The Role of Sell Frictions for the Cyclical Behavior of Inventories

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

Tiancheng Sun
University of Edinburgh

November 21, 2024

Abstract

Investment in inventories is typically omitted from business-cycle models even though it is responsible for a large share of fluctuations in GDP. The reason is its challenging 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 stockout avoidance motive. Specifically, firms strategically accumulate inventories by bringing more products to the market than they anticipate selling, thereby boosting expected sales. In response to monetary-policy (demand) shocks, 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. In response to TFP (supply) shocks, our framework can also replicate these key facts even though the direct effect of an increased supply would increase the inventory-sