benchmark, we assume that the acquisition cost parameters, Υ_g and Υ_s, are equal to the associated Leontieff weights.[91] Cao et al. (2022) estimate the annual depreciation rate for durables to be equal to 16%. We have only one type of consumption good, so the calibration of our depreciation rate, δ_c, takes into account that non-durable goods fully depreciate. Using the observed ratio of durable versus non-durable consumption, which is equal to 0.68, we obtain a quarterly depreciation rate of 0.6936 for our composite consumption good.

Gross investment in inventories is equal to 0.40% of GDP which pins down the depreciation rate for inventories, δ_x. We base our estimate for inventory maintenance costs, η_x, on Richardson (1995), but we do not include the costs related to the cost of money as this is part of our discount factor and we also exclude depreciation since we want average gross investment in inventories in our model to be consistent with national accounting data.[92] Remaining costs are clerical and inventory control, physical handling, warehouse expenses, insurance, and taxes. On an annual basis and as a fraction of the value of the inventory stock, the estimated ranges are 3-6%, 2-5%, 2-5%, 1-3%, and 2-6%, respectively. We use the upper estimate, i.e., 6.9395% on a quarterly basis. This is a conservative approach. The main mechanism responsible for the model to generate a procycal customer-finding rate in response to a TFP shock is the countercyclical fluctuation in the value of unsold goods. Quantitatively, this

[90] As in Gali (2015), we set the elasticities of substitution, ε_g and ε_s, equal to 6. Following Erceg et al. (2000), the elasticity of substitution among labor units, ε_w, is set to be the same as ε_g and ε_s. This value for ε_w is also consistent with those used in the literature, which typically range from 4 to 21; see Huo and Ríos-Rull (2020). The intertemporal substitution elasticity is set equal to 1. This is a common value in the literature and imposes balanced growth. Although not necessary to calculate IRFs or generate short simulations, balanced growth is necessary for our estimation procedure because it allows for a stationary-inducing transformation of the variables. The Taylor-rule coefficient related to inflation, Γ_π, and persistence, Γ_{lag}, are also standard. As shown in section 3, the value of Γ_y is important and we either estimate or calibrate it as discussed below. In appendix E.2, we consider results when the three coefficients are based on estimates from Mazelis et al. (2023).

[91] In appendix E.3, we confirm robustness of our results to alternative assumptions.

[92] As discussed in footnote 41, it is the sum of maintenance costs, η_x, and inventory depreciation, δ_x, that matters for model properties. The only exception is the calculation of GDP; whereas maintenance costs reduce GDP, depreciation does not.

Table 4: Benchmark calibration and estimation

commonly used values	target	M/O
discount factor: $\beta = 0.99$	-	-
intertemporal substitution elasticity: $\gamma = 1$	balanced growth	-
demand elasticity: $\varepsilon_g = \varepsilon_s = 6$	-	-
labor substitution elasticity: $\varepsilon_n = 6$	-	-
Taylor rule inflation response: $\Gamma_\pi = 1.5$	-	-
Taylor rule lag : $\Gamma_{\text{lag}} = 0.5$	-	-
based on data	target	M/O
Leontieff weight $i_s : \omega_{s,i} = 0.2415$	$\overline{i_{\text{intangibles}}/i}$	O
Leontieff weight $i_g : \omega_{g,i} = 1 - \omega_{s,i}$	-	O
Leontieff weight $c_g : \omega_{g,c} = 0.4229$	$\overline{c_g/c}$ and δ_c	O
Leontieff weight $c_s : \omega_{g,s} = 1 - \omega_{g,c}$	-	O
weight goods in search cost: $\Upsilon_g = \omega_{g,c}$	symmetry acquisition cost & expenditures	O
weight services in search costt: $\Upsilon_s = \omega_{g,s}$	symmetry acquisition cost & expenditures	O
depreciation goods: $\delta_c = 0.6936$	CCDHK22 and $\overline{c_{\text{durables}}/c}$	O
inventory depreciation: $\delta_x = 0.0040$	$\overline{\Delta x/y}$	O
inventory maintenance: $\eta_x = 0.0694$	R95	O
investment adjustment cost: $\eta_i = 0.1$	uniformly positive investment response	M
curvature production function: $\alpha = 0.7286$	$\overline{c/i}$	M
correlation TFP growth: $\rho_A = 0.35$	$\overline{\rho}(\Delta \ln A_t, \Delta \ln A_{t-1})$	O
price adjustment costs: $\eta_P = 0.10$	typical real response monetary shock	M
wage adjustment costs: $\eta_W = 0.10$	typical real response monetary shock	M
relative productivity: $A_g/A_s = 1.855$	$\overline{n_g/n_s}$	M
scaling goods search friction: $\mu_g = 0.5060$	$\overline{f}_g^f$	M
scaling services search friction: $\mu_g = 0.2295$	$\overline{f}_s^f$	M
based on normalization	normalization	M/O
TFP levels: $A_g = 0.8983$	$y_{ss} = 1$	M
scaling utility: $\xi_c = 0.8148$	$\lambda_{ss} = 1$	M
disutility working: $\xi_n = 0.3479$	$n_{ss} = 1$	M
disutility effort: $\xi_e = 0.0134$	$\theta_{g,ss} = 1$	M
estimation or matching key inventory, production, and sales moments		
$\partial R/\partial\text{output gap}$: Γ_y	0.0120 or in calibrated admissible range	
curvature search goods: ν_g	0.3469 or in calibrated admissible range	
curvature search services: ν_s	0.6713 or in calibrated admissible range	
standard deviation $\varepsilon_{A,t}$: σ_A	0.0038 or in calibrated admissible range	
standard deviation $\varepsilon_{R,t}$: σ_R	0.0036 or in calibrated admissible range	

Notes. R95 refers to Richardson (1995) and CCDHK22 refers to Cao et al. (2022). An upper bar indicates the corredponding estimated sample moment is used in the calibration. An O in the third column indicates that the parameter is pinned down using *only* the target mentioned in the second column. An M indicates that the calibration principle given in the second column is the *main* one to pin down this parameter, but its value is solved from a system of equations.

channel will only be relevant if its value is nontrivial relative to the value of newly produced goods and the higher the maintenance costs the lower the value. By showing that our model can generate a procyclical customer-finding rate *even* when we use a relatively high number for maintenance costs, we demonstrate the robustness of our mechanism. Our approach is also conservative, since these estimates are from the 1990s and inventory control costs are likely to have gone down since then.[93]

The curvature of the production function, α, is pinned down by the observed average for the ratio of consumption over investment. The relative magnitude of productivity in the two sectors, A_g/A_s, is chosen to match the observed relative employment shares in these two sectors.

The AR(1) coefficient in the law of motion for productivity growth is pinned down by the estimated auto-correlation using TFP data that are corrected for capacity utilization as described in Fernald (2014).[94]

The scaling coefficients of the search frictions, μ_g and μ_s, are chosen such that the model's steady-state values for the customer-finding rates are equal to the estimated average of their empirical counterpart.[95]

The price and wage adjustment cost parameters, η_P and η_W, are chosen such that a

[93] One could argue that we should not include taxes, since we abstract from taxes in our model. So an alternative calibration strategy would be to exclude taxes, but include obsolescence as well as deterioration and pilferage from Richardson (1995) to determine the value for η_x and set $\delta_x = 0$. If we would use the midpoint estimates, then we get a value for η_x equal to 7.44%. This is virtually identical to the combined effect of using our benchmark parameter values, i.e., $\eta_x + \delta_x + \eta_x * \delta_x$. As pointed out in footnote 92, it is this combined value that matters for virtually all model properties.

[94] Specifically, the estimated auto-correlation of adjusted annual TFP growth is equal to 0.112. Annual TFP growth displays a minor trend. In our model, TFP is non-stationary but TFP *growth* is stationary. When we correct for this (minor) trend in the data using an HP filter with a very high smoothing coefficient, namely $({}^{10000000}/_{4})^4$, then this correlation drops to 0.095. This correction only takes out extremely low frequency movements in the data, since we are interested in a suitable specification for *raw*, i.e., unfiltered, TFP growth that resembles the observed series except that it is stationary. Using $\rho_A = 0.35$ for our quarterly process, we find an auto-correlation for TFP growth (expressed on an annual basis) equal to 0.121 when a long series of $T = 100,000$ observations is used and a value equal to 0.096 for the mean across $N = 10,000$ short samples with length equal to the one of our data set. The latter matches the empirical counterpart and is the more relevant measure because the calculation of the correlation coefficient allows for the adjustment of the mean in the short sample just as is done in the data. For many of the moments we report the two approaches lead to similar answers, but this is an example where there is a more substantial difference. Note that these two ways to calculate higher-order moments could deviate even for large T and N.

[95] For the goods sector, the average customer-finding rate is equal to 0.501 and – as indicated by equation (19b) – is a simple transformation of the inventory-sales ratio. For the customer-finding rate for services, we use the Euro-Area capacity-utilization survey for services which gives an average of 0.89. It seems plausible that the customer-finding rate is substantially higher for services. After all, a good that is not sold ends up in inventories and could be still be sold at some future date, whereas a service provider that is idle has nothing to bring into the next period. As pointed out in section 2, this data series is very short and clearly not as ideal as what we have for the goods sector. Appendix E.4 documents that the target average customer-finding rate turns out to be not that important.

monetary-policy shock leads to a plausible outcome for the aggregate real economy.[96,97] Similarly, the value for the investment adjustment cost parameter, η_i, is such that the volatility of investment relative to GDP and consumption are empirically plausible.[98]

The next block of table 4 summarizes the calibration of parameters whose values are pinned down by normalizations. The parameters ξ_c, ξ_n, and ξ_e are scaling coefficients of the utility function, the dis-utility of working, and the dis-utility of effort, respectively. Those values are chosen to obtain unit steady-state values for the Lagrange multiplier of the household budget constraint, hours worked, and tightness in the goods sector. Model properties would not be affected if other targets are chosen. The final normalization is that aggregate output is equal to 1, which pins down the value of A_g.[99]

This leaves the standard deviations of the two shocks, σ_R and σ_A, and three parameters that are key for the qualitative and quantitative outcomes of inventory moments, Γ_y, ν_g, and ν_s. The parameter Γ_y captures the responsiveness of monetary policy to the output gap. As discussed in section 3, our model satisfies approximate divine coincidence for some basic versions of the model. The output gap and inflation response are then approximately zero and the value of this parameter does not matter. When standard features such as investment adjustment costs and wage stickiness are introduced, however, then this is no longer the case and the value of Γ_y does matter. A positive value for Γ_y implies a tightening when the output gap is positive, which – like a negative monetary-policy shock – would imply downward pressure on the customer-finding rate. If this effect is strong enough, then the customer-finding rate could become countercyclical in response to TFP shocks. The final two parameters are the curvature of the goods-market friction in the two markets, ν_g and ν_s. The value of ν_g directly af-

[96] A twenty-five basis points drop in the annual nominal policy rate leads to a drop in the production of goods of 0.72% and a drop in GDP of 0.37%. This is for the estimated values of ν_g, ν_s, and Γ_y, but very similar numbers are obtained when values for these three parameters are considered that are in the calibrated admissible range. Estimating the empirical impact of monetary-policy shocks is nontrivial and hampered by several challenges such as the difficulty to identify monetary-policy shocks and dealing with potentially time-varying and state-dependent outcomes. Figure 3 in Miranda-Agrippino and Ricco (2017) reports estimated peak responses for industrial production between roughly 2 and 3 percent for a 1 percentage point change in the policy rate, which means a range between 0.5 and 0.75 percent for a 25 basis point drop. Standard-error bands, however, are quite large. Thus, our theoretical responses are consistent with the data, but at the upper end of the empirical point estimates. In section 4.7.2, we use inventory data to evaluate the relative importance of monetary policy and TFP shocks for the volatility of output. It turns out to be the case that these inventory data present some challenges for monetary-policy shocks. It is, therefore, important that we give monetary-policy shocks the best possible chance and not understate their quantitative importance.

[97] As discussed in section 3, reducing the amount of price stickiness does not affect the responses to TFP shocks when the model satisfies divine coincidence. In terms of the monetary-policy-shock IRFs, reducing wage stickiness would scale down the IRFs and not affect our conclusions regarding the correlation properties of inventory, production, and sales data that we focus on.

[98] As discussed in section 3, setting $\eta_i > 0$ ensures that the initial investment response to a TFP shock is not negative.

[99] Recall that A_g/A_s is pinned down by relative sectoral employment shares. So knowing A_g and this ratio will imply a value for A_s.

fects the behavior of key goods-market variables related to inventories, production, and sales. Similarly, the value of ν_s directly affects the customer-finding rate of the service sector for which there is, unfortunately, little information. We adopt two strategies to determine the values of these three parameters.

Full-information Bayesian estimation strategy. The first strategy consists of estimating Γ_y, ν_g, ν_s, σ_R, and σ_A using a full-information Bayesian strategy. Specifically, we use the growth rates of inventories and sales to estimate these five parameters.[100] Details are given in appendix D.4. The data are remarkably powerful in identifying the parameters including the curvature parameter of search costs in the service sector, ν_s, even though no data for the service sector are used in the estimation. The posterior modes for these five parameter values are given in the bottom block of table 4.

Calibration instead of estimation. Full-information estimation has some disadvantages. One disadvantage is that the procedure is a bit of a black box. This is important for our paper because the objective is to see whether the model is consistent with a precise set of popular stylized facts that are highlighted in the inventory literature and it is not clear whether the full-information parameter estimates give the model the best possible chance to do this.[101]

Another disadvantage of full-information estimation methods is that they use all aspects of both the model and the data, which means that all shocks present in the model are considered simultaneously. The same would be true when a formal Simulated-Method-of-Moments method would be adopted. But we would like to investigate whether our model is consistent with key inventory, production, and sales facts for both monetary policy *and* TFP shocks and whether that is the true at the same set of structural parameters. This is important because there is a lot of empirical uncertainty regarding the relative importance of different types of shocks and we would like our results to not depend on a particular mix of demand and supply shocks, which in our model would require the right relative values of the two innovation standard deviations.

Finally, we would like to mention that full-information methods are quite ambitious, since they require that the model is correctly specified which – of course – is not the

[100]For the universe of firms for which we have inventory and sales data, we do not have production data, but the inventory accumulation identity implies production levels when given a value for the depreciation of inventories, δ_x. We prefer to use the calibrated value for δ_x that ensures that average investment in inventories is equal to its observed empirical counterpart. When we do add δ_x to the list of parameters to be estimated, however, then it has little effect on the outcomes for the other five parameters. Moreover, the estimated value of the posterior mode of δ_x is equal to 0.022 which is higher than the calibrated value which is equal to 0.0044, but also implies that inventories depreciate slowly.

[101]A relevant observation is the following. The full-information parameter estimates imply a value for the ratio of the standard deviation of the growth rate of inventories to the standard deviation of the growth rate of sales that is less than its empirical counterpart. Specifically, at the posterior modes the model-implied value is equal to only 68% of the corresponding value in the data used to estimate the parameters. This does turn out to be important as discussed in section 4.7.

case.[102]

The key elements of our calibration strategy to pin down the values of Γ_y, ν_g and ν_s are the following.[103] The key principle is that parameter values are determined using key properties emphasized in the inventory literature. The first one is that output is more volatile than sales, but not by more than what we see in the data taking into account sampling uncertainty.[104] The second is that the customer-finding rate (inventory-sales ratio) is procyclical (countercyclical), again taking into account sampling uncertainty of the empirical estimate. Moreover, we want that the model is consistent with these key properties for both a monetary policy and a TFP shock. This will create a tension for the appropriate choice for ν_g for the following reason. Section 3 showed that the customer-finding rate is procyclical in response to both types of shocks, but its response (relative to the same change in production) is bigger following a demand shock. Thus, getting the appropriate quantitative response would possibly require a different value for ν_g.

Because this calibration strategy is computationally involved, we limit ourselves to the case where $\nu_g = \nu_s$. The possible values that we obtain for ν_g and Γ_y are shown in figure 10. Our approach leaves us the flexibility to consider different combinations for the standard deviations of the two innovations, since model predictions are consistent with key inventory facts for both types of shocks. The choice for these standard deviations will be discussed in detail in the next section.

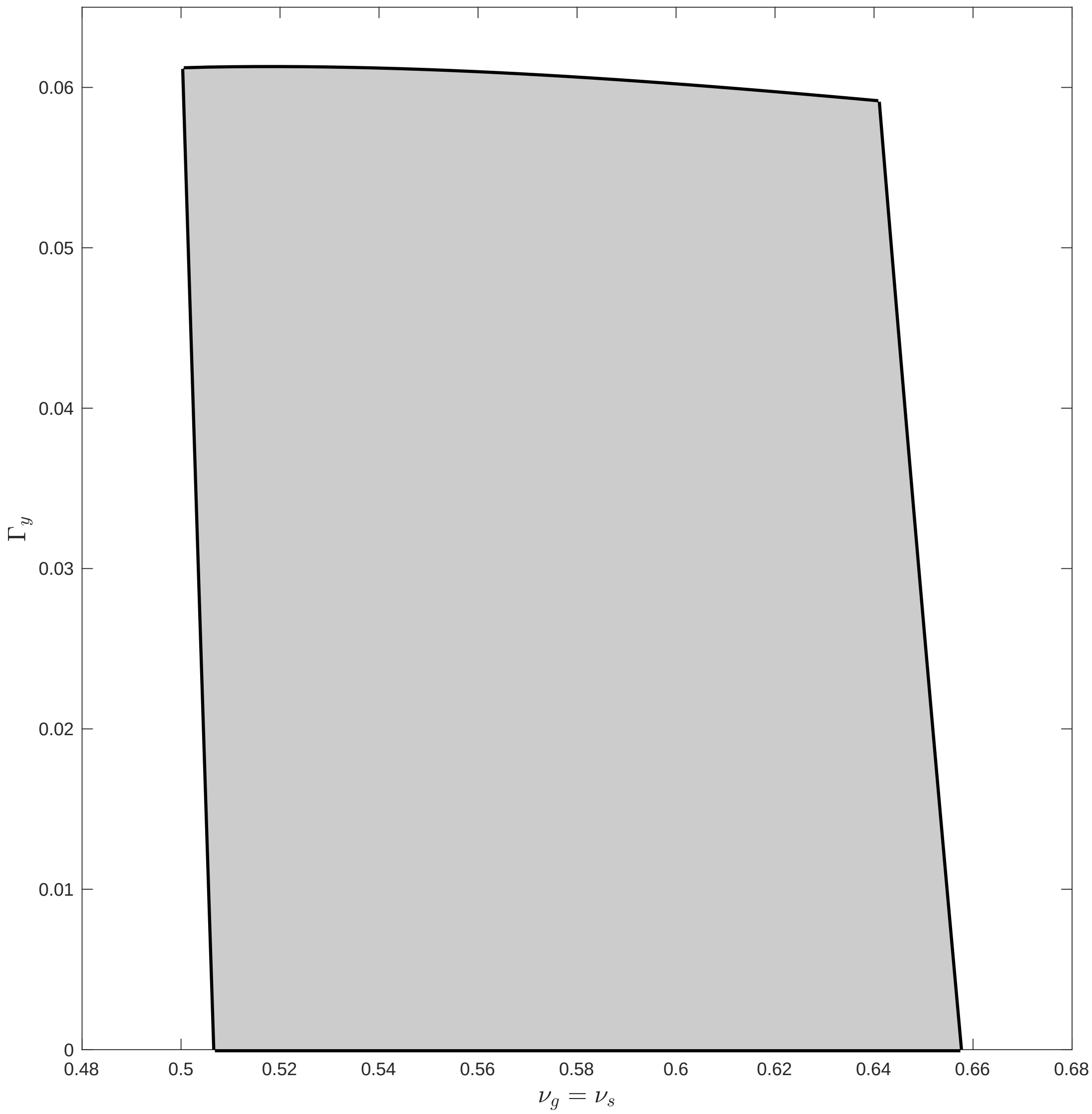
The value for ν_g obtained with the full-information estimation strategy is equal to 0.3469 which is quite a bit lower than the lower bound of the range obtained with the calibration strategy which is equal to 0.5006. The value for ν_s obtained with the full-information estimation is equal to 0.6713 which is above the upper bound. Nevertheless, model-implied properties are actually not that different except that the model based on parameters from the calibration strategy performs a bit better as discussed in the next section.

[102]Den Haan and Drechsel (2021) show that even a very minor misspecification of the empirical model can lead to large biases in parameter estimates, which in turn are associated with biased predictions of the theoretical model evaluated using estimated parameters.

[103]Details are given in appendix D.1.

[104]The lower (upper) bound of this ratio determines a lower (upper) bound for the curvature parameter ν_g.

Figure 10: admissible area for ν_g (=ν_s) and Γ_y



Notes. This figure plots the parameter values for which the model can match key empirical inventory, production, and sales facts for both monetary policy and TFP shocks.

4.7 Predictions of the full model with goods and services

This section discusses key properties of the full model. First, we present IRFs and model moments. Next, we show how inventory data have identifying information to determine the relative importance of demand and supply shocks for the cyclical variations in GDP. In the last subsection, we document that changes in ν_g, the parameter that controls the *quantitative* importance of cyclical variations in goods-market frictions, have a nontrivial effect on aggregate economic activity, *even* if we restrict ourselves to the relatively small variations that are possible within the calibrated admissible area, which ensures that model properties remain consistent with observed inventory facts.

4.7.1 Impulse response functions and moments

Figures 11 and 12 plot IRFs in response to a monetary policy and a TFP shock, respectively.[105] To generate these IRFs, we consider values for ν_g, ν_s, and Γ_y that are either the modes of the posterior from the full-information estimation procedure or a combination in the middle of the calibrated admissible area. The size of the initial shock is set to generate a peak 1 percent increase in GDP for both shocks and for both parameter choices.

The two sets of IRFs are based on parameter values that are quite different.[106] Nevertheless, the IRFs display a very similar shape and it is also true that most of the magnitudes are similar. The IRFs indicate that the customer-finding rate (inventory-sales ratio) is procyclical (countercyclical) and that the output response exceeds the sales response. And this is true for both types of shocks.[107] The response of the inventory stock following a monetary-policy shock does differ across the two parameter sets. It quickly turns negative when estimated parameter values are used whereas the response remains uniformly positive for the calibrated parameter values. That is, inventories are more procyclical for the calibrated parameter values. The reason is that the estimated value for ν_g is lower than the calibrated one, which implies a more responsive customer-finding rate, which in turn implies a stronger sales response. The calibration procedure is designed to get responses that are consistent with key inventory facts for *both* shocks. This is important to us because we want to illustrate that our model can be consistent with key inventory facts for both types of shocks. A model can, of course, be consistent with *unconditional* observed moments even if that is not the case for a particular type of shock. As long as the model is consistent with a particular statistic for one of the shocks and that shock is quantitatively important enough, then the model can be consistent with the empirical unconditional value. Nevertheless, we want to express doubt regarding the smaller – and at some point negative – inventory response associated with the estimated parameter values. At these parameter values, the model also under-predicts the volatility of the growth rate of inventories *even though* this growth rate is one of the two data series used to estimate the parameters.[108]

Investment in inventories is also procyclical. This is obvious for TFP shocks, since the initial increase in investment is followed by further gradual increases. The reason it is procyclical for monetary-policy shocks is that the initial increase dominates

[105] Model properties are based on a first-order perturbation numerical approximation.

[106] Specifically, $\nu_g = 0.3469$, $\nu_s = 0.6713$, and $\Gamma_y = 0.012$ for the IRFs based on the estimation procedure and $\nu_g = \nu_s = 0.565$ and $\Gamma_y = 0.03$ for the case based on the calibration.

[107] However, this is only true initially for the monetary-shock IRFs based on the estimated parameters.

[108] Specifically, the standard deviation of the growth rate of the inventory stock used in the estimation is equal to 0.0097, whereas the value implied by the model is equal to only 0.0060 (at the posterior mode values). Of course, it is not the objective of full-information estimation methods to match key moments of data used in the estimation. Nevertheless, it is not a desirable outcome.

the subsequent gradual decreases and this is true for both parameter sets.[109]

The IRFs also document that the model generates the usual relative volatility for GDP, consumption, and investment for both types of shocks. Moreover, output and sales of goods are more volatile than their counterparts of the service sector. The reason is that goods form a larger fraction of investment than services and investment is the more volatile expenditure component. Also, consumption goods are partially durable and the Leontief structure links consumption services to the *stock* of consumption goods not consumption goods expenditures.

The reasons the model can in principle match key observed inventory, production, and sales facts are really the same as the ones given in section 3 for the economy with no service sector and only a goods sector. Thus, it is more interesting to focus on the theoretical predictions of the service sector which are especially useful given the limited empirical data on goods-market frictions in the service sector. When considering the results, it is important to take into consideration that both the estimation and the calibration only relied on available data from the goods sector.

Following a monetary-policy shock, the increase in demand increases the customer-finding rate in the service sector for the same reasons as it does in the goods sector. On impact the response is slightly smaller in the goods sector. One dampening factor for the goods sector is the increase in the value of unsold goods making it more attractive for firms to set higher prices and dampen the increase in sales. However, the customer-finding rate increase is more persistent in the goods sector. This mirrors the persistence of investment which affects the goods sector more since the goods sector is more important for investment than the service sector.[110]

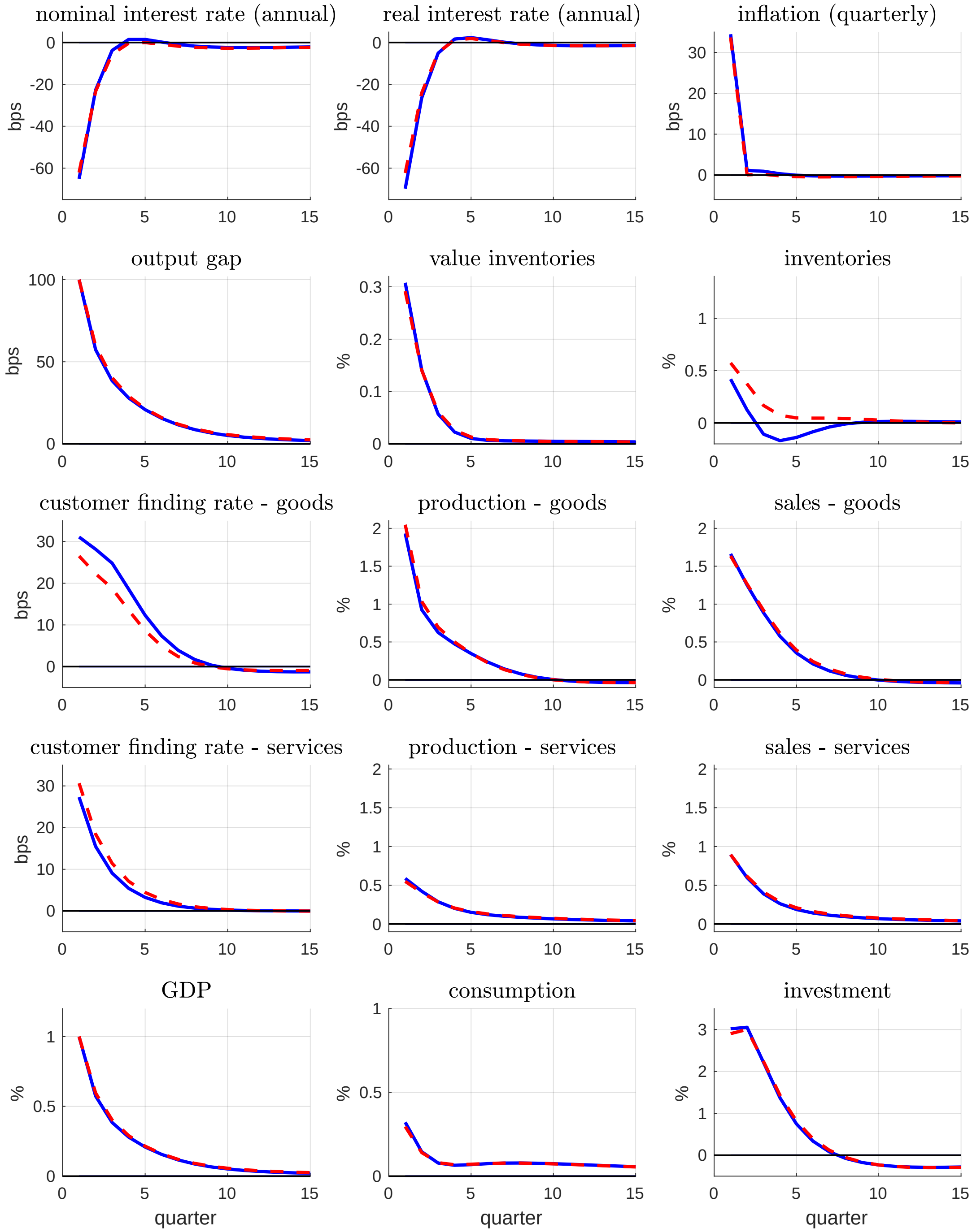
As discussed in section 4.5, the sign of the customer-finding rate in the service sector could be positive or negative following a TFP shock. Indeed, we find that the sign varies even within our relatively narrow range of admissible parameters. However, a better way to characterize the results is that the response of the customer-finding rate in the service sector following a TFP shock is always very small. In appendix E.1, we show that the customer-finding rate in the service sector displays a more robust procyclical response if productivity in the service sector lags the increase in the goods sector, consistent with the theoretical analysis of section 4.5.[111]

[109]Specifically, the correlation coefficient between the cyclical components of GDP and the growth rate of the inventory stock in a model with only monetary-policy shocks is equal to 0.53 when the calibrated parameter combination is used. Consistent with the IRFs, it is smaller and only equal to 0.29 when the estimated posterior modes are used.

[110]Recall that we calibrate the role of the service sector for investment using data on intangibles which is only 24.15% of total investment.

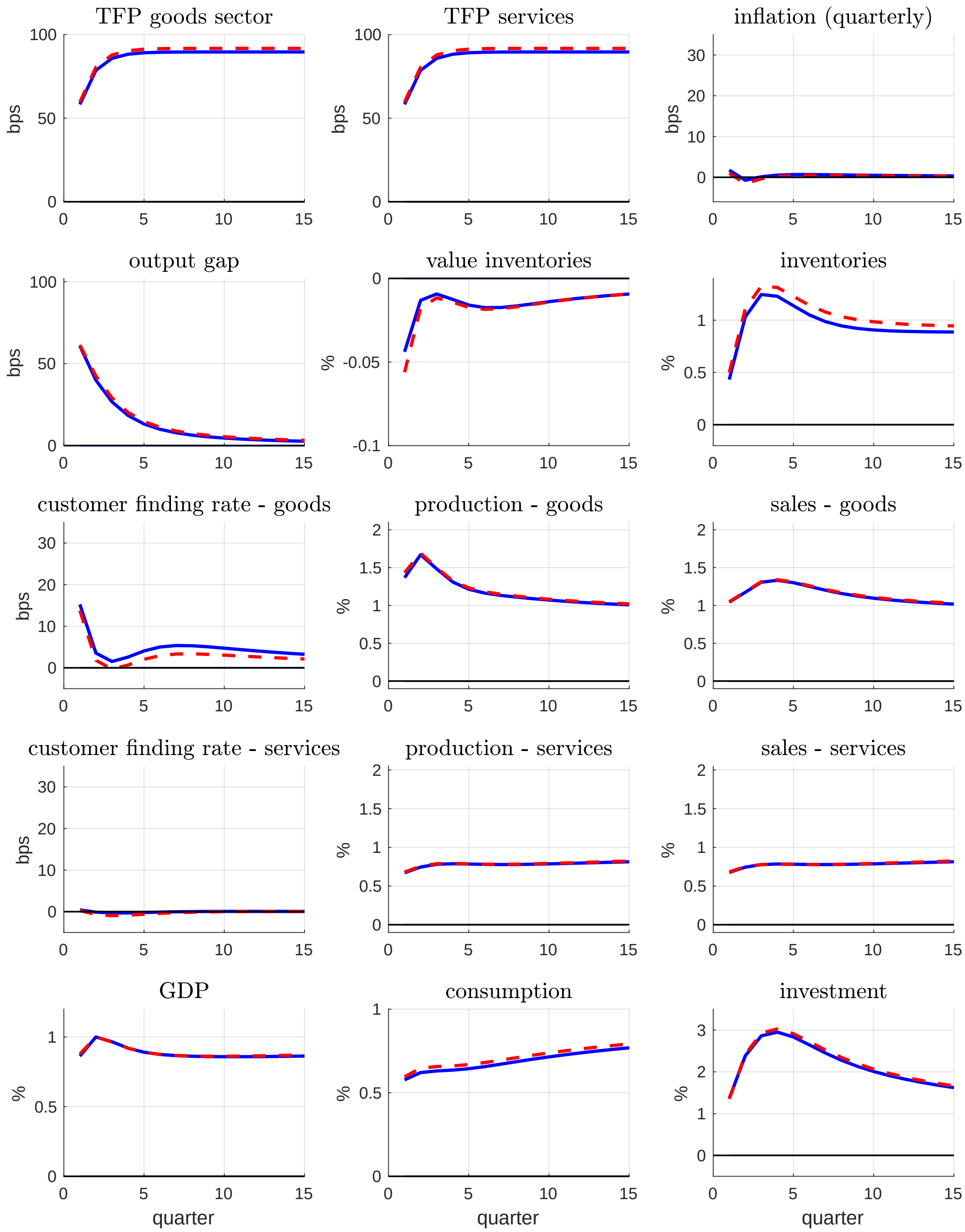
[111]The long-run response in the two sectors has to be equal to ensure balanced growth. When TFP in the service sector responds with a lag, then this would be initially similar to the case in which it responds by less.

Figure 11: monetary-policy shock; benchmark parameters



Notes. The solid lines correspond to the case when $\nu_g = 0.3469$, $\nu_s = 0.6713$, and $\Gamma_y = 0.012$, which are the values at the mode of the posterior. The dashed lines correspond to the case when $\nu_g = \nu_s = 0.565$ and $\Gamma_y = 0.03$, which are values in the middle of the calibrated admissible area. Shock size is calibrated to ensure a peak 1 percent increase in GDP in both cases.

Figure 12: TFP shock; benchmark parameters



Notes. The solid lines correspond to the case when $\nu_g = 0.3469$, $\nu_s = 0.6713$, and $\Gamma_y = 0.012$, which are the values at the mode of the posterior. The dashed lines correspond to the case when $\nu_g = \nu_s = 0.565$ and $\Gamma_y = 0.03$, which are values in the middle of the calibrated admissible area. Shock size is calibrated to ensure a peak 1 percent increase in GDP in both cases.

Table 1 presents the values of key moments together with their empirical counterpart. The same two set of parameter combinations are considered. All moments are based on the cyclical components of the series obtained using the HP filter. Model moments are the average across 10,000 replications of length 212, that is, the same length as our empirical data set. The number in brackets displays the standard deviation across replications.[112]

Economies with one type of shock. We start with a discussion of the results when fluctuations are due to either only monetary-policy shocks or only TFP shocks. Like the IRFs, the moments also make clear that the model can replicate key inventory facts for both types of shocks. Nevertheless, the results in the table indicate – like the IRFs do – that there are differences between an economy with monetary-policy driven fluctuations and one in which business cycles are due to TFP shocks.

Relative volatilities. The first key observation is related to volatilities. A procyclical customer-finding rate is a natural and robust outcome in response to demand shocks. A key message of this paper is that it is also a "natural" outcome in response to TFP shocks *if* the value of an inventory good is countercyclical which will be the case if the intertemporal marginal rate of substitution is countercyclical. Nevertheless, there is a quantitative difference and the customer-finding rate is less responsive to TFP than to monetary-policy shocks relative to the induced output response.

The problem with a strongly volatile customer-finding rate in response to monetary-policy shocks is that sales may turn out to be more volatile than output, whereas the observed value of ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ is significantly bigger than 1, an important stylized fact from the inventory literature. By contrast, the problem with a less volatile customer-finding

[112]Kydland and Prescott (1982) report model outcomes in the same way. It has two advantages relative to the alternative of presenting population moments, that is, the outcome consistent with a sample of infinite length. First, it makes more sense to compare each data moment with the *average* across replications of samples with similar length than with the population moment, since the observed moments are also obtained using a small sample in which the mean is the one for this small sample (and not the unknown long-run mean). For first-order moments one would get the same answer. For higher-order moments, however, the average of a statistic across replications does not have to be equal to the population moment (which would be equal to the average of a finite, but very long sample). For example, if a variable is very persistent, then the average of a set of variances calculated using small samples will be lower than the unconditional variance, since the means over the shorter samples adjust which reduces the variance. This actually turns out not to matter much for our model. One exception is the autocorrelation of TFP growth as discussed in footnote 94. The second advantage of this approach is relevant. Even if the underlying model is the true data-generating process, then the outcomes for a statistic of interest could still vary substantially across replications and, thus, not always be close to the empirical estimate. The reason is that the random numbers used to generate the model data according to the model differ, of course, across replications. By reporting standard deviations across replications, we gain insight into the question *how likely* it is that the model generates a statistic that is similar to the empirical estimate. So a proper evaluation of the model takes into account both the standard errors of the estimated moment, and – following Kydland and Prescott (1982) – also the standard deviations across replications.

rate in response to TFP shocks is that the implied value of ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ may be too big.

As shown in table 1, the value of ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ for the monetary-policy-shocks economy is equal to 1.081 for the parameter values in the middle of admissible range. This value is inside the 95% confidence band. By contrast, the value is only equal to 1.006 when estimated parameter values are used, which is substantially below the lower bound of the 95% confidence interval which is equal to 1.073. This does not mean that the model with these parameter values is inconsistent with the observed value of this moment. It just means that the model would need some TFP shocks to get the implied value inside the 95% confidence area. For TFP shocks, the implied value for ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ is somewhat above the upper bound of the 95% confidence interval. This is true for both parameter sets considered in table 1 but also for other parameter combinations in the admissible area. Again, this does not mean that the model is not consistent with this statistic. It just means that some monetary-policy shocks are needed to lower it. In section 4.7.2, we will provide a quantitative analysis to determine the necessary relative importance of monetary-policy and TFP shocks for the model to be consistent with this (and one other key) inventory statistic.

Correlation coefficients. Next, we turn our attention to correlation coefficients. In economies with only one type of shock, the *sign* of a correlation coefficient is much more important than its *magnitude*.[113] A striking outcome is that the sign of the correlation between the customer-finding rate with the beginning-of-period inventory stock, $\text{COR}(f_g^f, x_{-1})$, is positive for monetary-policy shocks and negative for TFP shocks. And this is a very robust result that we also found to hold for examined parameter value combinations that are outside the calibrated admissible area. The reason for the different sign is the following. For a TFP shock, the responses of the two variables move in opposite directions *after* the initial response. The change in the customer-finding rate is temporary and returs gradually to its steady-state value. By contrast, the inventory stock follows the time path of TFP and continues to increase before it stabilizes. This implies quite different trends and different cyclical components.[114] By contrast, the responses of the two series are both temporary following a monetary-policy shock.

The 95% confidence area of the estimated $\text{COR}(f_g^f, x_{-1})$ has a lower bound equal to -0.429 and an upper bound equal to -0.017. So although the point estimate is

[113]If two variables have different dynamics properties, then the correlation coefficient does not have to be minus or plus one and so the magnitude may still have some information.

[114]The customer-finding rate displays a small (barely visible) negative response in the third period when calibrated parameter values are used. For larger values of Γ_y, this temporary negative response increases somewhat. This would make the correlation coefficient even more negative. But key in understanding the negative value is the different trajectory after the initial uptick. Specifically, the response of the inventory stock is the smallest on impact after which it continues to increase and it does not return back to its initial value, but – like the TFP level – reaches a new higher steady-state level. Consequently, this means that initially the inventory stock is below its new trend value leading to a *negative* cyclical component initially. By contrast, the customer-finding rate's largest response happens on impact after which it basically returns back to its steady state.This implies a *positive* cyclical component when the cyclical component of the inventory stock is negative.

significantly different from zero, a small negative value is consistent with the data. Since it is a robust result that the value of the correlation coefficient is positive for monetary-policy shocks, at least some TFP shocks will be needed to get the model implied value inside this confidence region. Note that this correlation coefficient is not considered at all in our calibration procedure, a fact we exploit in section 4.7.2 to investigate the identifying information of inventory data to determine the relative importance of monetary-policy and TFP shocks.

Economies with both types of shocks. The table also reports outcomes when the fluctuations are due to both types of shocks. The discussion above already hinted at the need to have both shocks in order to get an excellent *quantitative* match with estimated moments. This means we have to take a stand on the relative magnitude of the two innovation standard deviations, ${}^{\sigma_R}/_{\sigma_A}$.[115] When we estimate ν_g, ν_s, and Γ_y, then we simultaneously estimate σ_R and σ_A. The results reported in the table for the calibrated parameter values are based on a value for ${}^{\sigma_R}/_{\sigma_A}$ such that model moments are consistent with the empirical estimates of ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ and $\text{COR}(f_g^f, x_{-1})$.There is a tension here. Since ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ is above the 95% confidence area for TFP shocks, a high value of ${}^{\sigma_R}/_{\sigma_A}$ is helpful. However, since $\text{COR}(f_g^f, x_{-1})$ is outside the 95% confidence area for monetary-policy shocks, a low value for ${}^{\sigma_R}/_{\sigma_A}$ is helpful. But it is possible to choose a value such that the model is consistent with both empirical findings.[116]

As documented in table 1, the model can generate values for seven of the nine inventory, production, and sales moments considered that are inside the 95% confidence band for the calibrated parameters. And the correlation of inventories and production is almost inside. Of course, we have chosen three parameters, ν_g, Γ_y, and ${}^{\sigma_R}/_{\sigma_A}$ to help with this and the moments are related to each other. We have challenged ourselves, however, by restricting parameter combinations to be in a calibrated region such that the model is consistent with key inventory facts for *both* the monetary policy and the TFP shock.

The one moment that the model does not capture well is the standard deviation of the inventory-sales ratio relative to the standard deviation of sales, ${}^{\sigma_{x/sg}}/_{\sigma_{sg}}$. The model can generate values inside the 95% confidence band for this moment, but at the cost of doing worse for other moments.[117] When estimated parameter values are used, then

[115] Only the ratio matters, since we solve the model with a first-order perturbation solution.

[116] We set ${}^{\sigma_R}/_{\sigma_A}$ equal to 0.5921. For the parameter combination in the middle of the admissible area considered in the table, the value of ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ is equal to 1.175 and $\text{COR}(f_g^f, x_{-1})$ is equal to -0.0184. That is, the tension between the two moments can still be resolved, but only just since values are close to the edges of the two 95% confidence intervals. As documented in section 4.7.2, *somewhat* higher values of ν_g can also resolve this tension if one increases Γ_y. For lower values of ν_g, the tension is easier to deal with and it is possible to have values for these two statistics that are deeper inside the 95% confidence interval. A more detailed discussion is given in section 4.7.2.

[117] This is not a popular moment in the inventory literature. We include it because it is discussed in Kryvtsov and Midrigan (2013). Recall that the inventory-sales ratio is a nonlinear function of just the customer-finding rate. Thus, volatility of the inventory-sales ratio relative to the volatility of sales is related to the volatility of the customer-finding rate relative to the volatility of production.

this statistic is closer to being inside the 95% confidence interval with a t-statistic of 2.5.[118] But now the correlation of the customer-finding rate with the lagged inventory stock is significantly different from its empirical counterpart.

The model has few bells and whistles, it has only two types of shocks, and there is no measurement error. Given the challenges that the literature has faced to build a business-cycle model that can replicate key inventory facts, it is promising that our relatively simple framework is successful in several dimensions.

4.7.2 Identifying power of inventory data: Demand versus supply shocks

Several papers in the business-cycle literature, address the question what type of shocks are the most important for business-cycle fluctuations.[119] The purpose of this section is *not* to put forward a new answer that competes with existing ones. Our model is too simple to do this, given that it has only two shocks, no heterogeneity, no financial frictions, no financial sector, and no fiscal variables. Instead, the purpose is to document that there is identifying information in inventory data for macroeconomic questions such as this one.[120]

We focus on two inventory moments. The first is the standard deviation of output relative to the standard deviation of sales, $\sigma_{y_g}/\sigma_{s_g}$. Recall that this moment is closely related to the volatility and cyclicality of the customer-finding rate and the accumulation of inventories. Although, the IRFs presented in section 4.7.1 document that the *qualitative* responses are quite similar for monetary policy and TFP shocks, there

As documented in the table, the model-generated values are below the point estimates for both, but outside the confidence band for the former and inside for the latter. There are two ways to get the estimate for $\sigma_{x/s}/\sigma_s$ inside the confidence band. The first is to increase σ_R/σ_A, but then the model fails in several other dimensions. Specifically, then the volatility of inventories relative to the volatility of sales is too low and the sign of the correlation of the customer-finding rate with the inventory stock is the opposite of what is observed in the data. The second is to lower ν_g which will increase the volatility of the customer-finding rate and the inventory-sales ratio. But then the model can no longer generate key inventory facts following a monetary-policy shock, specifically production is then substantially less volatile than sales.

[118] When we not only take into account sampling uncertainty of the observed value, but also take seriously that model moments for small samples display variations, then there is a match when results are based on estimated parameters. That is, although the mean across replications is outside the 95% confidence band it is not that far outside and the standard deviation across replications is nontrivial.

[119] A classic example is Smets and Wouters (2007) in which a structural business-cycle model is estimated and used to determine the importance of the seven structural shocks for the business-cycle fluctuations of model variables. The set of possible shocks considered in the literature has increased substantially over time. Related to our framework with a monetary (demand) and a TFP (supply) disturbance is Forni and Gambetti (2021) which finds that both demand as well as supply shocks have sizable effects on GDP.

[120] We are not the first to make this point. Kahn (1987) also addresses this question although the demand shock included is an idiosyncratic one capturing demand uncertainty that individual firms face. In Wang et al. (2014), it is documented that the role of demand shocks increases when inventory data are included in the analysis although supply shocks remain the dominant driving force for aggregate fluctuations, except possibly in the very short run.

are *quantitative* differences. Specifically, the increase in the customer-finding rate is smaller for a TFP shock, which in turn implies a much bigger response of output relative to sales (and a larger increase in inventories), that is, a higher value for σ_{yg}/σ_{sg}. The second moment is the correlation of the customer-finding rate with the beginning-of-period inventory stock, $\text{COR}(f_g^f, x_{-1})$.[121] Whereas the first moment is a "classic" in the inventory literature, the importance of the correlation with lagged inventories has not been highlighted. The discussion in section 4.7.1 makes clear, however, that this moment is useful because even the sign differs for the two types of shocks.

The exercise. Throughout this section, we restrict values of ν_g, ν_s, and Γ_y to be such that they are in the calibrated admissible area. The key exercise is to consider different values for the innovation standard deviations, σ_R and σ_A, and check for which combinations the model is consistent with the observed values of σ_{yg}/σ_{sg} and $\text{COR}(f_g^f, x_{-1})$, taking into account sampling uncertainty.[122] Let $[M]^x$ denote the value of a moment generated by shocks of type x where $x \in \{\text{MP,TFP}\}$.

Figure 13 displays the results. The horizontal-axis variable is the parameter ν_g which is assumed to be equal to ν_s. The vertical axis displays the fraction of GDP fluctuations that is due to TFP fluctuations for that value of ν_g. It is a range because (i) we allow Γ_y to vary and take on values that are in the admissible area for that value of ν_g and (ii) "being consistent with" the estimated value of the moment takes into account sampling variation.[123]

The σ_{yg}/σ_{sg} restriction. The blue area in figure 13 displays possible outcomes for the role of TFP shocks for GDP fluctuations when the model is consistent with this relative volatility restriction. The bottom line is that the role of TFP shocks could be zero, but also as high as 94.0%. As discussed above, the customer-finding rate could be too volatile in response to monetary policy shocks, which would reverse the desired relative volatility of output and sales. But our calibration procedure is such that parameters, and especially ν_g, are chosen such that this does not happen. Consequently, the model is consistent with just monetary policy shocks when parameters take on values in the admissible area. For TFP shocks, it is also true that output is more volatile than sales in the admissible area (and in fact for quite a few values outside as well). But the value of $[\sigma_{yg}/\sigma_{sg}]^{\text{TFP}}$ is actually slightly above the upper bound of the 95% confidence interval for the estimate of σ_{yg}/σ_{sg}. Consequently, the model will need some monetary-policy shocks to match the empirical counterpart. At the upper bound for ν_g in the admissible area,

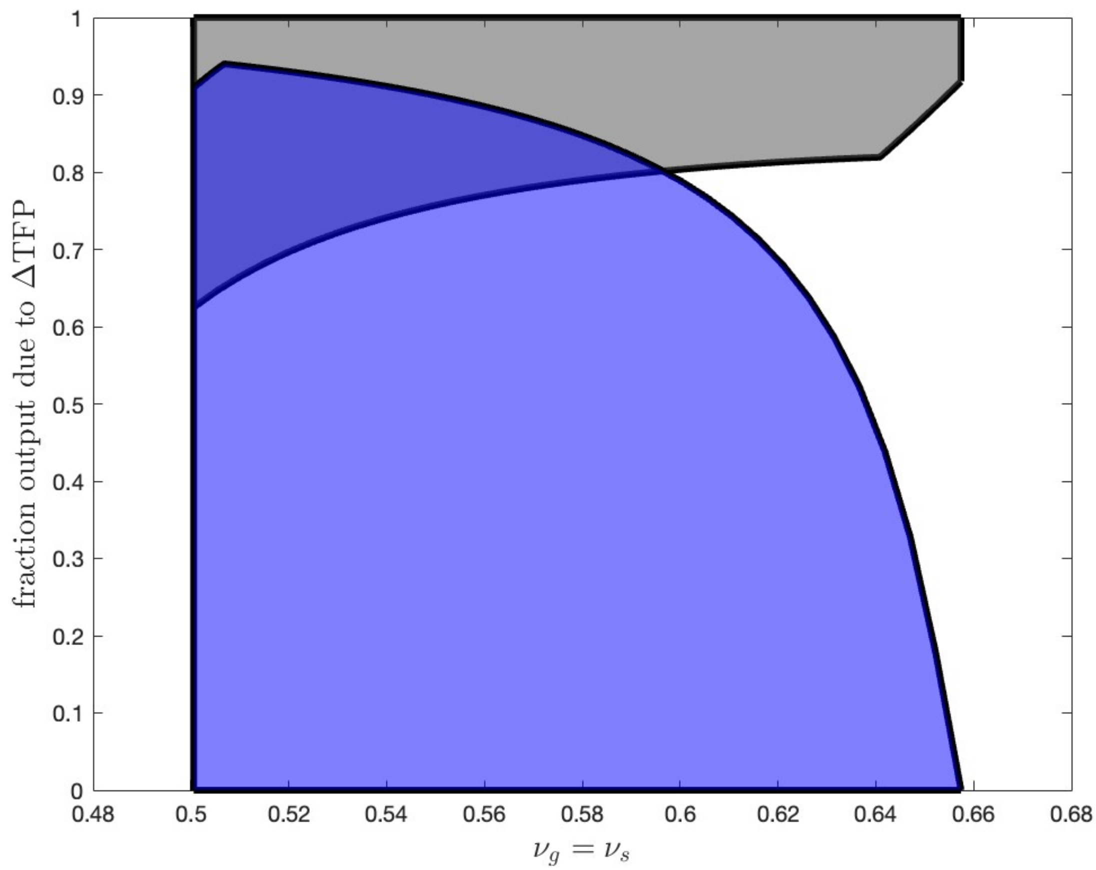
[121]In terms of the notation of our model, x_{t-1} is the inventory stock at the end of period $t-1$ and brought into period t.

[122]The advantage of a first-order perturbation solution method is that only the ratio of the two standard deviations matters for correlation coefficients and ratios of relative standard deviations.

[123]Consistent with the construction of the admissible area, the lower bound of σ_{yg}/σ_{sg}, we use a value of 1 and not the lower bound of the 95% confidence interval, which is equal to 1.073. In appendix D.3, we discuss this alternative and explain how this would only strengthen our conclusion regarding the importance of TFP shocks.

the value of $[\sigma_{yg}/\sigma_{s_g}]^{\text{MP}}$ has reached the upper bound of the 95% confidence interval and so there is no room for TFP shocks. But as ν_g falls, the amount of space for TFP shocks increases rapidly. An in-depth explanation of the figure including the non-monotonicity in the top border of the blue area is given in appendix D.3.

Figure 13: monetary-policy shock; benchmark calibration



Notes. The blue area indicates for a given value of ν_g which fraction of GDP can possibly be generated by TFP shocks when the model is consistent with the empirical estimate of σ_{yg}/σ_{s_g} taking into account sampling uncertainty and the range of values for Γ_y that is in the admissible area for that value of ν_g. The grey area is constructed in the same way, but uses the empirical outcome for $\text{COR}(f_g^f, x_{-1})$. Since there are only two shocks in the model, 1 minus the value on the vertical axis represent the fraction of GDP that can possibly be generated by monetary-policy shocks.

The COR(f_g^f, x_{-1}) **restriction.** The grey area in figure 13 displays the range of values for the relative importance of TFP shocks when we consider the estimate of $\text{COR}(f_g^f, x_{-1})$ (and its sampling uncertainty). The bottom line is that matching the observed value of this moment requires that TFP shocks are responsible for at least 62.5% of the fluctuations in GDP and their role could be as high as 100%. The explanation is the following. For any combination in the admissible area, it is true that the implied value for $[\text{COR}(f_g^f, x_{-1})]^{\text{TFP}}$ is inside the 95% confidence band and the value for $[\text{COR}(f_g^f, x_{-1})]^{\text{MP}}$ is outside. Since the value for $[\text{COR}(f_g^f, x_{-1})]^{\text{TFP}}$ is never at the bounds of the 95% confidence interval, there is always some space for monetary policy shocks.

A detailed explanation for the particular lower bound of the grey area is given in appendix D.3, but the main insights are the following. What matters for the lower border of the grey area is the value of Γ_y. The value of ν_g actually doesn't matter that much for the outcome of this correlation coefficient, except that as ν_g increases, the maximum possible value of Γ_y decreases, first gradually along the upper bound of the upper bound at the admissible area and then sharply along the right border. And a fall in Γ_y leads to a sharp increase in $[\mathrm{COR}(f_g^f, x_{-1})]^{\mathrm{TFP}}$ and a more moderate increase in $[\mathrm{COR}(f_g^f, x_{-1})]^{\mathrm{MP}}$ and both imply that there is less space for monetary-policy shocks.[124]

Combining the restrictions. The most interesting part of the figure is the intersection of the blue and the grey area which indicates that the role for TFP shocks must play a dominant role if the model is consistent with both restrictions. It tells us that the fraction of GDP fluctuations due to TFP shocks is at least 62.5% and at most 94%. Moreover, the results also imply a restriction on possible values for ν_g. That is, to be consistent with both empirical restrictions, the value of ν_g has to be less than 0.595, since the model cannot satisfy both restrictions for higher values no matter what combination of shock standard deviations is chosen.

Why does observed inventory behavior favor TFP shocks in a model with a goods-market friction? The result that inventory behavior favors TFP shocks in a model with a goods-market friction may – at first sight – be surprising. After all, an increase in output has a direct negative effect on the customer-finding rate which is the opposite of observed cyclical behavior. A key result of our paper is that this is no longer true if the value of storing a good in inventories is sufficiently countercyclical. In fact, for appropriate choice of ν_g – and to some extent Γ_y – the model can replicate that output is more volatile than sales for both monetary-policy and TFP shocks. As discussed in section 4.7.1, however, the negative value of $\mathrm{COR}(f_g^f, x_{-1})$ is consistent with TFP shocks, but not with monetary-policy shocks.

Qualifying comments. The reader may not agree with our choice of moments. And perhaps prefer model implications based on the parameter values from the full-information estimation.[125] But we hope to have made clear to all readers that key inventory moments do have identifying information for model parameters like σ_A/σ_R

[124] A higher value of Γ_y means more dampening, but this has a stronger impact on the procyclicality of the customer-finding rate than the inventory stock, especially for TFP-driven fluctuations.

[125] For our Bayesian full-information estimation exercise, we find that the fraction of GDP fluctuations due to TFP shocks is equal to 62.5% when the estimated posterior modes for σ_R and σ_A are used and equal to 52.1% when monetary-policy shocks are given an advantage by setting σ_R equal to the upper bound of its 90% HPD interval and σ_A to the lower bound of its 90% HPD interval.

and, thus, for the role of the different types of shocks for fluctuations in economic aggregates.

4.7.3 The quantitative importance of goods-market frictions

In section 4.7.1, we discussed key model properties related to the behavior of inventory, production, and sales. By construction of the calibrated admissible area, the model replicates key facts stressed in the inventory literature for parameter combinations inside this region. In this section, we ask the question whether variations of parameter values *inside* the admissible area matter for the volatility of GDP. We will show that this is the case, even when restricted to remain within the calibrated admissible area. The advantage of focusing on the calibrated admissible region instead of the posterior density is that we get a transparent link from the possible values of model moments (taking into account sampling uncertainty) to the parameters capturing variability in the severity of the goods-market frictions, ν_g and ν_s, to what this means for magnification of shocks.

We do not consider variations in Γ_y in the discussion, since they do not matter much for the magnification of the shocks. Thus, to focus the analysis we discuss model properties using parameter values at the two bottom corners of the admissible area where the value of Γ_y is constant (and equal to zero).[126]

Increasing the values of ν_g and ν_s reduces the impact of tightness on the customer-finding rates. Consequently, firms would face less of a dampening effect if they increase production. Thus, this channel would imply higher levels of volatility at higher levels of ν_g and ν_s. On the other hand, an increase in the value of these parameters would make the cost of searching from the buyers' perspective *more* sensitive. This would dampen the increase (decrease) in effort during expansions (recessions) and indicate that increases in ν_g and ν_s could also lead to lower volatility.

Table 5 reports the standard deviations of the cyclical components of GDP, output in the goods sector, $y_{g,t}$, and potential output in the service sector, $y_{s,t}$, for an economy with either only monetary-policy shocks or only TFP shocks.[127] It reports these numbers, not only for the value of ν_g at the two bottom corners of the admissible area, but also when ν_g is equal to 0.595, which is the highest value consistent with the exercise in the previous subsection. Thoughout the discussion, we impose that $\nu_g = \nu_s$. The table also reports corresponding numbers for the customer-finding rate in the two sectors.

The main conclusion is that as ν_g increases from the lowest value in the admissible area to the highest value, there are modest but nonnegligible increases in volatility. The largest increases are observed for goods production. Specifically, its standard deviation

[126]The same conclusions can be drawn when we consider the two top corners, but then the discussion is less clean because the values of Γ_y at those two corners is slightly different.

[127]Both $y_{g,t}$ and $y_{s,t}$ are the values that come out of the production function. The goods-market friction implies that only a fraction of this amount will lead to sales and that is true for both sectors. The difference is that unsold goods in the goods sector end up in inventories and will lead to sales in future periods, although there is a depreciation of goods over time.

increases with 16%, both when the economy is driven by monetary-policy shocks and when they are driven by TFP shocks. The increases in GDP volatility are 6.6% and 11.4% when shocks are due to monetary-policy and TFP shocks, respectively.[128]

When we increase ν_s and ν_g, then the volatility of potential output of the service sector does increase for the model with TFP shocks, but is basically not affected for the one with monetary-policy shocks. As pointed out above, the volatility could either increase or decrease. And the direction could very well depend on the type of shock. But the intriguing observation is that – despite the Leontief structure – the goods sector and service sector can respond differently to a change in this parameter. That is, volatility displays a notable increase in the goods sector and not in the service sector with monetary-policy shocks. What is the reason and why do the two sectors respond in a similar way when fluctuations are due to TFP shocks?

Table 5: Goods-market friction variations and aggregate volatility

TFP shocks			
	$\nu_g = \nu_s = 0.5068$	$\nu_g = \nu_s = 0.595$	$\nu_g = \nu_s = 0.6574$
GDP	0.01000	0.01090	0.01139
y_g	0.01648	0.01821	0.01918
y_s	0.00794	0.00852	0.00881
f_g^f	0.00126	0.00110	0.00099
f_s^f	0.00012	0.00011	0.00010
monetary-policy shocks			
	$\nu_g = \nu_s = 0.5068$	$\nu_g = \nu_s = 0.595$	$\nu_g = \nu_s = 0.6574$
GDP	0.01000	0.01041	0.01066
y_g	0.01936	0.02132	0.02249
y_s	0.00593	0.00595	0.00594
f_g^f	0.00418	0.00312	0.00249
f_s^f	0.00351	0.00305	0.00273

Notes. This table documents how the volatility of GDP, production in the goods sector, $y_{g,t}$, potential production in the service sector, $y_{s,t}$, and the customer-finding rates in the two sectors, $f_{g,t}^f$ and $f_{s,t}^f$ vary with changes in $\nu_g(= \nu_s)$. An increases in this parameter reduces the curvature of the customer-finding rate as a function of tightness and, thus, the volatility of the customer-finding rate. The value of Γ_y is kept constant and set equal to 0; $\nu_g = \nu_s = 0.5068$ corresponds to the bottom-left corner of the calibrated admissible area, $\nu_g = 0.6574$ to the bottom-right corner, and $\nu_g = \nu_s = 0.595$ to the highest value of $\nu_g(= \nu_s)$ in the calibrated admissible area such that there is a mix of monetary policy and TFP shocks such that the model is consistent with the two key restrictions imposed in section 4.7.2. Since these corners are obtained using population moments (calculated using a sample of 100,000 observations), the numbers reported here are calculated in the same way. The standard deviations of the innovations are set such that the standard deviation for the cyclical component of GDP is normalized to be equal to 0.01 for the lowest value of $\nu_g(= \nu_s)$.

In the remainder of this section, we explain why $y_{g,t}$ and $y_{s,t}$ can move independently even though we impose a Leontief structure for consumption and investment. There

[128]The standard deviations of the exogenous innovations are set such that the volatility for the cyclical component of GDP is normalized to be equal to 0.01 for the low value of ν_g.

are three reasons why this is the case. The first reason is that the relative importance of consumption and investment is different for the two sectors and the two expenditure components have different variability over the business-cycle. The second reason is that the consumption good is partially durable and the Leontief structure applies to the stock, not the purchases of the consumption good. But even if there was only one type of good, say non-durable consumption, then movements in $y_{g,t}$ and $y_{s,t}$ would still not have to be synchronized.[129]

4.7.4 Alternative specifications

In appendix E, we discuss alternative specifications of our model. In appendix E.1, we discuss several alternative specifications of the TFP process. Specifically, we consider model properties when TFP is a stationary process. That is, one in which TFP eventually returns to its pre-shock level. As discussed in section 3, our specification is more consistent with the data, but a stationary TFP process is quite popular in the business-cycle literature.[130] We also consider the case when TFP in the service sector lags TFP in the goods sector.[131]

Given the importance of the responsiveness of monetary policy to the output gap, i.e., the parameter Γ_y, we also consider the results when we use an estimated Taylor rule. This is discussed in appendix E.2.

In our benchmark specification, we assume that the relative importance of goods and services for search costs is equal to the observed ratio of consumption of goods

[129] Suppose that the Leontief structure is such that $c_{g,t} = c_{s,t}$. This implies that $f^f_{g,t}(y_{g,t} + (1-\delta_x)x_{t-1}) = f^f_{s,t}y_{s,t}$. Depending on the behavior of $f^f_{g,t}$, $f^f_{s,t}$, and x_{t-1}, the behavior of $y_{g,t}$ and $y_{s,t}$ can still be quite different. Since $f^f_{g,t}$ is affected by $\lambda^f_{x,t}$ and $f^f_{s,t}$ is not, these two customer-finding rate will not behave in exactly the same way. The special case when $f^f_{g,t}(y_{g,t} + (1-\delta_x)x_{t-1}) = f^f_{s,t}y_{s,t}$ also sheds light on the question why $y_{s,t}$ is more closely linked to $y_{g,t}$ in response to TFP shocks. The discussion of the IRFs in section 4.7.1 makes clear that the response of x_t is only temporary for monetary-policy shocks, but large and persistent in response to a TFP shock. The larger and more persistent response of x_{t-1} implies a larger and more persistent response of $f^f_{g,t}(y_{g,t} + (1-\delta_x)x_{t-1})$ which pushes up sales and production in the service sector because of the constraint that $c_{g,t} = c_{s,t}$.

[130] Working with a stationary process is easier than working with a non-stationary ones. A justification for adopting the simpler stationary process is given in Christiano and Eichenbaum (1990) which shows that business-cycle properties for a model with a persistent but stationary TFP process are similar to one in which the TFP process has a unit root. But that paper only considers typical business-cycle variables. The conclusion turns out to be not true for the behavior of inventories in our model since it depends crucially on an asset price, namely the end-of-period value of inventories, λ^f_x. Although the *change* in consumption is stationary in our model, its response to TFP shocks is persistent as is made clear by the persistent response of λ^f_x. And as documented in Bansal and Yaron (2004), this is important for asset prices even if this persistent component is quantitatively not that important.

[131] To ensure balanced growth, the long-run percentage response of TFP in the service sector must be equal to the long-run response of TFP in the goods sector. The exercise in which TFP in the service sector responds with a lag is the best *feasible* stand-in for the exercise in which TFP in the service sector responds by less than TFP in the goods sector when there is an aggregate productivity shock.

relative to consumption of services. But it does not seem implausible that services are more important. In appendix E.3, we show that our results are robust when services are more important. In fact, results are even very similar when search costs consists solely of services.

Based on data of the Euro-Area capacity-utilization index and the inventory-sales ratio we reached the conclusion that the customer-finding rate is substantially higher in the service sector. In appendix E.4, we consider the case when the means are the same across the two sectors. In appendix E.5, we discuss alternative assumptions regarding the curvature parameter in the search-friction function.

Finally, we consider lower maintenance costs of holding inventories in appendix E.6. This is important, since these may very well have fallen over time.

Summary of robustness exercises. The appendix makes clear that the only variation that matters is the persistence of the TFP process. When deviations in the TFP level are temporary and TFP is assumed to revert back to its pre-shock level, then it is still possible to generate a procyclical customer-finding rate response to TFP shocks, but it is a less robust outcome, at least in our relatively simple model which excludes modifications such as the inclusion of habits to robustly generate a hump-shaped consumption response.

5 Areas for future research

We have shown that our relatively simple framework is capable of generating behavior that is consistent with key inventory, production, and sales data for both monetary policy and TFP shocks. Given that the cyclical behavior of investment in inventories is *and* systematic *and* quantitatively important, our model can be used to shed light on a variety of business-cycle related questions. In this paper, we have already shown how inventory facts have identifying information for what drives economic fluctuations and that the severity of the goods-market friction matters for the volatility of output even if parameters are restricted to remain in the relatively narrow range of admissible values.

Since a large fraction of value added is generated in the service sector, we decided to add a service sector. Whereas the inventory-sales ratio provides a direct measure of the fraction of available goods sold in the goods sector, no comparable measure is available for the service sector.[132] But sell frictions are likely to be relevant in the service sector as well. And this is indeed what we assumed in this paper. But it would, of course, be great if reliable data would become available to study the cyclical behavior of the customer-finding rate for the service sector, that is, how the gap between actual and potential sales move over the business cycle.

[132] The survey data for the European Union discussed in section 2 provides some insights, but even if this is the right measure, then it is only available for a short sample.

The analysis in the main text is based on the assumption that a productivity shock affects TFP in the two sectors in the same way. This is a sensible benchmark and allows us to show that the interaction between the two sectors is then quantitatively not important. In appendix E.1, however, we show that stronger interaction effects are possible when an aggregate TFP shocks causes changes in the *relative* productivity levels of the two sectors. And this is true *even* though the Leontief structure is still in place. So another place where additional data would be helpful to make progress in understanding the role of sell frictions is knowing whether sectoral TFP fluctuations are synchronized and how their magnitudes compare.

References

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. The Journal of Finance LIX, 1481–1509.

Beaudry, P., Portier, F., 2006. Stock prices, news and economic fluctuations. American Economic Review 96, 1293–1307.

Bils, M., Kahn, J.A., 2000. What inventory behavior tells us about business cycles. American Economic Review 90, 458–481.

Blinder, A., Hotz-Eakin, D., 1986. Inventory fluctuations in the united states since 1929, in: Gordon, R.J. (Ed.), The American Business Cycle: Continuity and Change. University of Chicago Press, Chicago, pp. 183–236.

Blinder, A.S., Maccini, L.J., 1991. Taking stock: A critical assessment of recent research on inventories. Journal of Economic Perspectives 5, 73–96.

Cao, J., Cui, C., Dinger, V., Holm, M.B., Kang, S., 2022. Identifying the depreciation rate of durables from marginal spending responses. Norges Bank Working Paper 1/2022.

Christiano, L.J., Eichenbaum, M., 1990. Unit roots in real gdp: Do we know, and do we care. Carnegie-Rochester Conference Series on Public Policy 32, 7–62.

Coen-Pirani, D., 2004. Markups, aggregation, and inventory adjustment. American Economic Review 94, 1328–1353.

Den Haan, W.J., Drechsel, T., 2021. Agnostic structural disturbances (asds): Detecting and reducing misspecification in empirical macroeconomic models. Journal of Monetary Economics 117, 258–277.

Den Haan, W.J., Levin, A., 1997. A practioner's guide to robust covariance matrix estimation, in: Handbook of Statistics 15. Elsevier, Holland, pp. 291–341.

Eichenbaum, M., 1989. Some empirical evidence on the production level and production cost smoothing models of inventory investment. American Economic Review 79, 853–864.

Erceg, C.J., Henderson, D.W., Levin, A.T., 2000. Optimal monetary policy with staggered wage and price contracts. Journal of Monetary Economics 46, 281–313.

Fernald, J., 2014. A quarterly, utilization-adjusted series on total factor productivity. Federal Reserve Bank of San Francisco Working Paper 2012-19 (updated).

Forni, M., Gambetti, L., 2021. Policy and business cycle shocks: A structural factor model representation of the US economy. Journal of Risk and Financial Management 14, 1–21.

Fuhrer, J.C., 2000. Habit formation in consumption and its implications for monetary-policy models. American Economic Review 90, 367–390.

Gali, J., 2015. Monetary policy, inflation, and the business cycle: an introduction to the New Keynesian framework and its applications. Princeton University Press, Princeton, US.

Hansen, G.D., 1985. Indivisible labor and the business cycle. Journal of Monetary Economics 16, 309–327.

Huo, Z., Ríos-Rull, J.V., 2020. Sticky wage models and labor supply constraint. American Economic Journal: Macroeconomics 12, 284–318.

Kahn, J.A., 1987. Inventories and the volatility of production. American Economic Review 77, 667–679.

Kahn, J.A., 1992. Why is production more volatility than sales? Theory and evidence on the stockout-avoidance motive for inventory-holding. Quarterly Journal of Economics 107, 481–510.

Khan, A., Thomas, J.K., 2007. Inventories amd the business cycle: An equilibrium analysis of (S,s) policies. American Economic Review 97, 1165–1188.

Kryvtsov, O., Midrigan, V., 2013. Inventories, markups, and real rigidities in menu cost models. Review of Economic Studies 80, 249–276.

Kydland, F., Prescott, E.C., 1982. Time to build and aggregate fluctuations. Econometrica 50, 1345–1370.

Lindé, J., 2009. The effects of permanent technology shocks on hours: Can the RBC-model fit the var evidence. Journal of Economic Dynamics and Control 33, 597–613.

Mazelis, F., Motto, R., Ristiniemi, A., 2023. Monetary policy strategies for the euro area: optimal rules in the presence of the ELB. Working Paper Series No. 2797.

McMahon, M., 2011. Inventories in motion: A new approach to inventories over the business cycle. Unpublished Manuscript, University of Warwick.

Miranda-Agrippino, S., Ricco, G., 2017. The transmission of monetary policy shocks. Bank of England Staff Working Paper No. 657.

Ramey, V.A., 1989. Inventories as factors of production and economic fluctuations. American Economic Review 79, 338–354.

Ramey, V.A., 1991. Nonconvex costs and the behavior of inventories. Journal of Political Economy 99, 306–334.

Ramey, V.A., 2016. Macroeconomic shocks and their propagation, in: Taylor, J.B., Uhlig, H. (Eds.), Handbook of Macroeconomics. Elsevier. volume 2, pp. 71–162.

Ramey, V.A., West, K.D., 1999. Inventories, in: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics. Elsevier. volume 1, pp. 863–923.

Ravn, M.O., Uhlig, H., 2002. On adjusting the Hodrick-Prescott filter for the frequency of observations. Review of Economics and Statistics 84, 371–375.

Richardson, H., 1995. Control your costs then cut them. Transportation and Distribution , 94–96.

Rogerson, R., 1988. Indivisible labor, lotteries and equilibrium. Journal of Monetary Economics 21, 3–16.

Rotemberg, J.J., 2003. Technical progress, smooth trends, and nearly distinct business cycles. American Economic Review 93, 1543–1559.

Rupert, P., Sustek, R., 2019. On the mechanics of New-Keynesian models. Journal of Monetary Economy 102, 53–69.

Smets, F., Wouters, R., 2007. Shocks and frictions in us business cycles: A bayesian dsge approach. American Economic Review 97, 586–606.

Wang, P., Wen, Y., Xu, Z., 2014. What inventories tell us about aggregate fluctuations–A tractable approach to (S,s) policies. Journal of Economic Dynamics and Control 44, 196–217.

Wen, Y., 2005. Understanding the inventory cycle. Journal of Monetary Economics 52, 1533–1555.

Wen, Y., 2008. Inventories, liquidity, and the macroeconomy. Working Paper 2008-045A, Federal Reserve Bank of St. Louis.

A Additional information empirical section

A.1 Data sources.

- Burea of Economic Analysis (BEA)
  - Table 1.1.6: GDP and its components.
  - Tables 5.8.6A & 5.8.6B: Wholesale trade inventories, retail trade inventories, and final sales of goods and structures of domestic business.[133]

[133] At the end of 1996, there is a change in the allocation of inventories across industries. For all inventory series there are 5 quarters available (1996Q4 till 1997Q4) for which observations are available for both the old and the new definition. To obtain a consistent time series, we use the average relative magnitude for the two approaches over these five quarters to scale the pre-1996Q4 observations.

A.2 Variance Decomposition.

When $X_t = X_{t,1} + X_{t,2}$, then

$$\mathbb{V}\text{ar}[X_t] = \mathbb{V}\text{ar}[X_{t,1}] + 2\mathbb{C}\text{ov}[X_{t,1}, X_{t,2}] + \mathbb{V}\text{ar}[X_{t,2}] = \mathbb{C}\text{ov}[X_t, X_{t,1}] + \mathbb{C}\text{ov}[X_t, X_{t,2}]. \quad (46)$$

Thus, the total variance of an aggregate variable can be decomposed as the sum of the covariances between the individual components and the aggregate. This is the method we use to decompose the fluctuations in total finished-goods inventories in the three sectoral components and also for the quantitative importance of investment in inventories for GDP.

A.3 Relative volatility of inventories and sales.

Kryvtsov and Midrigan (2013) report a low (implied) elasticity of inventories relative to sales at business-cycle frequencies. In particular, they find an elasticity of inventories with respect to sales equal to 0.24 for the retail sector and 0.16 for manufacturing and trade.[134] By contrast the elasticity implied by our empirical estimates is somewhat higher and equal to 0.563. There are several differences in how KM and we calculate statistics: (i) KM use monthly data, whereas we use quarterly, (ii) KM take the log of the inventory-sales ratio, whereas we do not, and (iii) our data set uses an additional ten years of data.[135] None of these differences matter. The different outcomes are due to how the data are detrended to get the business-cycle component.

The first column in Table 6 replicates the findings of KM for the retail sector, one of the two cases considered in KM.[136] KM use monthly data and detrend the data using an HP filter with a smoothing parameter, λ_{hp}, equal to 14,400.[137] The second column in the table reports the results when the same monthly data are used, but the data are detrended using the smoothing parameter equal to 129,600 which is the value proposed in Ravn and Uhlig (2002) for monthly data.[138] The third column presents the results

[134] KM calculate the elasticity as follows. First, the elasticity of the inventory-sales ratio with respect to sales is defined as the slope coefficient in a regression of the log inventory-sales ration on log sales. Second, one is added to get the elasticity of inventories with respect to sales. That is, the elasticity is equal to $1 + \rho(x/s, s)\sigma_{x/s}/\sigma_s$.

[135] Thus, KM measure changes in this ratio as percentage changes, we measure them as percentage points changes.

[136] The same conclusions can be drawn when manufacturing and trade data are used.

[137] The motivation for this value is the following. The HP filter contains a quadratic penalty term on changes in the growth rate of the trend. The starting point is the well-established value for quarterly data which is equal to $1,600$.The reason for using 14,400 for monthly data is that the coefficient in front of the penalty term is first adjusted by a factor of three, since there are three months in a quarter, and then squared because the penalty term is quadratic. Then we get $1,600 \times 3^2 = 14,400$.

[138] This is equal to $1,600 \times 3^4$. Ravn and Uhlig (2002) show that the frequency domain representation of the filter with this adjustment applied to monthly data is approximately the same as that of the one for quarterly data. That is, both extract that part of the data associated with business-cycle frequencies. Ravn and Uhlig (2002) also discuss a motivation based on a time domain perspective. The idea is that adjustment of the smoothing parameter should be such that the ratio of the variance

for quarterly data using the standard smoothing parameter for quarterly data, $1,600$.

The following observations can be made. First, there are nontrivial quantitative differences for the results based on a smoothing parameter equal to $14,400$ and $129,600$. Specifically, the implied elasticity of inventories with respect to sales is equal to 0.262 when $\lambda_{\text{hp}} = 14,400$ and equal to 0.630 when $\lambda_{\text{hp}} = 129,600$. However, the ratio of the standard deviation of inventories to the standard deviation to sales is similar for both detrending methods. The second observation is that the results for monthly data using $\lambda_{\text{hp}} = 129,600$ are similar to those for quarterly data using $\lambda_{\text{hp}} = 1,600$.[139]

Using a different value for λ_{hp} simply means a focus on a somewhat different aspect of the series and different researchers may have different preferences. So there is no right or wrong value of λ_{hp}.[140] Also, this statistic does not play a key role in our analysis. For completeness, however, we wanted to bring to the surface why our estimate is somewhat higher than the one reported in KM.

Table 6: Comparison with Kryvtsov and Midrigan (2013)

	monthly		quarterly
	$\lambda_{\text{hp}} = 14,400$	$\lambda_{\text{hp}} = 129,600$	$\lambda_{\text{hp}} = 1,600$
$\frac{\sigma_{x/s}}{\sigma_s}$	1.189	0.888	0.810
$\rho\left(\frac{x}{s}, s\right)$	-0.621	-0.417	-0.343
implied elasticity x w.r.t. s	0.262	0.630	0.722
$\frac{\sigma_x}{\sigma_s}$	0.969	1.024	1.049

Notes. In this table, we calculate inventory statistics in exactly the same way as in KM and for the same sample period. The only difference is that KM use monthly data and $\lambda_{\text{hp}} = 14,400$ whereas we use $\lambda_{\text{hp}} = 129,600$ for the monthly series and the standard 1,600 for the quarterly series.

A.4 The customer-finding rate after demand and supply shocks

In section 2, it was shown that the correlation between customer-finding rate and aggregate activity is positive. This means that there is positive comovement when averaged across *all* shocks and leaves open the possibility that there is a negative comovement in response to some shocks. Kryvtsov and Midrigan (2013) document that the inventory-sales ratio decreases during a monetary expansion, which implies that the customer-finding rate increases. Given the dominant role that TFP shocks are

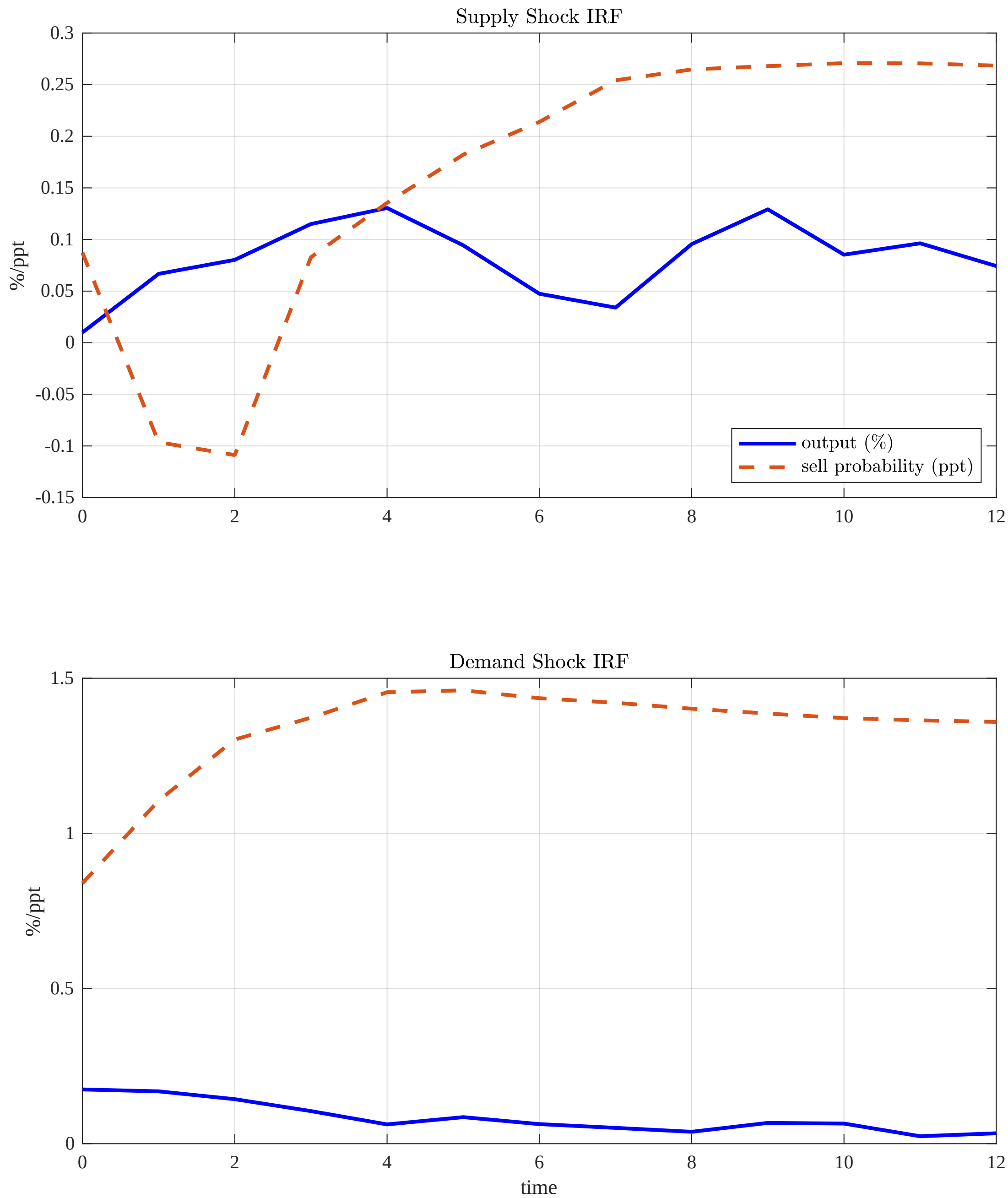
of the cyclical component and the variance of the change in the growth rate of the trend are the same if monthly instead or quarterly data is used. This motivation indicates to use $1,600 \times 3^3$ which is also higher than $14,400$. We use $\lambda_{\text{hp}} = 1,600 \times 3^4$ because we find the frequency domain motivation most appealing.

[139] This is what one could expect, since Ravn and Uhlig (2002) proposed $\lambda_{\text{hp}} = 129,600$ for monthly data so that the filter for monthly data would extract (approximately) the same frequencies as the standard HP filter for quarterly data.

[140] Of course, one should use the same value of λ_{hp} when applying the filter to observed and model-generated data.

believed to have for business-cycle fluctuations, it would be helpful to know whether TFP driven fluctuations also imply a procyclical customer-finding rate.

Figure 14: the customer-finding rate: Demand and Supply shocks



Notes. These panels plot the IRFs of the goods-sector's customer-finding rate and output in response to demand and supply shocks identified using the Blanchard-Quah decomposition.

In this section, we use the Blanchard-Quah decomposition to extract "demand" and "supply" shocks and investigate how the customer-finding rate for the goods-sector responds to these two shocks. Specifically, we use a bivariate VAR with output per hour and hours as the two variables. The Blanchard-Quah identifying assumption is that demand shocks do not have a permanent effect on productivity. This assumption is subject to critique and one obviously should take that into account when interpreting the results. In the next step, we regress the change in the goods-sector's customer-finding rate on the current and twelve lags of the either the demand or the supply shocks. Figure 14 plots the IRFs for both the level of the goods-sector's customer-finding rate and output.

The figure shows that output and the customer-finding rate are positively correlated in response to both types of shocks. Thus, it is supportive of the view that the customer-finding rate may very well be procyclical in response to both types of shocks. It is interesting to note that the response of the customer-finding rate *relative* to the response of output is much larger for the demand shock, which is also a prediction of our model.

We want to stress that one should be careful in drawing strong conclusions from this exercise given the massive challenge in credibly identifying structural shocks. However, there is another relevant observation. Using several different VAR specifications with both identified and unidentified shocks that the prominent finding is that the customer-finding rate response has the same sign as the output response. These results indicate that – consistent with the positive unconditional correlation coefficient – the customer-finding rate is procyclical for a variety of (combination of) shocks.

B Proofs for the propositions

In section 3, a subsystem of three equations was given that determine tightness, θ_t, the price of the intermediate good i, $P_{i,t}/P_t$,[141] and marginal costs, MC_t, as a function of the value of an unsold good, $\lambda^f_{x,t}$, and a measure of inflationary pressure, $\lambda^f_{d,t}$. For the convenience of the reader we repeat that system and the expressions for $\lambda^f_{x,t}$ and $\lambda^f_{d,t}$.

$$1 = \frac{\xi_e}{f^b(\theta_t)} + \frac{P_{i,t}}{P_t} \tag{47a}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = \left(\frac{P_{i,t}}{P_t} - \lambda^f_{d,t} - \lambda^f_{x,t}\right) f^f(\theta_t) \tag{47b}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = \varepsilon \lambda^f_{d,t} \frac{\nu}{1-\nu} \xi_e \theta_t \tag{47c}$$

$$\lambda^f_{x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\begin{array}{c} f^f(\theta_{t+1})\lambda^f_{s,t+1} \\ +(1-f^f(\theta_{t+1}))\lambda^f_{x,t+1}\end{array}\right)\right], \tag{47d}$$

$$1 - \varepsilon\lambda^f_{d,t} = \eta_P \frac{P_t}{P_{i,t}}\left(\begin{array}{c} \left(\frac{P_{i,t}}{P_{i,t-1}} - 1\right)\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \\ -\beta\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\frac{P_{i,t+1}}{P_{i,t}} - 1\right)\left(\frac{P_{i,t+1}}{P_{i,t}}\right)\frac{s_{t+1}}{s_t}\right]\end{array}\right), \tag{47e}$$

Recall from equation (19b that the customer-finding rate is an increasing function of tightness only. Two propositions were put forward which we repeat here with their proofs.

Proposition 1 $\frac{\partial f^f(\theta_t)}{\partial \lambda^f_{x,t}} < 0$. *That is, an increase in the value of carrying an unsold good into the future as inventory is associated with a reduction (increase) in the customer-finding rate (inventory-sales ratio).*

Proposition 2 $\frac{\partial f^f(\theta_t)}{\partial \lambda^f_{d,t}} < 0$. *That is, an increase in inflationary pressure (relative to expected future inflation) is associated with an increase (decrease) in the customer-finding rate (inventory-sales ratio).*[142]

Proofs. Using equations (47a) and (47b) we can substitute out $\frac{P_{i,t}}{P_t}$ and MC_t and rewrite equation (47c) as

$$\left(1 - \frac{\xi_e}{f^b(\theta_t)} - \lambda^f_{x,t} - \lambda^f_{d,t}\right) = \varepsilon\lambda^f_{d,t}\frac{\nu}{1-\nu}\xi_e\frac{\theta_t}{f^f(\theta_t)},$$

[141] Although $P_{i,t}$ is the same for each firm, $P_{i,t}/P_t$ is not equal to 1 in the symmetric equilibrium because of search costs.

[142] This proposition is only relevant when $\eta_P > 0$, that is, when prices are sticky, because $\lambda^f_{d,t}$ is a constant when $\eta_P = 0$.

Using that ${}^{1}\!/_{f^b(\theta_t)} = {}^{\theta_t}\!/_{f^f(\theta_t)} = \mu^{-1}\theta_t^{\nu}$ gives

$$1 = \lambda^f_{x,t} + \lambda^f_{d,t} + \xi_e \left(1 + \varepsilon\lambda^f_{d,t}\frac{\nu}{1-\nu}\right)\frac{\theta_t^{\nu}}{\mu}, \tag{48}$$

which shows that θ_t is an implicit function of $\lambda^f_{x,t}$ and $\lambda^f_{d,t}$. Rewriting and taking the partial derivative immediately gives the desired result that $\frac{\partial\theta_t}{\partial\lambda^f_{x,t}} < 0$, and $\frac{\partial\theta_t}{\partial\lambda^f_{d,t}} < 0$. Since $f^f(\theta_t)$ is an increasing function of θ_t, we also have that the customer finding rate $f^f(\theta_t)$ decreases with $\lambda^f_{x,t}$ and $\lambda^f_{d,t}$. ■

Proposition 3 *MC_t increases with $\lambda^f_{x,t}$ and decreases with $\lambda^f_{d,t}$ locally around the steady state.*

Proof. We first prove that MC_t increases with $\lambda^f_{x,t}$ locally around the steady state. After substituting out $\frac{P_{i,t}}{P_t}$ and $\lambda^f_{x,t}$, we get

$$MC_t = \left(1 - \lambda^f_{d,t}\right) + \xi_e\left(\varepsilon\lambda^f_{d,t}\frac{\nu}{1-\nu}\theta_t - \left(1+\varepsilon\lambda^f_{d,t}\frac{\nu}{1-\nu}\right)\frac{\theta_t^{\nu}}{\mu}\right). \tag{49}$$

At the steady state, $\lambda^f_d = \frac{1}{\varepsilon}$ and $f^f(\theta_{\text{ss}}) = \mu\theta_{\text{ss}}^{1-\nu} < 1$. Using this and taking the derivative of MC_t with respect to θ_t gives

$$\left.\frac{\partial MC(\theta_t)}{\partial\theta_t}\right|_{\theta_t=\theta_{\text{ss}}} = \frac{\nu}{1-\nu} - \left(1+\frac{\nu}{1-\nu}\right)\frac{\nu\theta_{\text{ss}}^{\nu-1}}{\mu} = \frac{\nu}{1-\nu}\left(1 - \frac{1}{\mu\theta_{\text{ss}}^{1-\nu}}\right) < 0. \tag{50}$$

Thus, MC_t is a decreasing function of θ_t (around the steady state). Since θ_t itself is a decreasing function of $\lambda^f_{x,t}$ according to Proposition 1, an increase in $\lambda^f_{x,t}$ would cause a decrease in θ_t, and thus an increase in MC_t. In other words, MC_t increases with $\lambda^f_{x,t}$ locally around the steady state.

Next, we prove that MC_t decreases with $\lambda^f_{d,t}$ locally around the steady state. Recall that equation (48) shows that θ_t is an implicit function of $\lambda^f_{x,t}$ and $\lambda^f_{d,t}$. Holding $\lambda^f_{x,t}$ constant and differentiating equation (48) locally around the steady state gives

$$\left.\frac{d\theta_t}{d\lambda^f_{d,t}}\right|_{\theta_t=\theta_{\text{ss}}} = \frac{\theta_{\text{ss}}\varepsilon}{\nu}\left(\frac{-\xi_e\left(1+\frac{\nu}{1-\nu}\right)}{\left(\xi_e\left(\frac{\nu}{1-\nu}\right) + \frac{\mu\theta_{\text{ss}}^{-\nu}}{\varepsilon}\right)}\right)^{-1} < 0. \tag{51}$$

In addition, holding $\lambda^f_{x,t}$ constant and differentiating equation (47c) locally around the steady state gives

$$\left.\frac{dMC_t}{d\theta_t}\right|_{\theta_t=\theta_{\text{ss}}} = \frac{\xi_e\nu}{1-\nu} + \frac{\xi_e\nu}{1-\nu}\theta_{\text{ss}}\varepsilon\left.\frac{d\lambda^f_{d,t}}{d\theta_t}\right|_{\theta_t=\theta_{\text{ss}}}$$

$$= \frac{\xi_e\nu}{1-\nu}\left(1+\frac{-\xi_e\left(1+\frac{\nu}{1-\nu}\right)}{\left(\xi_e\left(\frac{\nu}{1-\nu}\right)+\frac{\mu\theta_{\text{ss}}^{-\nu}}{\varepsilon}\right)}\nu\right), \tag{52}$$

which shows a positive relationship between MC_t and θ_t because

$$1+\frac{-\xi_e\left(1+\frac{\nu}{1-\nu}\right)}{\left(\xi_e\left(\frac{\nu}{1-\nu}\right)+\frac{\mu\theta_{\text{ss}}^{-\nu}}{\varepsilon}\right)}\nu = \frac{\frac{\mu\theta_{\text{ss}}^{-\nu}}{\varepsilon}}{\left(\xi_e\left(\frac{\nu}{1-\nu}\right)+\frac{\mu\theta_{\text{ss}}^{-\nu}}{\varepsilon}\right)} > 0. \tag{53}$$

Since θ_t and $\lambda^f_{d,t}$ are negatively related, MC_t decreases with $\lambda^f_{d,t}$ locally around the steady state. ■

C One-period model

The purpose of this appendix is to highlight the additional degree of freedom that firms have in our framework and how that affects the firm problem. We use a simple static partial-equilibrium version of our model.

Partial-equilibrium static environment. The consumer problem is given by

$$\max_{c_i,s_i,e_i} \ln(c_i)$$

s.t.

$$p_i s_i = \omega - \eta e_i, \tag{54a}$$

$$c_i = s_i, \tag{54b}$$

$$s_i = f^b_i e_i, \tag{54c}$$

where e_i stands for effort, p_i for the price, c_i for consumption, s_i for sales, and $1/f^b_i$ the amount of effort needed to obtain 1 unit of good i. Resources of the consumer are a fixed endowment, ω, but those are diminished if more effort is put into acquiring goods.[143]

[143]The alternative adopted in the main text that search costs reduce amount available for consumption leads to a more cumbersome first-order condition.

Substituting out c_i and s_i, we get

$$\max_{e_i} \ln\left(\frac{e_i}{f_i^b}\right)$$
s.t.
$$p_i f_i^b e_i = \omega - \eta e_i. \tag{55a}$$

The first-order conditions are this constraint and

$$\frac{1}{e_i} = (p_i f_i^b + \eta)\lambda. \tag{56}$$

From these two equations, we get the following demand equation:

$$s_i = \frac{\omega}{p_i + \frac{\eta}{f_i^b}}, \tag{57}$$

which is decreasing in the price and search costs, $1/f_i^b$.

The firm problem is given by

$$\max_{s_i, y_i, p_i, \theta_i} p_i s_i - \alpha y^2$$
s.t.
$$s_i = \frac{\omega}{p_i + \frac{\eta}{f^b(\theta_i)}}, \tag{58a}$$
$$s_i = f^f(\theta_i) y_i, \tag{58b}$$

where y_i denotes production, $\theta_i = {}^{e_i}/_{y_i}$ denotes tightness, and $f^f(\theta_i$ denotes the firm's customer-finding rate. For simplicity we have assumed that any unsold goods have zero value in this static example.

The firm is a monopolist and understands that its choices affect household behavior. Consequently, it takes optimal household behavior into account. Specifically, one of the firm's constraints is the household's demand equation, which indicates that demand is not only affect by the price the firm charges, but also by search cost, $1/f^b(\theta_i)$ which the firm affects by choosing tightness, $\theta_i = {}^{e_i}/_{y_i}$. What about the household constraint $s_i = f^b(\theta_i)e_i$? This constraint is automatically satisfied, since $s_i = f^f(\theta_i)y_i$ is a firm constraint and $s_i = f^b(\theta_i)e_i = f^f(\theta_i)y_i$.

From this maximization problem, we get a system of six equations in the following variables: s_i, y_i, p_i, θ_i and the two Lagrange multipliers associated with the two constraints, λ_d and λ_s. Given the functional form for $f^b(\theta_i)$ and $f^f(\theta_i)$, this is a closed system.

To solve for e_i, we just have to add the definition of tightness, $\theta_i = e_i/y_i$. The inverse of the search cost for the buyer and the customer-finding rate are given by

$$f_i^b = f(\theta_i) = \mu \left(\frac{e_i}{y_i}\right)^{-\nu}, \tag{59a}$$

$$f_i^f = f^f(\theta_i) = \mu \left(\frac{e_i}{y_i}\right)^{1-\nu}. \tag{59b}$$

One can obtain c_i from $c_i = s_i$ and by combining the household first-order condition with the one remaining constraint, equation (55a), one gets that $\lambda = 1/\omega$.

D Additional details and explanations

D.1 Details of the calibration

This appendix provides additional motivation and details for our calibration procedure with which we determine a range of appropriate values for three structural parameters that we know are important for key model properties regarding inventories, production, and sales. Those are the two curvature parameters of the functions characterizing the goods-market frictions, ν_g and ν_s, and the responsiveness of monetary policy to the output gap.

We would like the customer-finding rate in the goods market, $f_{g,t}^f$, to be not only procyclical in response to monetary-policy shocks, but also in response to TFP shocks. Section 3, however, made clear that the customer-finding rate is more responsive (relative to the change in aggregate activity) to monetary-policy shocks than to TFP shocks. In fact, $f_{g,t}^f$ is countercyclical in our model for some parameter values. Thus, the appropriate value for ν_g would depend on the type of shock considered if one would want to match a specific observed target. This difference in the procyclicality of the customer-finding rate carries over to differences in the relative volatility of sales and output. However, the estimated values of these statistics are subject to sampling variation. Thus, the relevant question is whether there is a *range* of values for these three parameters that are consistent with key observed statistics when we take into account sampling variation and that we want the model to replicate some key facts emphasized in the literature for both types of shocks.

Our calibration procedure constructs an "admissible" set of combinations of Γ_y, ν_g, ν_s such that the model is consistent with the following.

- Parameters are such that model-generated output is more volatile than sales. That is, we want ${}^{\sigma_{y_g}}/_{\sigma_{s_g}} > 1$. This is a robust empirical finding that has been challenging for the theoretical inventory literature. This is a relatively weak requirement because the estimated lower bound of the 95% confidence interval for this statistic is equal to 1.072 as reported in table 1.[144] On the other hand,

[144] Within our range of admissible parameters there are several that are such that ${}^{\sigma_{y_g}}/_{\sigma_{s_g}}$ is not only bigger than one, but also above the 1.072 lower value. In fact, table 1 documents that at this is true

it is a tough requirement because we want it to hold for *both* monetary policy and TFP shocks. We think this is important because there is so much debate regarding the empirical relevance of demand and supply shocks.[145]

- The value of ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ should also not be too high. Specifically, it should *not* be bigger than the upper bound of the 95% confidence interval (i.e., $1.124 + 1.96 \times 0.026$) in *both* the economy with only monetary-policy shocks and the economy with only TFP shocks. If the ratio would be higher for both types of shocks, then it would be impossible to find a combination of the two innovation standard deviations with which the model can generate a value for ${}^{\sigma_{yg}}/_{\sigma_{sg}}$ inside the 95% confidence interval of the empirical estimate.[146] The focus of the literature has been on the possibility that ${}^{\sigma_{yg}}/_{\sigma_{sg}} > 1$ and it has not worried about the possibility that it could be too big. But this requirement does help in narrowing the set of admissible parameters in our model.

- The observed inventory-sales ratio is countercyclical and we want this to be true in our model as well and for both types of shocks. Consistent with our theoretical framework, we focus on the customer-finding rate, which is a monotone negative transformation of the inventory-finding rate. Thus, we want the customer-finding rate to be procyclical. In models with just one shock, correlation coefficients can be quite high, even close to 1, especially when dynamics of variables are similar. In a model with only monetary policy shocks it is Indeed the case that the correlation coefficient is not only always positive, but also very high. For TFP shocks, however, the correlation can be either positive or negative and when positive not necessarily large. So the question really is whether the customer-finding rate in the model with TFP shocks is procyclical and sufficiently so. Since this is a one-side test, we check whether the model-generated correlation coefficient is bigger than 0.514 (the point estimate) minus 1.645×0.109 (the standard error).

We also considered imposing that inventories are procyclical, but it turns out that this is automatically true when the three listed requirements are satisfied.[147] All model-generated moments are calculated using HP-filtered data series, consistent with how their empirical analogues were calculated.[148] We impose that ν_g is equal to ν_s, since

for our calibrated parameter combination.

[145]Although there are clear limitations to the Blanchard-Quah decomposition, the empirical results in appendix A.4 are consistent with the view that the customer-finding rate is procyclical for both demand and supply shocks.

[146]Thus, it is fine if the relative volatility is much higher than what we see in the data for *just one* of the shocks, because that still allows the model to be within the confidence interval of the empirical unconditional moment by putting less weight on this particular shock.

[147]Intuitively, inventories will be procyclical whenever output is more volatile than sales which is guaranteed by our first requirement.

[148]Model-generated moments are calculated using a long sample of 100,000 observations. Since our procedure requires solving several nonlinear equations, we do not use the more computationally-intensive method of the average across many short replications.

the calibration procedure is quite computationally intensive. But it turns out that even with this restriction, we get several parameter values for which we can satisfy all criteria. Appendix E.5, documents that results are quite similar when nu_s is not equal to ν_g. Also, the estimation procedure allows the two parameter values to differ.

One particular inventory moment is not used in this calibration strategy and that is the correlation between the customer-finding rate and the beginning-of-period inventory stock. This moment seems to have been overlooked by the inventory literature. Of course, the data have already been quite challenging for model builders. But this comovement statistic is interesting for two reasons. First it is negative and although it is not that precisely estimated, it is significantly negative. Thus, the customer-finding rate is negatively correlated with available goods that were produced in the past and positively with newly produced goods. This seems intriguing. Also, it turns out that this moment is where our model with monetary-policy shocks differs *qualitatively* from the one with TFP shocks in that in response to monetary-policy shocks it predicts a positive comovement and with TFP shocks a negative one. As shown in section 4.7.2, this means that we can use this aspect of the inventory data to study the quantitative importance of different types of shocks for business-cycle fluctuations.

Calibrated admissible range. The range of admissible values for ν_g (and ν_s) is between 0.5007 and 0.6574 and for Γ_y between 0 and 0.0612. The joint set is displayed in figure 15. There are four boundaries. $\Gamma_y > 0$ is a natural theoretical restriction. The following explains what data properties pin down the other three boundaries.

First, consider the upper bound on ν_g. The customer-finding rate is less responsive to TFP shocks which means that the volatility of output relative to sales in the model with only TFP shocks, $[\sigma_{yg}/\sigma_{sg}]^{\text{TFP}}$ is higher than the corresponding number in the model with only monetary policy shocks, $[\sigma_{yg}/\sigma_{sg}]^{\text{MP}}$. As ν_g increases, the relative volatility of output to sales increases for both types of shocks. When $\nu_g = 0.6574$ and $\Gamma_y = 0$, we have that $[\sigma_{yg}/\sigma_{sg}]^{\text{TFP}}$ exceeds $[\sigma_{yg}/\sigma_{sg}]^{\text{MP}}$ which is exactly equal to the upper bound of its unconditional empirical counterpart. Thus, for higher values of ν_g, there is no combination of the two shock standard deviations that could match the unconditional empirical counterpart. An increase in Γ_y mainly dampens the expansion induced by a negative interest rate shock, but reduces the volatility of sales by slightly more than the volatility of output which means that the relative volatility of output to sales increases. Thus, as Γ_y increases, the value of ν_g has to drop to ensure that the relative volatility does not exceed the upper bound imposed. However, the right border is basically vertical, since the effects are quantitatively small.

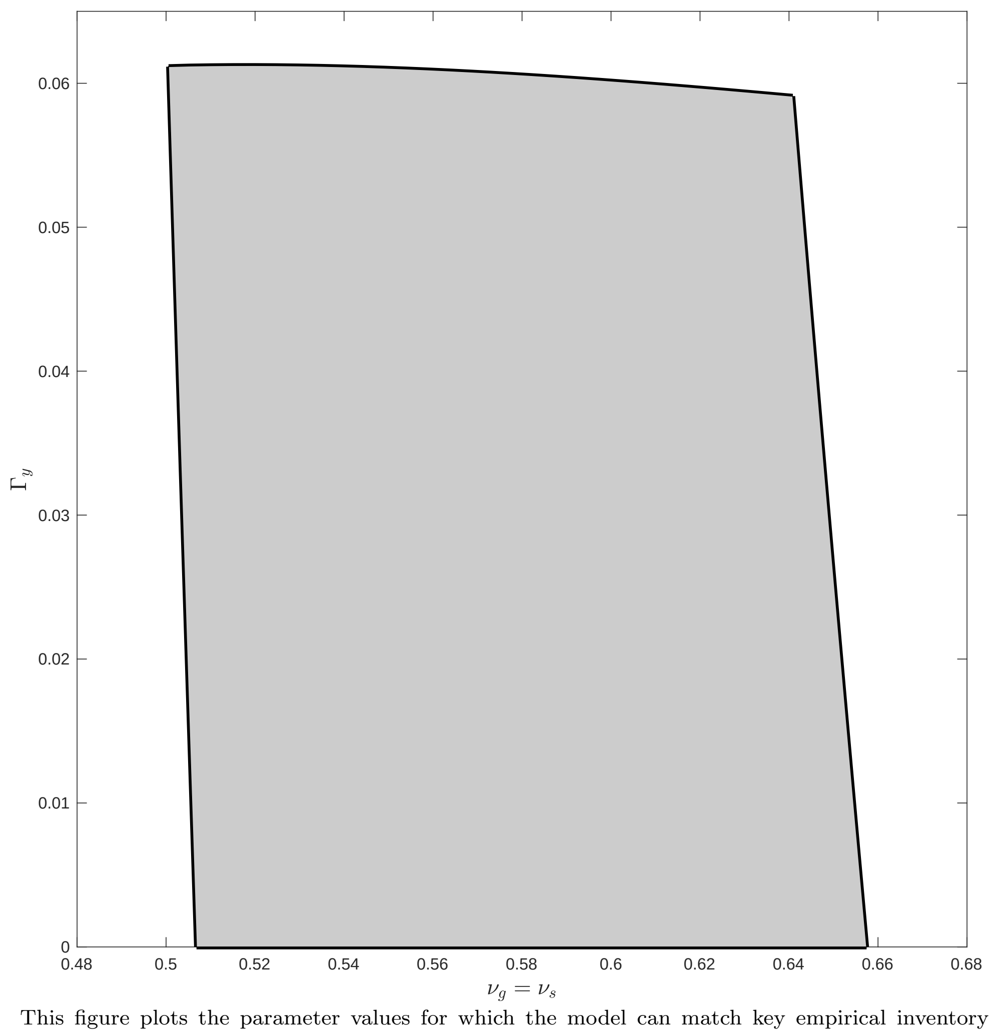
Now consider the left border. For low values of ν_g, it is less likely that output is more volatile than sales. Specifically, if $\nu_g = 0.5068$ and $\Gamma_y = 0$, then it is no longer possible for the model with just monetary-policy shocks to generate the prediction that output is more volatile than sales.[149] The reason the border on the left is downward

[149]Recall that our approach is restrictive in that we want to document that the model replicate this key inventory facts independent of one's believes on the relative importance of shocks. At lower

sloping is based on the same reasoning given above for the border on the right.

At the upper bound, the requirement that the correlation coefficient of the customer-finding rate with output is sufficiently positive becomes binding for TFP shocks.[150] As shown in section 3, the response of the customer-finding rate following a TFP shock falls and may turn negative at some point as Γ_y increases, which would imply a correlation coefficient with output that is too low relative to its empirical counterpart, taking into account sampling variation.

Figure 15: admissible area for ν_g (=ν_s) and Γ_y



Notes. This figure plots the parameter values for which the model can match key empirical inventory facts. Specifically, output is more volatile than sales for both shocks, the customer-finding rate is sufficiently procyclical for both shocks, and the model can be in the 95% confidence interval for the ratio of output to sales volatility for a combination of shocks.

values of ν_g the model can still predict that output is more volatile than sales, but one could not rely on having only monetary or similar demand-type shocks. This first part of the calibration strategy is weak in that it is qualitative in nature and only requires output to be more volatile than sales not be close to the point estimate.

[150] Since the correlation for monetary-policy shocks is always sufficiently positive.

D.2 Inventory stylized facts and model predictions for different values in calibrated range

In section 4.7.1, we discussed the model properties presented in table 1 when we used a parameter combination in the middle of the calibrated admissible range. That is, $\nu_g = \nu_s = 0.565$ and $\Gamma_y = 0.03$. In section 4.7.2, we provided an argument to prefer values for ν_g (and ν_s in the lower half of the calibrated admissible range, because then the model cannot only explain the traditional inventory facts emphasized in the inventory literature (and used to construct the admissible range), but also the correlation of the customer-finding rate and the beginning-of-period inventory stock. So a value for ν_g equal to $\nu_g = 0.565$ turns out to be a bit below the upper bound of our preferred region when we take this additional moment into consideration. To document robustness of our results, we report in table 7 the key inventory, production and sales statistics the curvature parameters take on the lowest value of the admissible area, i.e., $\nu_g = \nu_s = 0.50684848$. The table also repeats the results for the case when $\nu_g = \nu_s = 0.565$.

The results are fairly similar, but there are some quantitative differences. When $\nu_g = 0.565$, which is the benchmark considered in the paper, then the model does better in replicating the relative volatility of output and sales for *both* types of shocks. In fact, the value for ${}^{\sigma_{y_g}}/_{\sigma_{s_g}}$ is above the lower bound of the 95% confidence interval for both types of shocks. When ν_g takes on the lower value, however, then output is more volatile than sales in response to monetary policy shocks, but at a value of 1.006 only slightly so. It is, thus, clearly outside the 95% confidence band of the empirical estimate.

As documented in figure 13, at the lower values for ν_g, there is a wider gap between the minimum and maxium contribution of TFP shocks for GDP fluctuations. This figure is shown again in the next section and discussed in more detail.

Table 7: Inventory stylized facts and model predictions for different values in calibrated range

customer-finding rate statistics

	DATA	MODEL benchmark values			MODEL $\nu_g = \nu_s = 0.50683848$		
		TFP&R	TFP	R	TFP&R	TFP	R
$\mathbb{E}[f_g^f]$	0.506	=	=	=	=	=	=
	(0.002)						
$\frac{\sigma_{f_g^f}}{\sigma_{y_g}}$	0.170	0.099	0.066	0.164	0.124	0.077	0.213
	(0.157)	(0.009)	(0.005)	(0.005)	(0.011)	(0.006)	(0.005)
$\rho(f_g^f, y_g)$	0.514	0.594	0.444	0.935	0.599	0.452	0.949
	(0.109)	(0.067)	(0.023)	(0.005)	(0.069)	(0.021)	(0.005)
$\rho(f_g^f, x_{g,-1})$	-0.223	-0.018	-0.222	0.645	-0.057	-0.220	0.530
	(0.105)	(0.115)	(0.078)	(0.049)	(0.117)	(0.075)	(0.044)

inventory, sales, and production statistics

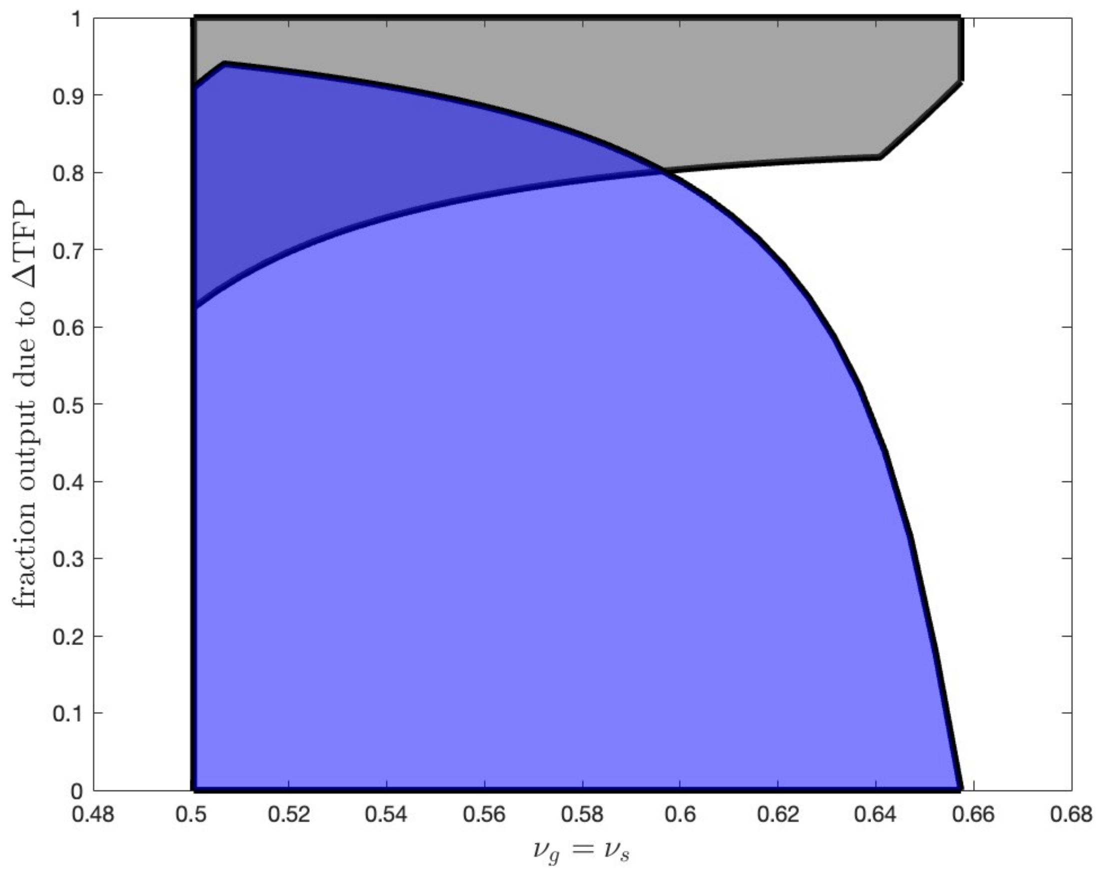
	DATA	TFP&R	TFP	R	TFP&R	TFP	R
$\mathbb{E}[\frac{x}{s_g}]$	0.976	=	=	=	=	=	=
	(0.077)						
$\frac{\sigma_{y_g}}{\sigma_{s_g}}$	1.124	1.175	1.212	1.081	1.149	1.206	1.006
	(0.026)	(0.030)	(0.034)	(0.020)	(0.029)	(0.033)	(0.014)
$\frac{\sigma_x}{\sigma_{s_g}}$	0.835	0.812	0.936	0.313	0.777	0.923	0.196
	(0.054)	(0.041)	(0.011)	(0.009)	(0.050)	(0.012)	(0.011)
$\frac{\sigma_{x/s_g}}{\sigma_{s_g}}$	0.749	0.453	0.313	0.692	0.557	0.364	0.837
	(0.045)	(0.044)	(0.033)	(0.008)	(0.054)	(0.037)	(0.011)
$\rho(\frac{x}{s_g}, s_g)$	-0.583	-0.599	-0.355	-0.991	-0.633	-0.387	-0.989
	(0.113)	(0.069)	(0.033)	(0.001)	(0.067)	(0.030)	(0.001)
$\rho(s_g, y_g)$	0.941	0.594	0.444	0.935	0.599	0.452	0.949
	(0.033)	(0.067)	(0.023)	(0.005)	(0.069)	(0.021)	(0.005)
$\rho(x, y_g)$	0.630	0.839	0.867	0.977	0.795	0.853	0.845
	(0.095)	(0.026)	(0.019)	(0.001)	(0.036)	(0.021)	(0.016)

Notes. Inventory series are based on finished goods in the manufacturing, wholesale, and retail sector. Sales are final sales in the sector producing goods and structures. The customer-finding rate, f_g^f, is calculated using equation (3). Also, x denotes inventories, s_g sales, and y_g output of the goods and structures sector. The DATA column reports standard errors in parentheses; these are calculated using the VARHAC procedure of Den Haan and Levin (1997) which corrects for serial correlation and heteroskedasticity. The columns for model-generated statistics report the means across 10,000 replications of length 212 (same length as the data set) as well as – in brackets – the standard deviation across replications. The column labeled "TFP&R" uses a mix for the two innovation standard deviations as discussed in the main text. In the other columns only one type of shock is driving fluctuations. In addition to the benchmark values, i.e. $\nu_g = \nu_s = 0.565, \Gamma_y = 0.03$, the table also reports model outcomes when a value for ν_g ($= \nu_s$) is used that is at the lower bound of the calibrated admissible area. The value of σ_R/σ_A is set equal to 0.5921 which ensures that the means of both $\rho(f_g^f, x_{g,-1})$ and $\sigma_{y_g}/\sigma_{s_g}$ across replications are inside the empirical 95% confidence intervals of their empirical counterpart. Throughout this paper, we extract business-cycle components using the HP filter with a smoothing coefficient of 1,600.

D.3 Understanding the kinks in figure 13

Section 4.7 provides an explanation for the general shape and location of the blue and grey area in figure 13 which is duplicated here as figure 16. In this appendix, we provide a more detailed explanation for the kinks and the non-monotonicity present in the borders of the two areas.

Figure 16: monetary-policy shock; benchmark calibration



Notes. The blue area indicates for a given value of ν_g which fraction of GDP can possibly be generated by TFP shocks when the model is consistent with the empirical estimate of σ_{yg}/σ_{sg} taking into account sampling uncertainty and the range of values for Γ_y that is in the admissible area for that value of ν_g. The grey area is constructed in the same way, but uses the empirical outcome for $\text{COR}(f_g^f, x_{-1})$. Since there are only two shocks in the model, 1 minus the value on the vertical axis represent the fraction of GDP that can possibly be generated by monetary-policy shocks.

The figure displays the results of an exercise that uses the restrictions imposed by two key estimated inventory-sales moments to study the relative importance of monetary and TFP disturbances for business-cycle fluctuations, taking into account sampling uncertainty. The first restriction is given by the standard deviation of output relative to the standard deviation of sales, σ_{yg}/σ_{sg}. The second moment is the correlation of the customer-finding rate with the beginning-of-period inventory stock, $\text{COR}(f_g^f, x_{-1})$.

The exercise. For the convenience of the reader, we repeat the specifics of the exercise. The key element of the exercise is to consider different values for the innovation standard deviations, σ_A and σ_R, and check for which combinations the model is consistent with the observed values of $\sigma_{y_g}/\sigma_{s_g}$ and $\text{COR}(f_g^f, x_{-1})$ taking into account sampling uncertainty. In both cases, we restrict values of ν_g, ν_s, and Γ_y to be such that they are in the calibrated admissible area. Let $[M^x]$ denote the value of a moment generated by shocks of type x, where $x \in \{\text{MP,TFP}\}$.

Results of the exercise. Figure 16 displays the results. The horizontal-axis variable is the parameter ν_g. The vertical axis displays the fraction of GDP fluctuations that is due to TFP fluctuations for that value of ν_g. It is a range because (i) we allow Γ_y to vary and take on values that are in the admissible area for that value of ν_g and (ii) "being consistent" with the estimated value of the moment takes into account sampling variation.

The first restriction. Our model can replicate the intriguing fact from the inventory literature that output is more volatile than sales and can do so for both TFP and monetary-policy shocks. However, there is a quantitative difference in that the customer-finding rate fluctuates less strongly in response to TFP shocks which means that $\sigma_{y_g}/\sigma_{s_g}$ will be higher for TFP shocks.[151] The blue area in the figure is based on values of the innovation standard deviations, σ_A and σ_R, such that the standard deviation of goods production relative to the standard deviation of the sale of goods, $\sigma_{y_g}/\sigma_{s_g}$ is not too large and not too small. "Not too large" means that the unconditional moment predicted by the full model with both shocks should not be above the upper bound of the 95% confidence interval.[152] "Not too small" means the unconditional moment should not be less than 1. Alternatively, we could restrict it to be not smaller than the lower bound of the 95% confidence interval. This would actually strengthen our conclusion regarding the relative importance of TFP shocks.[153]

[151] As explained in section 3, if the customer-finding rate is constant, then output will definitely be more volatile than sales in our framework with a goods-market friction. If the customer-finding rate fluctuates sufficiently strongly over the business cycle, however, then sales could be more volatile.

[152] Using the numbers presented in table 1, this means it cannot exceed 1.175. Also, for this exercise, we only have to consider variations in the ratio of standard deviations, σ_A/σ_R, since our model solution is based on a first-order approximation.

[153] The 95% confidence interval is equal to $[1.073, 1.175]$ so the moment $\sigma_{y_g}/\sigma_{s_g}$ is estimated quite precisely. If we would restrict the model generated value for $\sigma_{y_g}/\sigma_{s_g}$ to be above 1.073 in this exercise *and* also in the construction of the admissible area, then this would not affect the qualitative shape of the figure. However, both areas would shrink with $\nu_g = 0.556$ as the lower bound instead of $\nu_g = 0.5$ and the peak of the top border of the blue area would occur at $\nu_g = 0.567$. This would reduce the maximum possible role for monetary-policy shocks and increase the minimum role for TFP shocks. If we would restrict the model-generated value for $\sigma_{y_g}/\sigma_{s_g}$ to be above 1.073 but only in this exercise and not in the construction of the admissible area (which would thus remain the same), then the lower bound of the blue area would no longer be zero for values of ν_g below 0.556. Thus, we are giving monetary-policy shocks again the best possible chance. Although we are lenient by using 1 as the

The following is important for understanding the lower bound of the blue area. First, $[\sigma_{yg}/\sigma_{sg}]^{\text{TFP}}$ exceeds the upper bound of the confidence area for *all* values of ν_g and Γ_y in the admissible area. By contrast, $[\sigma_{yg}/\sigma_{sg}]^{\text{MP}}$ is always below the upper bound.[154] Thus, to be consistent with the estimated value of σ_{yg}/σ_{sg} – and taking into account sampling variation – it must be the case that at least some of the fluctuations are due to monetary-policy shocks. Second, the admissible area is constructed such that $\sigma_{yg}/\sigma_{sg} > 1$, both when fluctuations are due only to TFP shocks and when they are due only to monetary shocks. Thus, we know that the model with only monetary-policy shocks will generate a value for σ_{yg}/σ_{sg} that is neither too low nor too high. This explains that the lower bound of the blue area is the zero line, that is, the first restriction is consistent with a model in which *all* fluctuations are driven by monetary-policy shocks and none by TFP shocks.

Now that we have established the lower bound of the blue area, we turn to the upper bound, that is, the maximum fraction of GDP fluctuations that could be driven by TFP shocks in a model that is consistent with the restriction that σ_{yg}/σ_{sg} is not too low and not too large. To understand the upper bound, the following two properties are important. Keeping Γ_y constant, a decrease in ν_g increases the volatility of the customer-finding rate for both shocks which leads to a decrease in σ_{yg}/σ_{sg}. Similarly, a lower value for Γ_y means less dampening, which again implies a more volatile customer-finding rate and a lower value for σ_{yg}/σ_{sg}.

Let's start at the highest possible value for ν_g in the admissible area, that is, the bottom-right corner of the admissible are for which ν_g is equal to 0.6574 and $\Gamma_y = 0$. At this point, the model generates a value for $[\sigma_{yg}/\sigma_{sg}]^{\text{MP}}$ that is equal to the upper bound of the empirical confidence interval and a value for $[\sigma_{yg}/\sigma_{sg}]^{\text{TFP}}$ that is above this value. Thus, only monetary-policy shocks are possible when $\nu_g = 0.6574$, since a model with both types of shocks would generate a value for σ_{yg}/σ_{sg} that would exceed the upper bound of the empirical confidence interval. The same is true when the $\{\nu_g, \Gamma_y\}$ combination is on the right border of the admissible area, that is, $\nu_g < 0.6574$ and $\Gamma_y > 0$. The reason is the following: Along that border, the economy with only monetary-policy shocks hits the constraint that σ_{yg}/σ_{sg} cannot be outside the 95% confidence interval. Thus, to create space for TFP shocks, the value of $[\sigma_{yg}/\sigma_{sg}]^{\text{MP}}$ has to fall. As pointed out above, when we lower ν_g, then σ_{yg}/σ_{sg} falls. And it would fall by more if we keep Γ_y equal to its lowest possible value, since a reduction in Γ_y also lowers this ratio.

Thus, the upper bound of the blue area, indicating the maximum role of TFP shocks for GDP fluctuations is pinned down as follows. Starting on the right at the

lower bound for σ_{yg}/σ_{sg} instead of 1.073, we are strict in that we require that *both* the model with only TFP shocks and the model with only monetary-policy shocks have to satisfy this restriction. Given the importance of this restriction in the literature, we wanted it to hold no matter what kind of shocks are driving fluctuations.

[154]Recall from the discussion in section 4.7, that the customer-finding rate is more responsive following a monetary policy than a TFP shock which would increase the volatility of sales and decrease the gap between the two.

highest possible value for ν_g, it is first pinned down by the bottom-left corner of the admissible area. As ν_g falls the upper bound of the blue area is determined by the bottom border of the admissible area, that is, where $\Gamma_y = 0$. As ν_g reaches the bottom-left corner border, the upper bound of the blue area reaches a peak. As ν_g continues falling, the upper bound is determined by the left border of the admissible area. The explanation is that monetary policy has the best possible chance of being important when Γ_y is as low as possible. Lowering ν_g along the bottom border of the admissible area (so keeping Γ_y at zero) increases the space for TFP shocks because – as pointed out above – the value of ${}^{\sigma_{yg}}/_{\sigma_{s_g}}$ drops for TFP shocks and also because it drops for monetary-policy shocks. The upper bound of the blue area has a non-monotonicity and reaches a peak which corresponds to the parameter combination at the bottom-left corner of the admissible area. The reason is the following. Along the left border of the admissible area, $[{}^{\sigma_{yg}}/_{\sigma_{s_g}}]^{\text{MP}}$ has hit its lowest allowable value. So lowering ν_g no longer creates additional space for TFP shocks when Γ_y is kept fixed. The restriction of the admissible area that ${}^{\sigma_{yg}}/_{\sigma_{s_g}} = 1$ remains satisfied if the value of Γ_y increases as ν_g falls. But this leads to a sharp *increase* in the value of ${}^{\sigma_{yg}}/_{\sigma_{s_g}}$ for TFP driven fluctuations. Consequently, the maximum fraction of GDP fluctuations that could be due to TFP shocks has to decline.

The second restriction. The second moment we consider is the correlation of the customer-finding rate with the beginning-of-period inventory stock, $\text{COR}(f_g^f, x_{-1})$. Its point estimate is negative and significantly different from zero, although not by much. This is a useful moment to identify the role of different types of shocks, since our model robustly predicts a negative correlation for TFP shocks and a positive one for monetary-policy shocks. The latter means that a model with *only* monetary-policy shocks cannot be consistent with this negative estimate, again taking sampling uncertainty taking into consideration. But since the upper bound of the 95% confidence interval is quite close to zero, the role of monetary-policy shocks could in principle still be substantial.

The results are displayed in the grey area of figure. The model with only TFP shocks generates values for $\text{COR}(f_g^f, x_{-1})$ that are inside the estimated confidence region for all values of ν_g and Γ_y inside the admissible area. Consequently a model with only TFP shocks is consistent with this empirical finding which means that the top border of the grey area is the 100% line.

To understand the other borders, start at the (ν_g, Γ_y) combination that is at the bottom-right corner of the admissible area. At this parameter combination, the value for $[\text{COR}(f_g^f, x_{-1})]^{\text{TFP}}$ is strictly inside the confidence interval. The latter means that there is space for monetary-policy shocks, but having no monetary-policy shocks is also consistent with the estimated confidence interval. This means that the right border of the grey area is vertical. Specifically, at the highest value of ν_g in the admissible area the fraction of GDP fluctuations generated by monetary-policy shocks can be as low as 0% and is at most 13%.

The lower bound of the grey area is pinned down by (ν_g, Γ_y) combinations that are located on the right and top border of the admissible area. Starting at the highest possible value of ν_g of the admissible area and then lowering it increases the space for monetary-policy shocks, first sharply as we move along the right border of the admissible area and then more gradually as we move along the top border. The explanation is that as ν_g falls and we move along the right and then the top border of the admissible area, the value of Γ_y increases. And an increase in Γ_y leads to a (substantially) more negative value of $[\text{COR}(f_g^f, x_{-1})]^{\text{TFP}}$ and a (somewhat) smaller positive value of $[\text{COR}(f_g^f, x_{-1})]^{\text{MP}}$. Both imply that there is more space for monetary shocks to be consistent with the restrictions imposed by this empirical estimate of $\text{COR}(f_g^f, x_{-1})$.[155] Since the value for Γ_y increases sharply along the right border of the admissible area and gradually along its top border, we see a corresponding change in the slope of the bottom borders of the grey area.[156]

D.4 Full-information estimation of key parameters

In the main text, we consider both a calibration and an estimation procedure to pin down values for a set of key parameters that are important for the behavior of inventories, production, and sales. Those are the curvature parameters of the customer-finding rate relationship with tightness, ν_g and ν_s, the responsiveness of monetary policy to the output gap, Γ_y, and the shock innovations, σ_R and σ_A. In this appendix, we provide details of the estimation procedure.

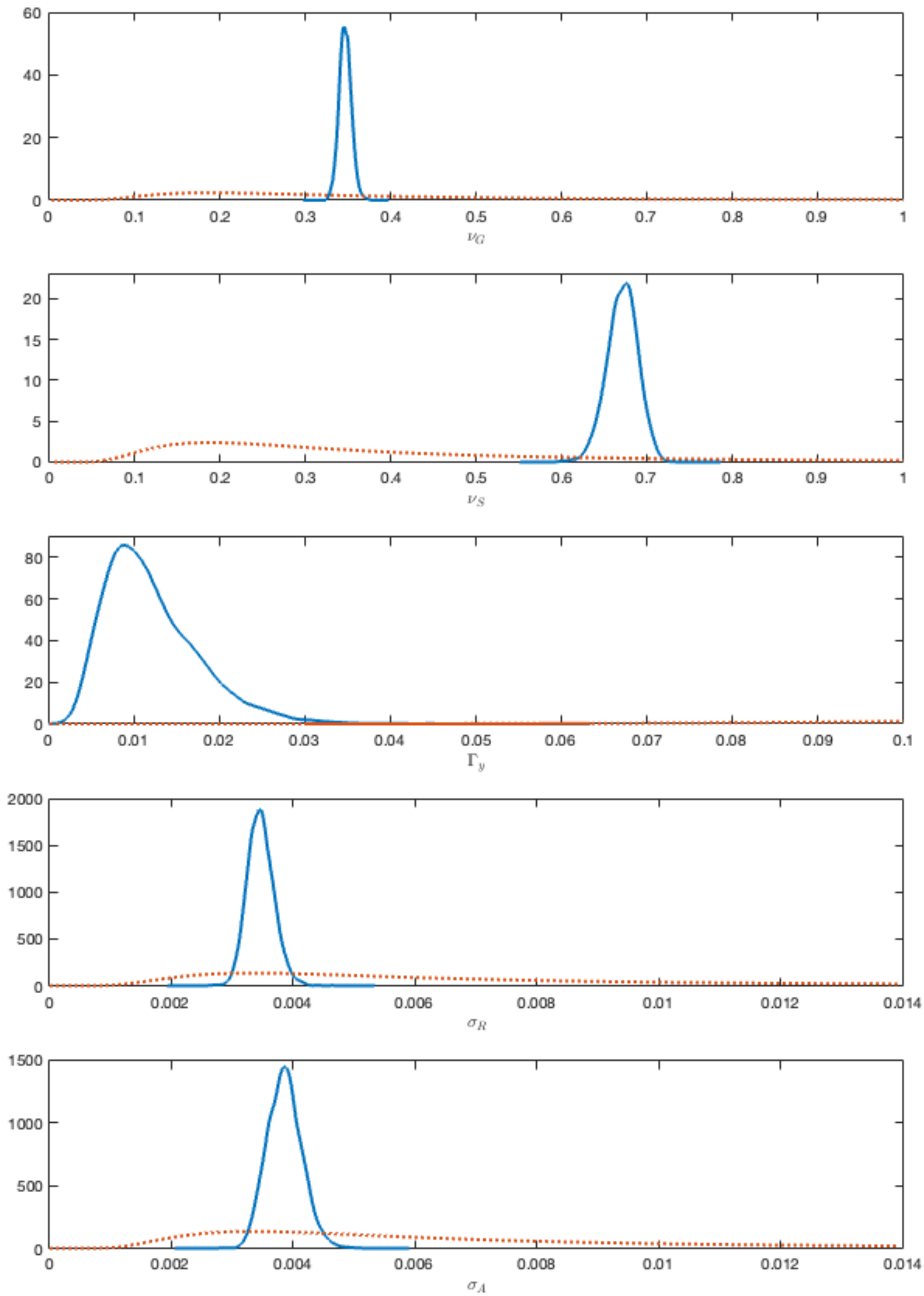
Data. The data used consists of the growth rates of inventories and final sales for the sector producing goods and structures where inventories include finished-goods inventories in the manufacturing, wholesale, and retail sector. The driving processes in our model have a unit root, but no drift. To be consistent with our model, we demean these two growth rates.

Estimation procedure. We use Dynare to implement a full-information Bayesian estimation procedure. Figure 17 plots the prior and posterior densities and table 8 provides summary information regarding the prior and the posterior. The prior is an Inverse Gamma with infinity variance so quite diffuse. The means of the prior are the based on the calibration procedure and in particular the ones used to generate the model properties in table 1. The posterior is obtained using five MCMC sequences with 10,000 observations each. Both the table and the figure document that the posterior is much more concentrated than the quite diffuse prior. That is, these two data series are quite informative about these five parameter values.

[155] A higher value of Γ_y means more dampening, but this has a stronger impact on the procyclicality of the customer-finding rate than the inventory stock, especially for TFP-driven fluctuations.

[156] That is, the kink in the bottom border corresponds to the top-right corner of the admissible area.

Figure 17: prior and posterior densities



Notes. The two panels display the prior and the posterior for the five parameters. Information about the prior is given in table 8. The posterior is obtained using five MCMC sequences with 10,000 observations each.

Table 8: Prior and posterior density summary

parameter	prior	prior mean (variance)	posterior mean	posterior 90% HPD
Γ_y	inverse Gamma	0.03 (∞)	0.0120	[0.0040,0.0198]
ν_g	inverse Gamma	0.565 (∞)	0.3469	[0.3348,0.3580]
ν_s	inverse Gamma	0.565 (∞)	0.6713	[0.6432,0.7018]
σ_R	inverse Gamma	0.01 (∞)	0.0035	[0.0031,0.0038]
σ_A	inverse Gamma	0.01 (∞)	0.0039	[0.0034,0.0043]

Notes. This table reports key information regarding the prior and the posterior. The 90%-HPD range gives the shortest interval that contains 90% of the probability density. The posterior is obtained using five MCMC sequences with 10,000 observations each.

The disadvantage of a full-information estimation method is that it is a bit of a black box and not clear what aspects of the data series used are important for the estimation outcomes. Specifically, the presence of outliers and misspecification can introduce bias.[157] Table 9 reports implied model outcomes using posterior mean parameter values for the two series used in the estimation, i.e., the growth rates of the inventory stock and sales. Whereas the implied standard deviation for sales growth rates is somewhat higher than the empirical counterpart, the implied standard deviation for the growth rates of the inventory stock is only 62% of the corresponding data series. Implications of this underestimation is discussed in section 4.7.

Table 9: Standard deviations of data used and implied values

data series	data	implied by model
growth rate of sales	0.0127	0.0116
growth rate of inventory stock	0.0097	0.0600

Notes. This table reports the standard deviations of the data series used in the estimation as well as the model-implied value when we set parameter values equal to their posterior means as reported in tqble 8.

E Additional results for the full model

E.1 Alternative TFP processes

E.1.1 Conventional stationary TFP

The business-cycle literature typically adopts a persistent, but stationary process for TFP. As explained in the main text, we adopt a more realistic non-stationary one with a serial correlation in the growth rate that matches its empirical counterpart. As shown by Christiano and Eichenbaum (1990), a model with the computationally

[157]See Den Haan and Drechsel (2021) for a discussion.

more convenient persistent stationary process and one with a non-stationary alternative have very similar predictions for the business-cycle characteristics of real variables like output, that is, after the data are filtered to exclude low-frequency moments including any trend. But Bansal and Yaron (2004) show that low-frequency movements can matter a lot for asset prices even if their volatility is small.

In our framework, the value of an unsold good, $\lambda^f_{x,t}$, is an asset price and we have shown that its countercyclical movement is key in generating a procyclical customer-finding rate in response to TFP shocks.[158] Our benchmark process for TFP ensures a robust procyclical response in consumption growth, which implies a countercyclical response in the marginal rate of substitution, which in turn implies a countercyclical $\lambda^f_{x,t}$, and a procyclical customer-finding rate. We will show in this section that this is also possible if TFP follows a stationary process, but it is then no longer a robust outcome.

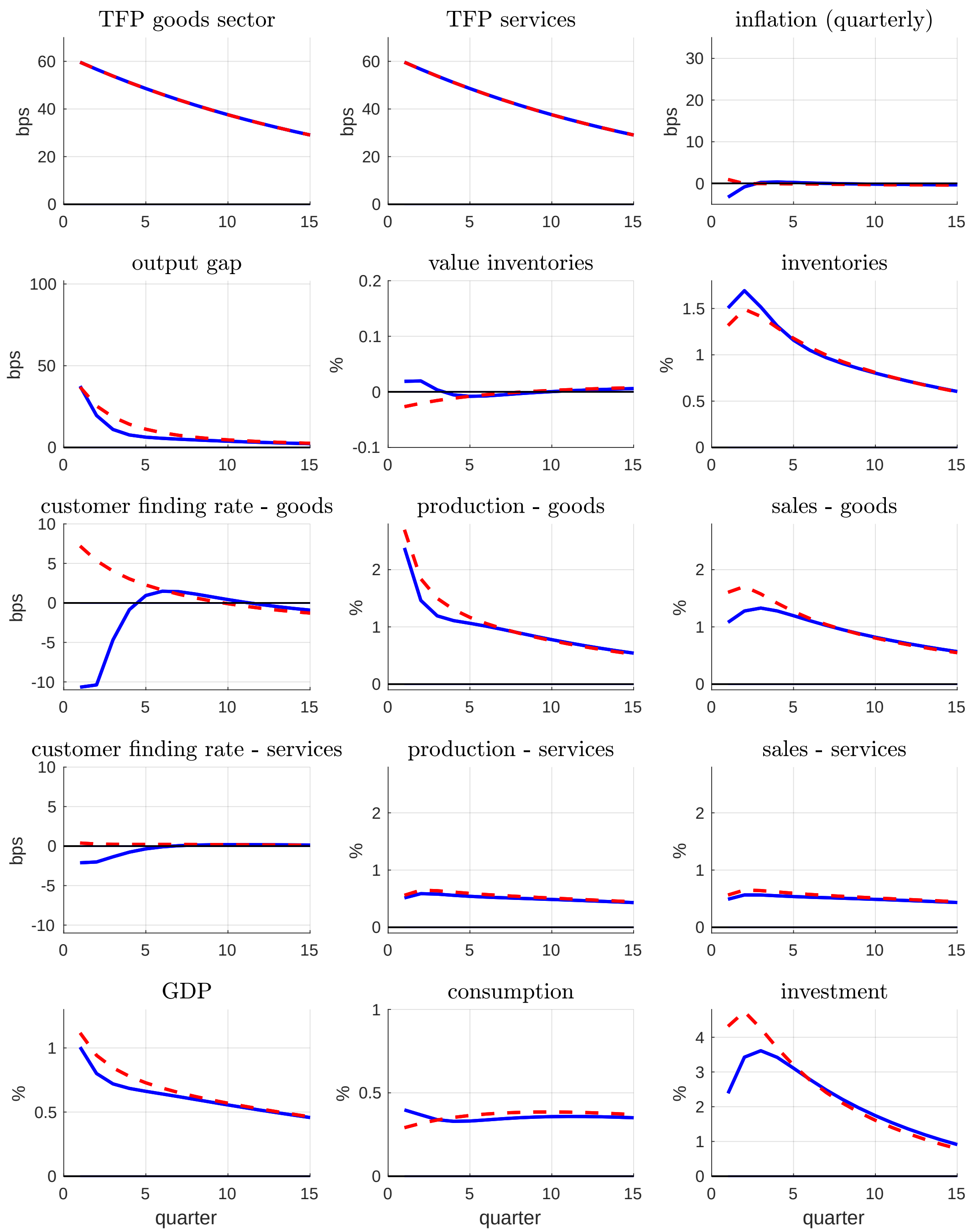
Figure 18 plots the IRFs for two cases. TFP follows a stationary process in both cases with the usual auto-regressive coefficient equal to 0.95. The blue solid line corresponds to the case where all other parameters are identical to the ones used to generate the IRFs in figure 12. We see that the value of an inventory good increases on impact and the customer-finding rate drops. To generate a procyclical $\lambda^f_{x,t}$ it is not needed that consumption keeps on increasing after the shock as in our benchmark model. If the consumption IRF displays a hump-shaped pattern, then $\lambda^f_{x,t}$ will be procyclical when it matters, that is, during the first couple periods. In fact, the literature is keen to generate such a hump-shaped pattern, because it resembles empirical estimates.[159]

There are many ways in which one can enrich the model to generate such a hump. One example is to add habits as in Fuhrer (2000). Without any such modification, our framework allows for a hump if we set the investment cost parameter, η_i, equal to zero. The corresponding IRFs are also displayed in figure 18. The figure shows that consumption now does display a (long-lasting) hump and that the customer-finding rate is once again procyclical. Instead of exploring modifications that generate a hump-shaped pattern for the consumption IRF with a stationary TFP process, we prefer to rely on the more realistic non-stationary TFP process.

[158] It is important to realize that an inventory good differs from assets such as equity in that it doesn't have procyclical dividends. Given that an expansion goes together with expected growth, agents would like to borrow which reduces the value of assets like inventory goods, whereas the value of equity with increased expected earnings are likely to increase.

[159] Ramey (2016) documents that estimated consumption IRFs do display such a hump-shaped pattern for several empirical specifications.

Figure 18: stationary TFP process; with (-) and without (- -) investment adjustment costs



Notes. TFP follows a stationary process with an auto-regressive coefficient equal to 0.95. Other parameter values are set equal to the calibrated values used to generate figure 12 except that in the case without investment adjustment costs, we set $\eta_i = 0$.

E.1.2 Productivity differences across sectors

In section 4.5, it was shown that an increase in the productivity level of the goods sector *relative* to the service sector – and, thus, a relative change in the opposite direction for marginal costs – creates an upward effect on the customer-finding rate of the service sector that could possibly overturn the (small) downward effect in the benchmark economy with flexible prices.

To ensure balanced growth, we have to assume that the long-run effect of a shock to service-sector productivity, $A_{s,t}$, is the same as that to goods-sector productivity. Thus, to study the impact of a relative change in $A_{g,t}/A_{s,t}$, we consider the case in which $A_{s,t}$ lags $A_{g,t}$ but eventually catches up. This will ensure that $A_{s,t}$ is below $A_{g,t}$ when the "action" happens, that is, in the first couple periods. Specifically, we assume that the law of motion for $A_{g,t}$ and $A_{s,t}$ are determined by the following system.

$$\ln\left(\frac{A_{g,t}}{A_{g,t-1}}\right) = \rho_A \ln\left(\frac{A_{g,t}}{A_{g,t-1}}\right) + \varepsilon_{A,t} \tag{60a}$$

$$\ln\left(\frac{A_{s,t}}{A_{g,t}}\right) = \rho_{\text{gap}}\left(\frac{A_{s,t-1}}{A_{g,t-1}}\right) - \rho_{\text{gap}}\ln\left(\frac{A_{g,t}}{A_{g,t-1}}\right) \tag{60b}$$

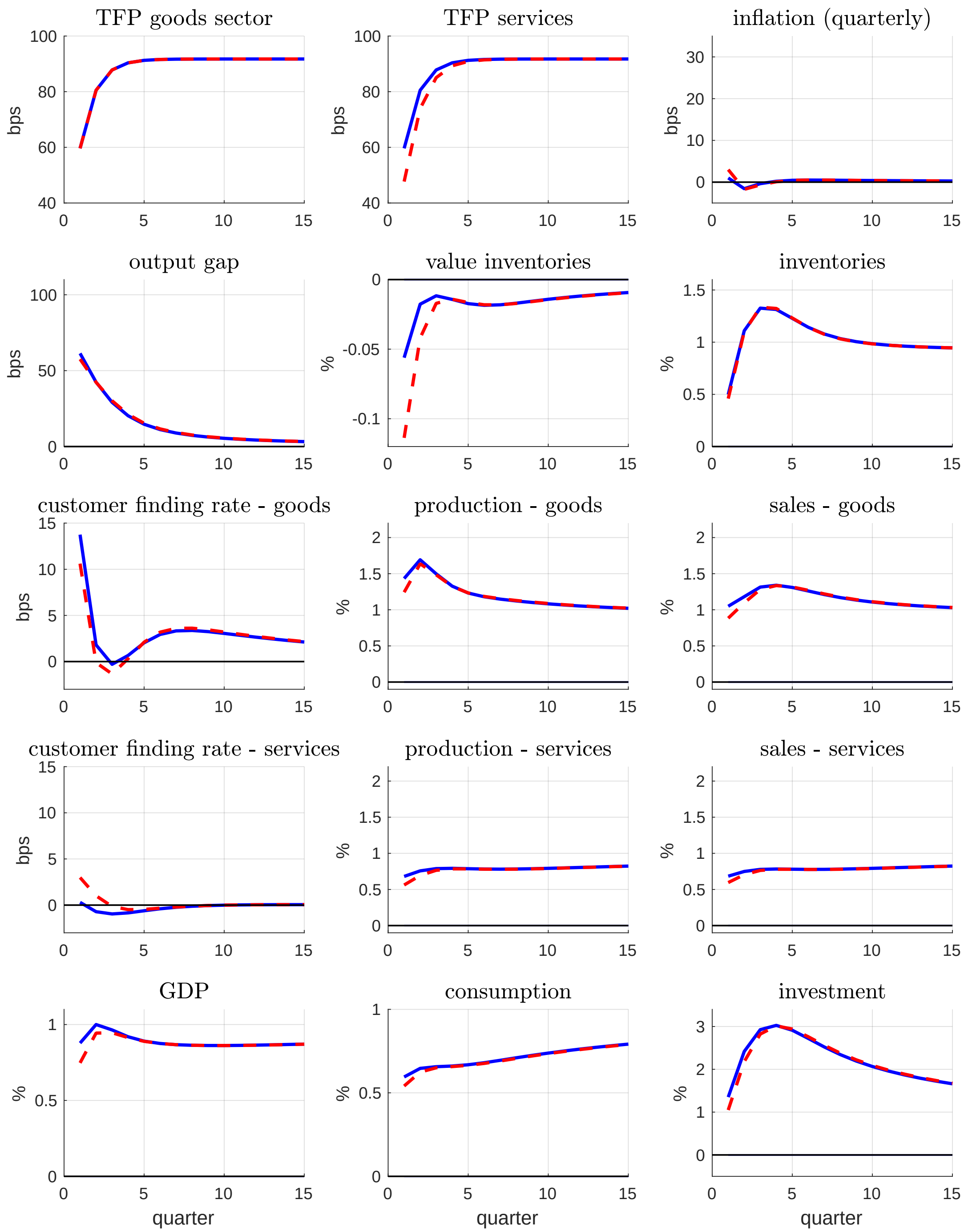
Thus, the law of motion for $A_{g,t}$ is unchanged. Following a shock, the change in $A_{s,t}$ is always less than the change in $A_{g,t}$ with the difference being the biggest on impact, but gradually going to zero.

Figure 19 displays the benchmark IRFs and the corresponding ones when ρ_{gap} is equal to 0.2 instead of 0. At the positive value for ρ_{gap}, the response of the service-sector customer-finding rate is equal to 2.99 instead of 0.30 basis points. Consistent with the analysis in section 4.5, the increase in the customer-finding rate in the goods sector is now smaller and equal to 10.61 instead of 13.80 basis points on impact. This pattern continues as we increase ρ_{gap}. In fact, when ρ_{gap} is increased enough, then the customer-finding rate in the goods sector, $f^f_{g,t}$, can display a countercyclical response. Specifically, when ρ_{gap} is equal to 0.72, then the response of $f^f_{g,t}$ is negative in the first 7 quarters after which there is only a very small positive one. But even then the customer-finding rate in the service sector increases by only 5.68 basis points.[160]

We want to point out that we show these exercises to learn more about the model and not to match an empirical counterpart, because unfortunately we don't know how the customer-finding rate of the service sector responds to TFP shocks given that the short sample discussed in section 2 only covers two recessions and demand factors are believed to have been important in both.

[160] A stronger response of $f^f_{s,t}$ is obtained when $A_{s,t} = A_{g,t-1}$, that is, the law of motion for $A_{s,t}$ is identical to the one of $A_{g,t}$ but with a one-period lag. Then the increase on impact is equal to 15.57 basis points. But again this comes at the cost of generating a negative response for the customer-finding rate in the goods sector.

Figure 19: TFP shock; $A_{s,t}$ lags $A_{g,t}$



Notes. Productivity in the service sector, $A_{s,t}$, lags productivity in the service sector as in equation (60) with $\rho_{\text{gap}} = 0.2$. Other parameter values are set equal to the calibrated values used to generate figure 12.

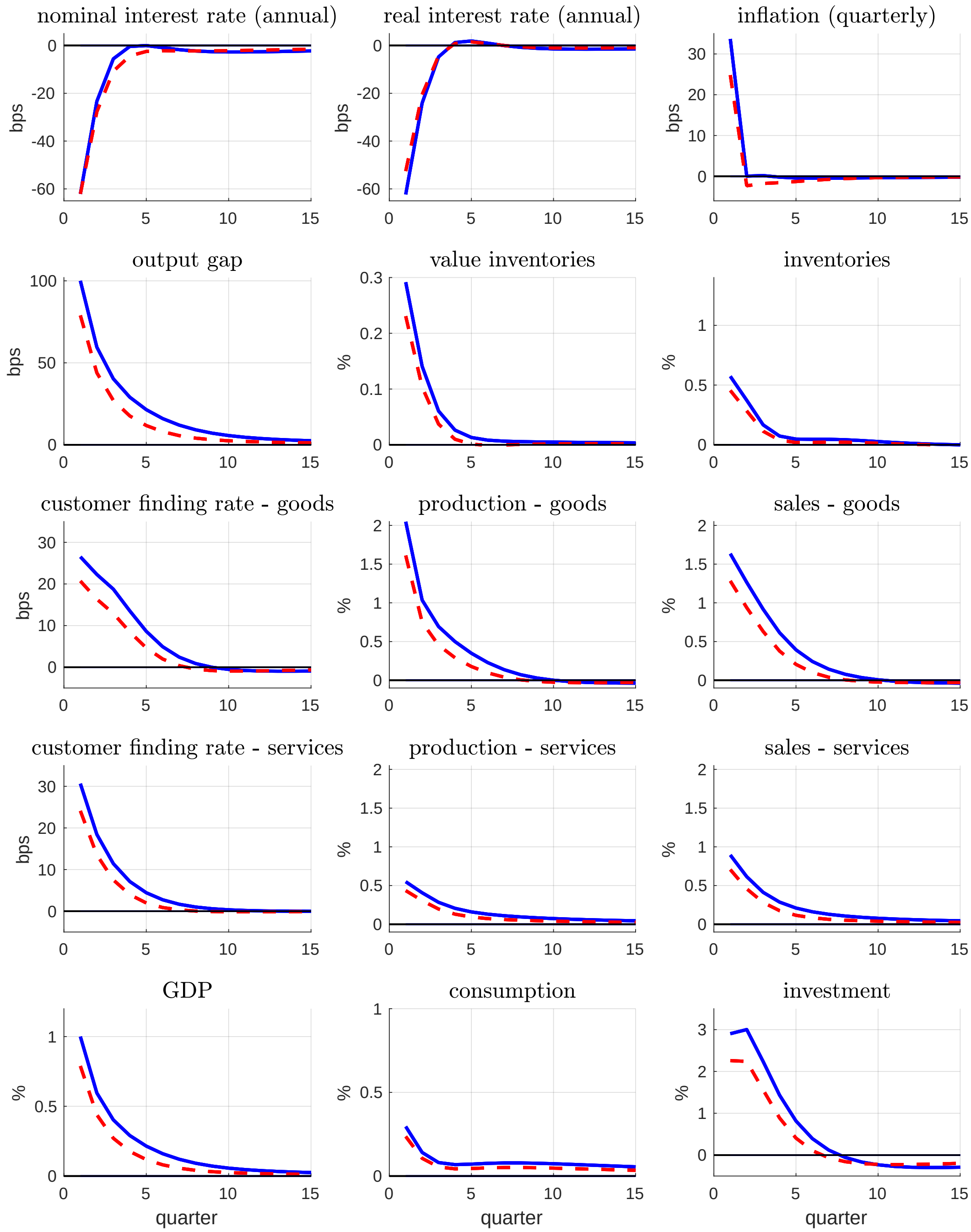
E.2 Alternative Taylor rule

Figure 20 shows the IRFs when we use a Taylor rule as estimated by Mazelis et al. (2023) together with our benchmark results. The associated coefficients are $\Gamma_\pi = 1.99$, $\Gamma_y = 0.24$, and $\Gamma_{\text{lag}} = 0.84$. So all coefficients are substantially larger than the ones used in our benchmark calibration. We have scaled σ_R up to ensure that the drop in the nominal interest rate is the same as in the benchmark. As expected, with a more hawkish Taylor rule, the responses of a monetary-policy shock are dampened across the board. For example, the customer-finding rate in the goods sector increases with 26.5 basis points and GDP increases with 1.00% for our benchmark calibrated parameter set. With the alternative Taylor rule, the customer-finding rate increases with 20.7 basis points and GDP increases with 0.79%. This corresponds to 26.2 basis points per per percentage point increase in GDP. So the *relative* responses are very similar.

Figure 21 displays the corresponding IRFs for a TFP shock. For our benchmark Taylor rule, the central bank avoids both inflationary and deflationary pressure following this supply-side shock. For our more hawkish Taylor rule, a TFP shock is accompanied with some deflationary pressure. According to proposition 2 in section 3.4, this should have a downward effect on the customer-finding rate. Indeed, this is what we find for both sectors. Specifically, the customer-finding rate in the goods sector still displays a sizable initial response, but is quickly followed by a (smaller) negative response after which there is a minor positive but persistent response.[161] The response of the customer-finding rate in the service sector is now uniformly negative. And although larger than the response found with our benchmark parameter values, still much smaller in absolute value than what is observed with a monetary-policy shock or with a TFP shock for the response of the customer-finding rate in the goods sector.

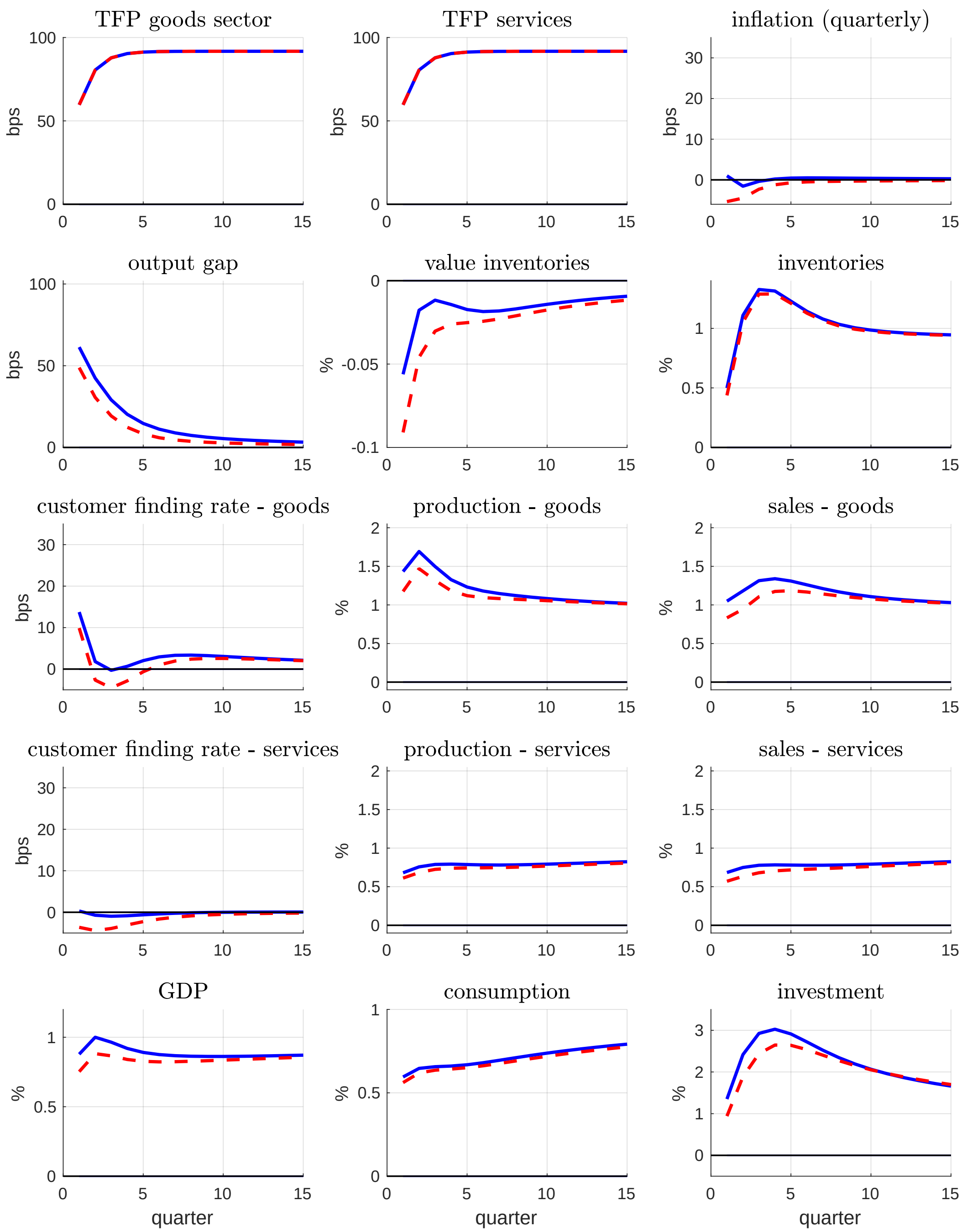
[161] And the correlation coefficient of the cyclical component of the customer-finding rate and production in the goods sector would now be slightly negative, namely -0.04, in the economy with only TFP shocks. Combined with monetary-policy shocks, however, the correlation coefficient would be positive and substantial, namely +0.58.

Figure 20: monetary-policy shock; alternative Taylor Rule



Notes. The dashed lines display IRFs of a monetary-policy shock using an alternative Taylor rule with $\Gamma_y = 0.24$, $\Gamma_\pi = 1.99$, and $\Gamma_{\text{lag}} = 0.84$. Other parameter values are set equal to the calibrated values used to generate figure 11. The solid lines display the IRFs using our benchmark parameters.

Figure 21: TFP shock; alternative Taylor Rule



Notes. The dashed lines display IRFs of a TFP shock using an alternative Taylor rule with $\Gamma_y = 0.24$, $\Gamma_\pi = 1.99$, and $\Gamma_{\text{lag}} = 0.84$. Other parameter values are set equal to the calibrated values used to generate figure 12. The solid lines display the IRFs using our benchmark parameters.

E.3 Alternative assumptions about search costs

The results in the main text are based on the assumption that search costs consists of a mix of goods and services and the calibrated value of the fraction of services, Υ_s, was set equal to the fraction of services in the consumption bundle, $\omega_{s,c} = 0.5771$.[162] This may underestimate the services component as getting advice, information acquisition, and transportation are important aspects of acquiring consumption and investment goods and services. To study robustness of our results, we consider an extreme case in which all search costs are in the form of services, that is, $\Upsilon_s = 1$.

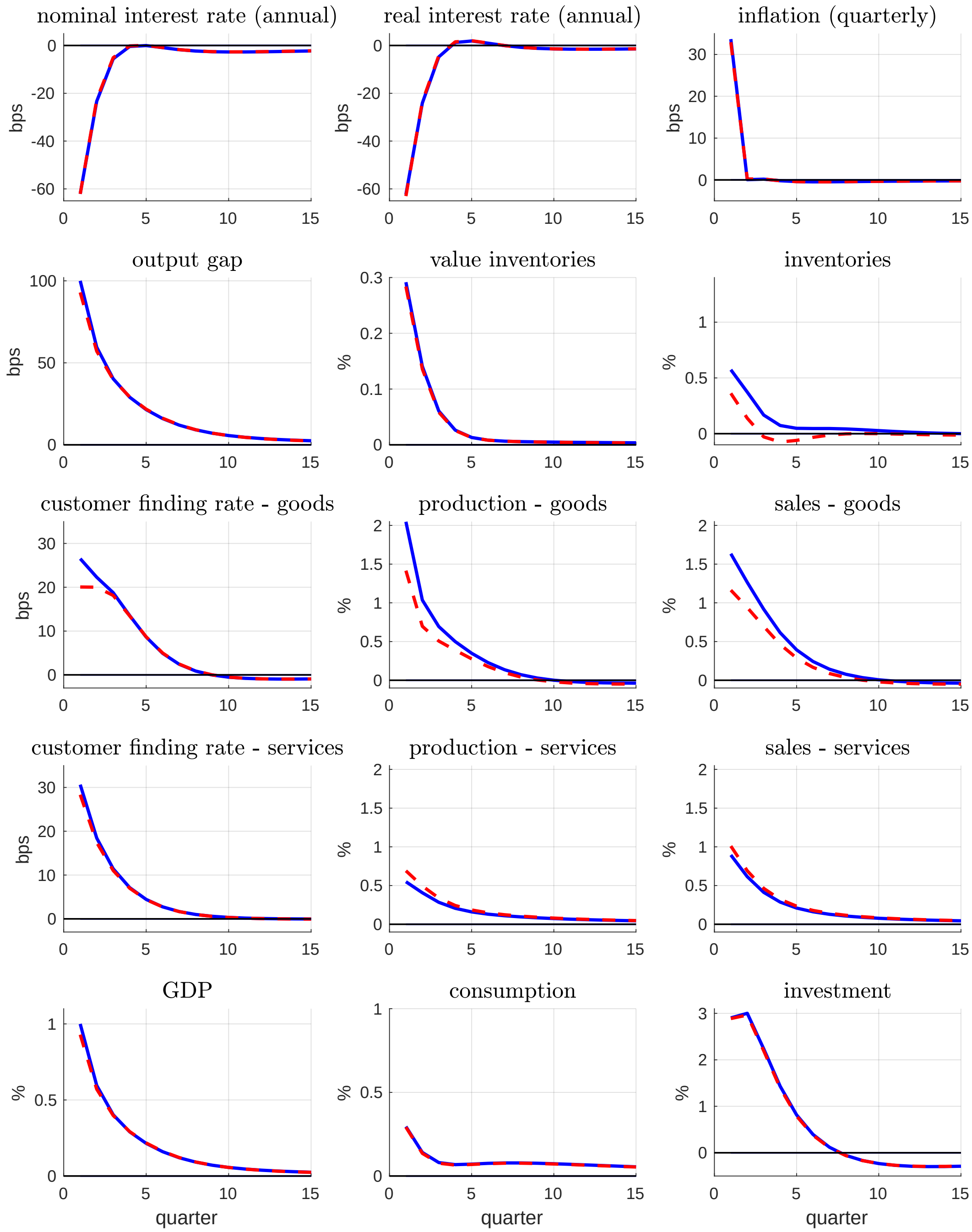
Figures 22 and 23 display the associated IRFs for this alternative parameterization as well as the ones for our benchmark calibration as displayed in figures 11 and 12. The IRFs for a monetary-policy shock show that the goods sector benefits less from an expansion when they are not needed for the acquisition of consumption and investment expenditures. This goes together with a slightly smaller response of the customer-finding rate in the goods sector,namely 20.1 instead of 26.5 basis points. This implies that the relative volatility of sales to output declines, which in turn implies that inventories go up by less and the initial positive response is followed by a very small negative one.[163]

This parameter change has virtually no effect on the TFP IRFs. An important difference between a monetary policy and a TFP shock is that the long-run quantitative impact is completely pinned down by the long-run increase in TFP which is not affected by the change in Υ_s, i.e., by the relative importance of goods in search costs. That is, the long-run response of all real aggregates remains the same. That is not true for the customer-finding rate which does return to its pre-shock value. But the fixed permanent long-run responses for variables like inventories, production, and sales do shape the temporary response for variables like the customer-finding rate. This explains why the TFP IRFs are not affected much by changes in the value for Υ_s.

[162] And the fraction spend on goods, Υ_g is equal to $1 - \Upsilon_s$.

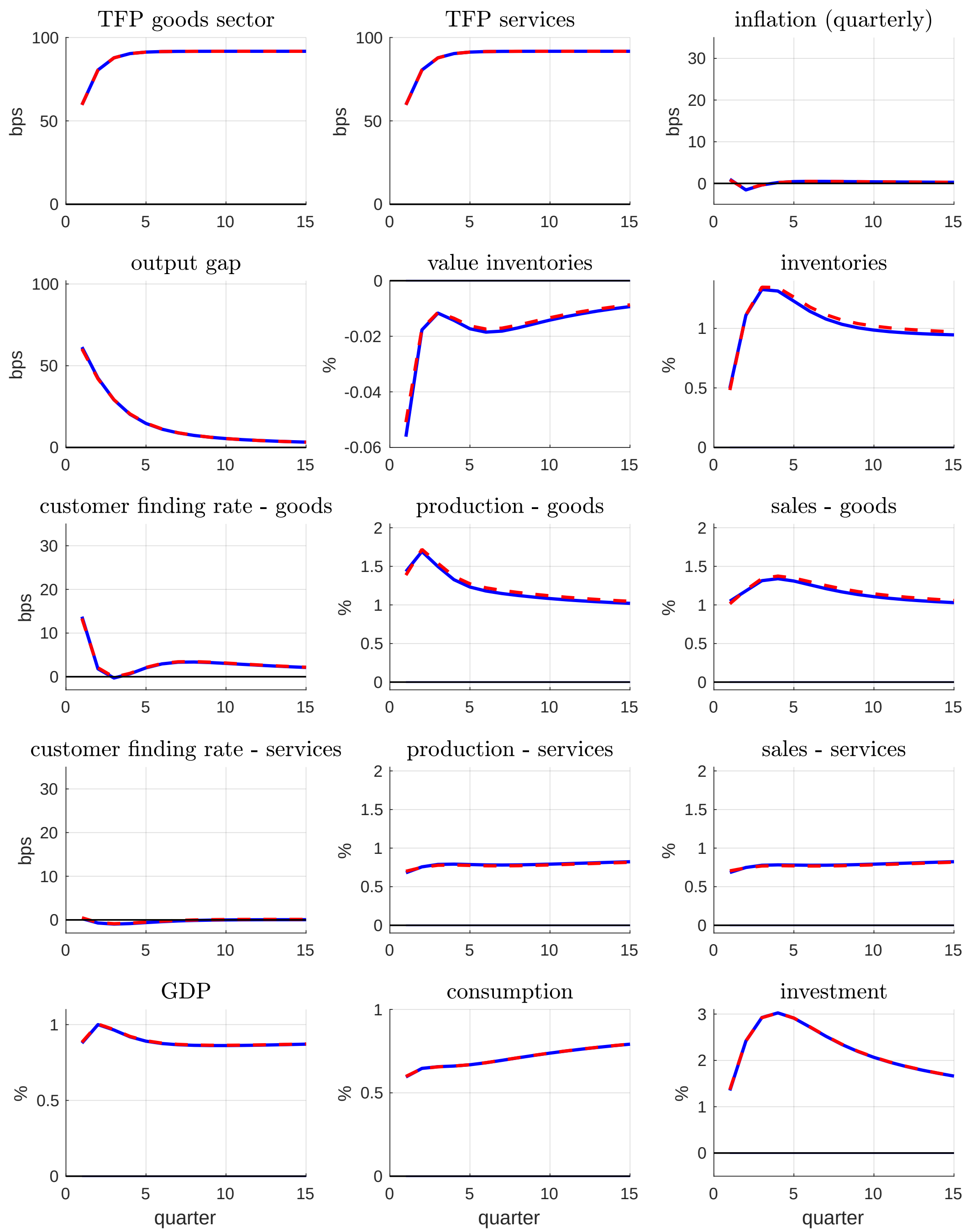
[163] Note that we have not redone the calibration. If we would lower the curvature parameter ν_g, then the customer-finding rate would be more volatile and the response of inventories would be stronger.

Figure 22: monetary-policy shock; search only requires services - $\Upsilon_s = 1$



Notes. The dashed lines display IRFs of a monetary-policy shock when search costs are only in the form of services, that is, $\Upsilon_s = 1$. The value of σ_R is set such that the response of the nominal interest rate on impact is the same as the one for our benchmark parameterization. Other parameter values are set equal to the calibrated values used to generate figure 11. The solid lines display the IRFs using our benchmark parameters.

Figure 23: TFP shock; Search only requires services - $\Upsilon_s = 1$



Notes. The dashed lines display IRFs of a TFP shock when search costs are only in the form of services, that is, $\Upsilon_s = 1$. Other parameter values are set equal to the calibrated values used to generate figure 12. The solid lines display the IRFs using our benchmark parameters.

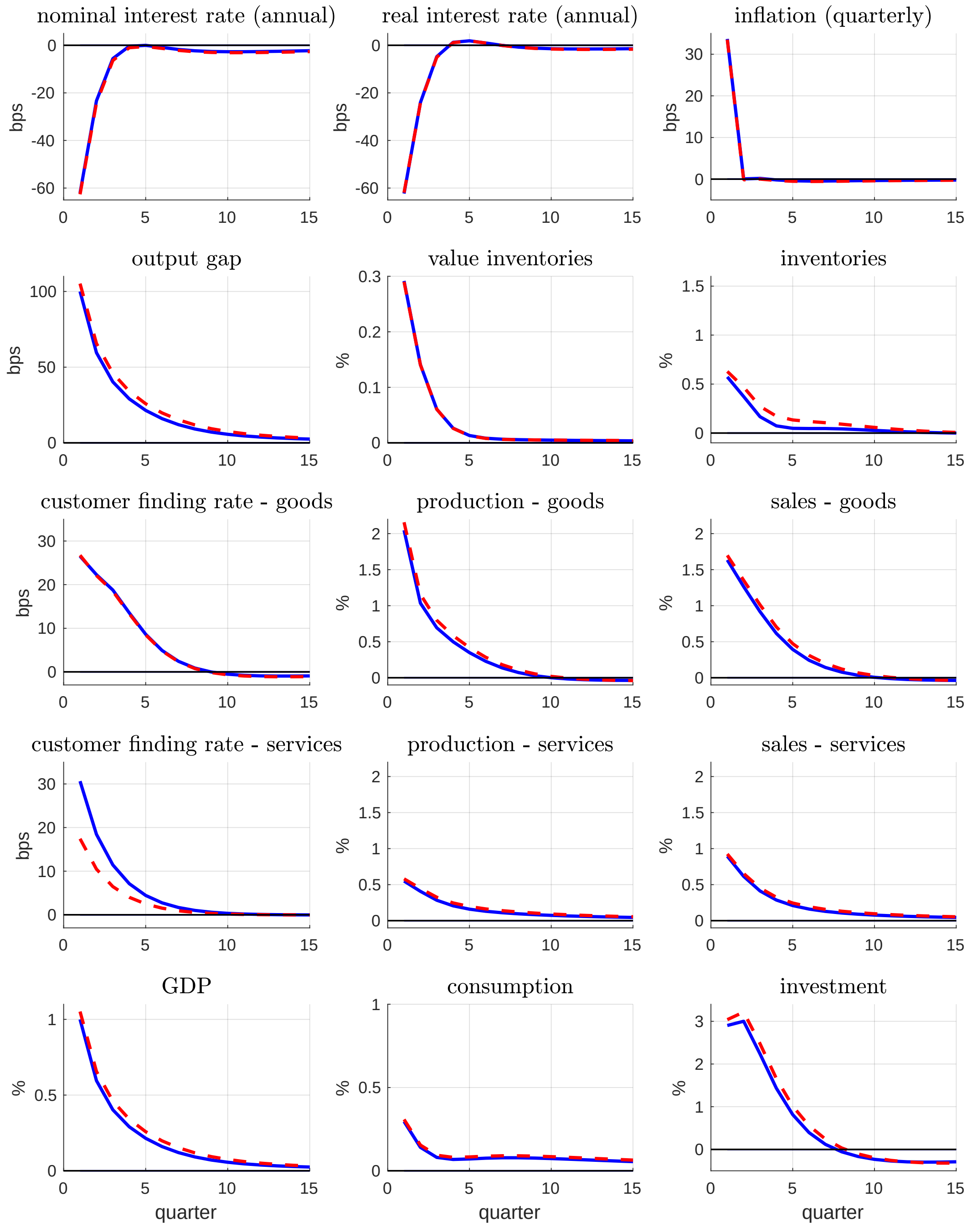
E.4 Alternative mean service-sector customer-finding rate

The results in the main text are based on parameters such that the mean customer-finding rate of services is equal to 0.888 which is substantially higher than the one in the goods sector which is equal to 0.506. Whereas the latter is based on a long sample for the inventory-sales dratio (and equation (19b), the former is based on a short sample of capacity survey data. It seems natural that the customer-finding rate (or sell fraction) is substantially higher in the service sector since an unsold service cannot be carried over to the next period, whereas that is possible for goods that don't fully depreciate. Nevertheless, it is useful to check whether our results depend on this assumption. As an alternative, we consider the case where the average customer-finding rate is the same in both sectors.

Figure 24 shows that the IRFs for a monetary-policy shock are barely affected except for the response of the customer-finding rate in the service sector. We have always expressed these responses as a change in basis points. But if we would have expressed them as percentage changes, then the IRFs are actually virtually identical.

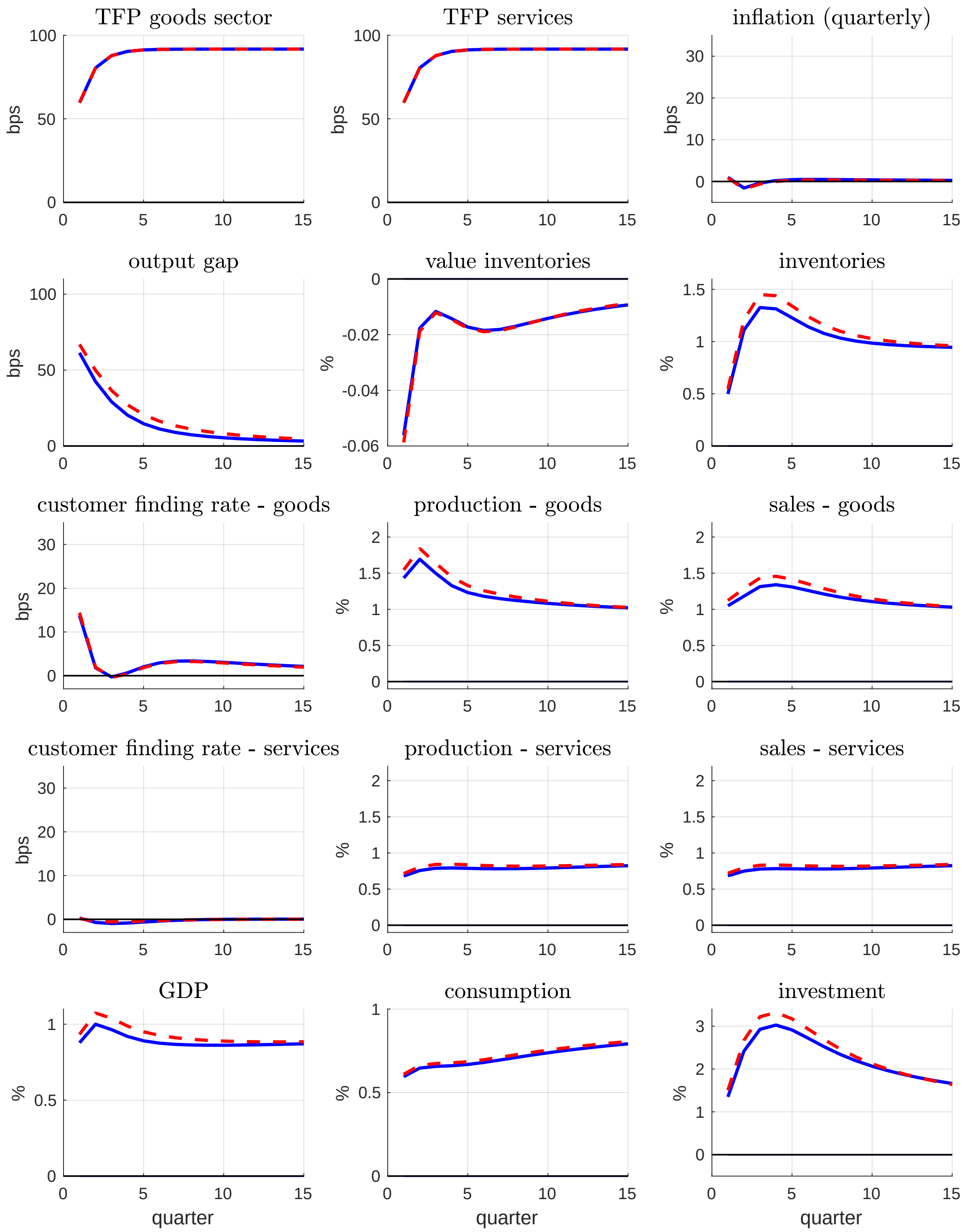
Although the effect is quantitatively small, the economy as a whole has become slightly more volatile. The reduction in the average sell fraction in the service sector has made the economy less efficient and the same employment level generates less value added. Figure 25 shows that this change in the mean customer-finding rate for the service sector has also made the economy more volatile in response to TFP shocks. Although the results are more visible in the figure than for monetary-policy shocks, the quantitative impact of this change in parameter values is still very minor.

Figure 24: monetary-policy shock; same average customer-finding rate across sectors



Notes. The dashed lines display IRFs of a monetary-policy shock when the mean customer-finding rate is the same in the service sector is equal to 0.506 instead of 0.888 so equal to the one in the goods sector. The value of σ_R is chosen to ensure the same drop in the nominal interest rate on impact. Other parameter values are set equal to the calibrated values used to generate figure 11. The solid lines display the IRFs using our benchmark parameters.

Figure 25: TFP shock; same average customer-finding rate across sectors



Notes. The dashed lines display IRFs of a TFP shock when the mean customer-finding rate is the same in the service sector is equal to 0.506 instead of 0.888 so equal to the one in the goods sector. The value of σ_R is chosen to ensure the same drop in the nominal interest rate on impact. The solid lines display the IRFs using our benchmark parameters.

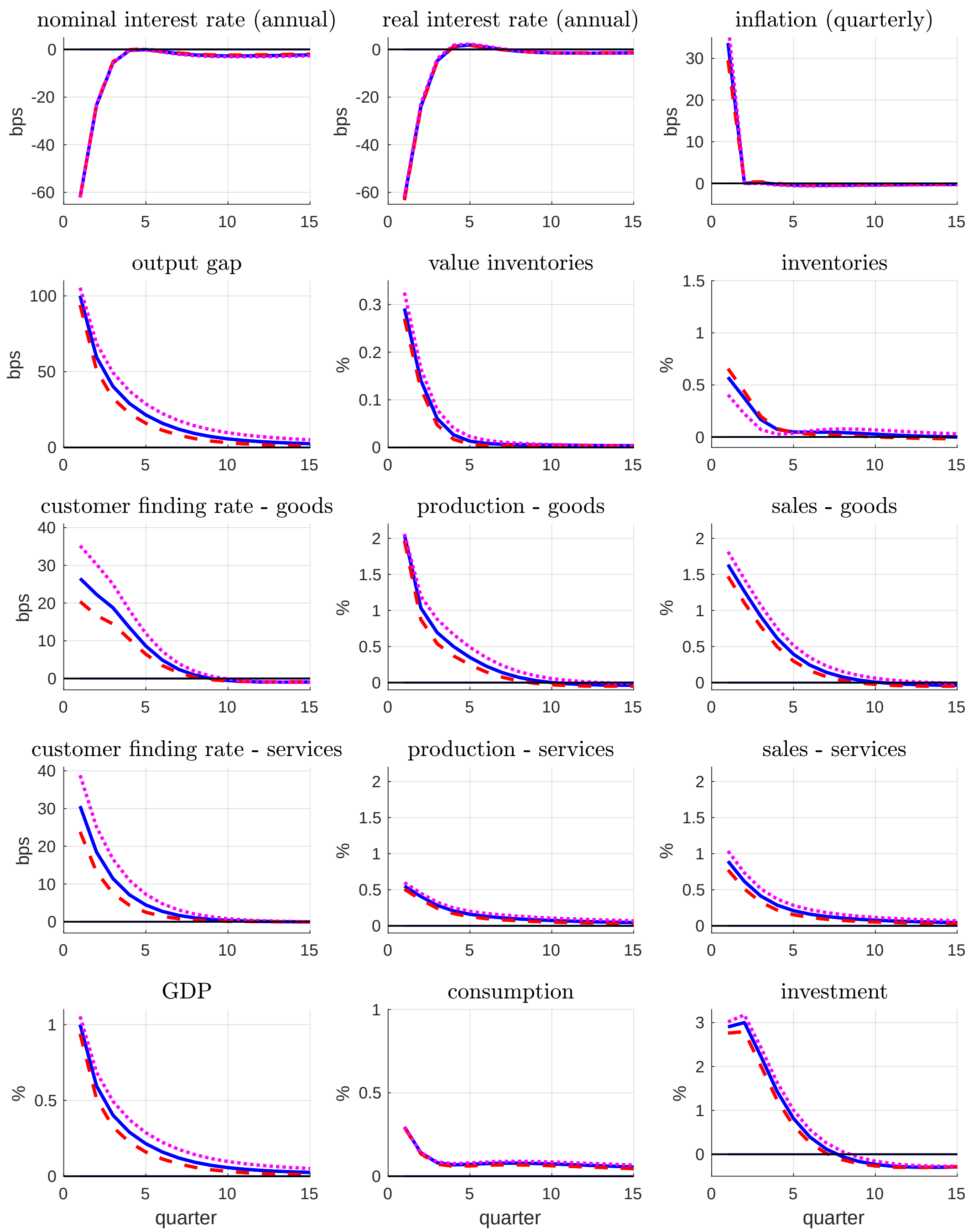
E.5 Alternative curvature in friction of selling services

Our calibration procedure was very careful in making sure that the range of values we considered for the curvature parameter that controls variations in the friction of selling goods, ν_g, was consistent with key inventory, production, and sales data. Results in the main text are based on the assumption that the analogue parameter for the service sector, ν_s, takes on the same value. Figures 26 and 27 show the IRFs for the benchmark and two alternatives, namely $\nu_s = \nu_g - 0.1$ (dotted line) and $\nu_s = \nu_g + 0.1$ (dashed line). The results for the customer-finding rate are as expected. That is, a lower value of ν_s implies that the customer-finding rate is a more nonlinear function of tightness which translates into a more volatile customer-finding rate. For a monetary-policy shock, the more volatile customer-finding rate results in slightly more volatile sales.

Since we keep the scale of the vertical axis the same for the two types of shocks, the impact of a change in ν_s on the goods-sector customer-finding rate responses to a TFP shock is less visible.[164] But there is an impact. When $\nu_s = \nu_g - 0.1$, then the response on impact is equal to 15.0 basis points, whereas it is only 12.8 basis points when $\nu_s = \nu_g + 0.1$. In contrast to what was observed for a monetary-policy shock, the higher response of the customer-finding rate allows firms to increase production by less.

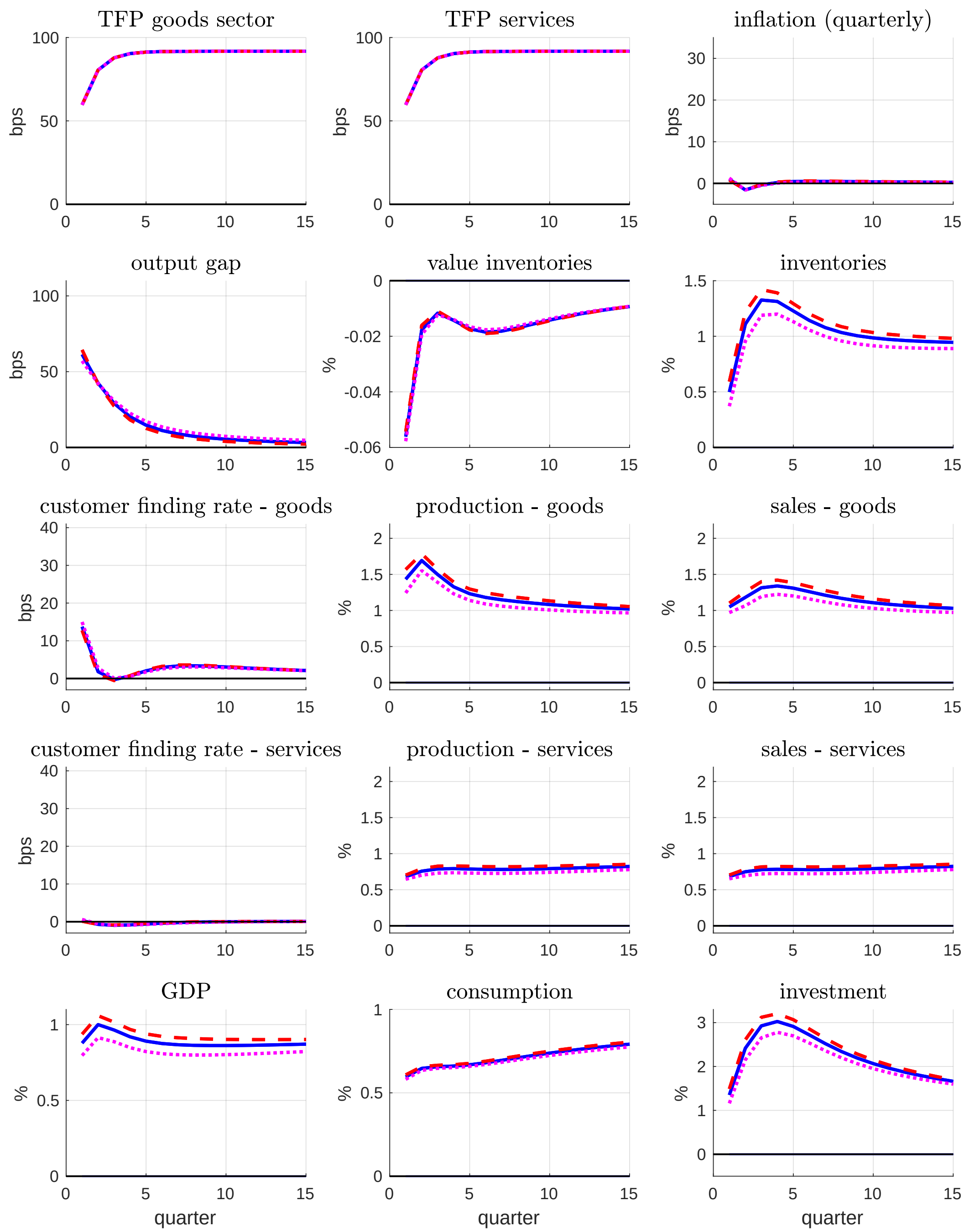
[164]The customer-finding rate response in the service sector remains tiny.

Figure 26: MP shock; higher (- -), lower (:), and benchmark curvature service sector friction



Notes. This figure displays the monetary policy IRFs for three different values of ν_s, the curvature parameter in the function that controls the friction of selling services. The solid line corresponds to the benchmark case, $\nu_s = \nu_g$, the dashed line to $\nu_s = \nu_g + 0.1$, and the dotted line to $\nu_s = \nu_g - 0.1$. The value of σ_R is chosen to ensure the same drop in the nominal interest rate on impact. The value of σ_R is chosen to ensure the same drop in the nominal interest rate on impact.

Figure 27: TFP shock; higher (- -), lower (:), and benchmark curvature service sector friction



Notes. This figure displays the TFP IRFs for three different values of ν_s, the curvature parameter in the function that controls the friction of selling services. The solid line corresponds to the benchmark case, $\nu_s = \nu_g$, the dashed line to $\nu_s = \nu_g + 0.1$, and the dotted line to $\nu_s = \nu_g - 0.1$. The value of σ_R is chosen to ensure the same drop in the nominal interest rate on impact.

E.6 Alternative assumptions about inventory maintenance costs

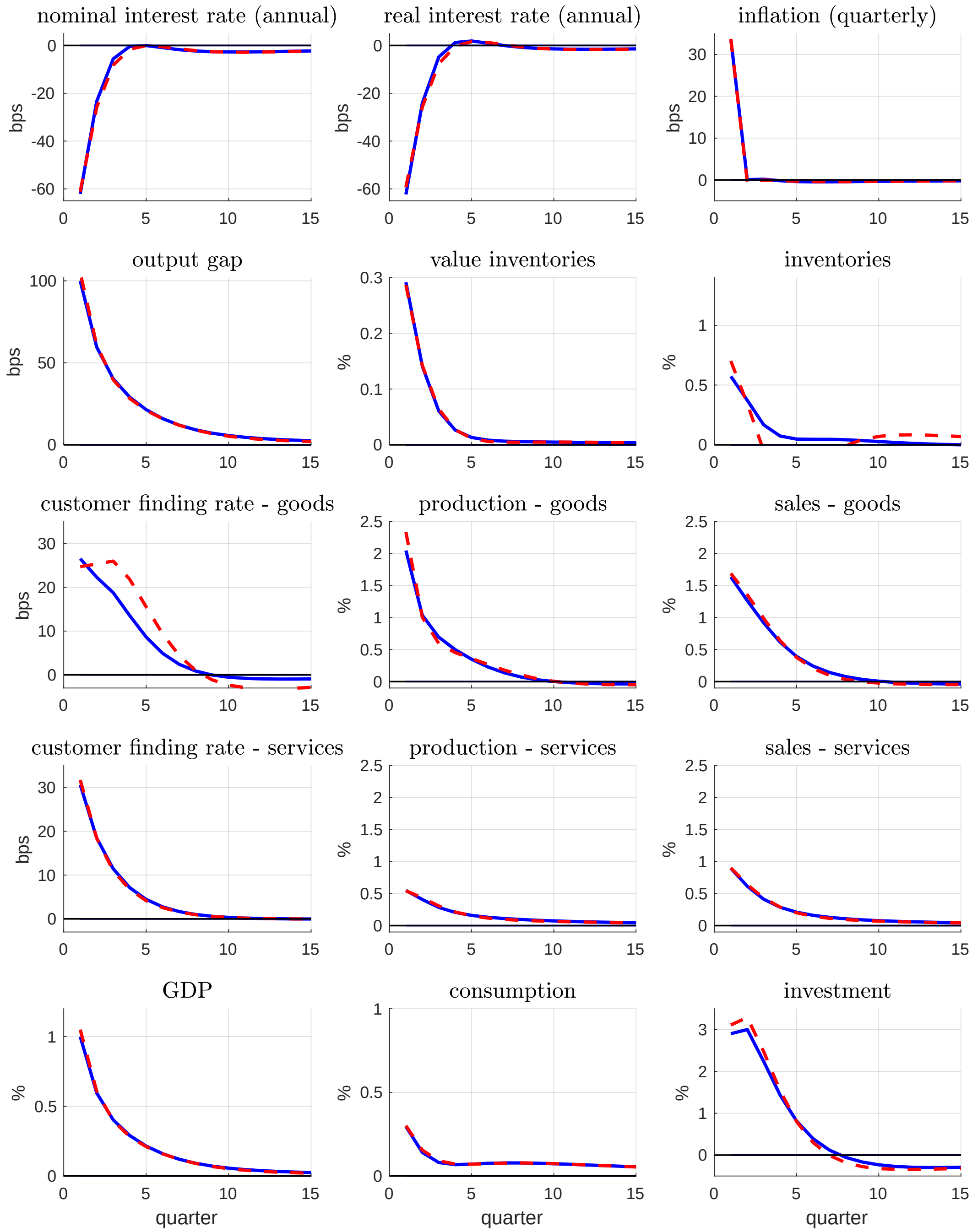
Our calibration of the maintenance cost parameter, η_x, is based on historical data. But such costs are likely to have become much smaller because of technological advances especially in inventory planning. To study how such a reduction in costs affect model predictions, we consider a value for η_x that is only 20% of its benchmark value. We consider such a large change to show that model predictions are very robust.

Figures 28 and 29 show the IRFs for this lower value of η_x for a monetary policy and a TFP shock, respectively. The results are remarkably robust even though this reduction in η_x implies a 13% increase in the steady-state value of an unsold good, λ_x^f. And recall that changes in λ_x^f are crucial to understand the procyclical behavior of the customer-finding rate in the goods sector in response to TFP shocks, so that the magnitude of λ_x^f is likely to matter. Indeed, the response of the customer-finding rate following a TFP shock is substantially more responsive for the lower value of η_x. And this implies that sales increase by 1.14% instead of 1.05% on impact. But although there are some differences, the overall picture remains very similar. One reason is that – although the steady-state value of an unsold good has gone up – the percentage increase is actually smaller, which is consistent with the flatter consumption profile.[165]

So the good news is that the predictions of our model are likely to remain valid even as inventory maintenance costs change.

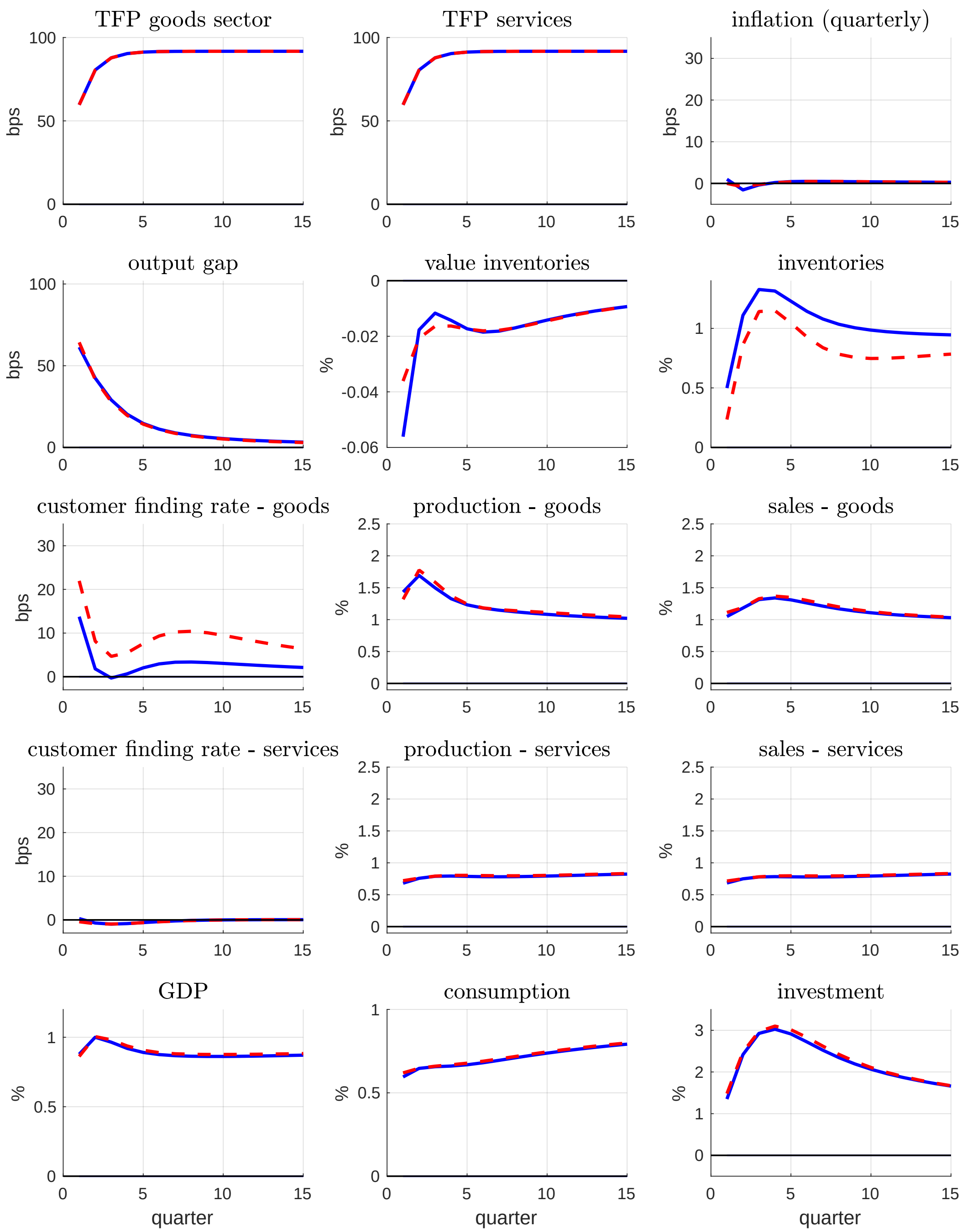
[165] A flatter consumption profile implies a smaller drop in the marginal rate of substitution and thus a smaller drop in asset prices like $\lambda_{x,t}^f$.

Figure 28: monetary-policy shock; lower maintenance costs inventories



Notes. The dashed lines display IRFs of a monetary-policy shock when the maintenance costs of inventories are only 20% of the benchmark value. The value of σ_R is chosen to ensure the same drop in the nominal interest rate on impact. The solid lines display the IRFs using our benchmark parameters.

Figure 29: TFP shock; lower maintenance costs inventories



Notes. The dashed lines display IRFs of a TFP shock when the maintenance costs of inventories are only 20% of the benchmark value. The value of σ_R is chosen to ensure the same drop in the nominal interest rate on impact. The solid lines display the IRFs using our benchmark parameters.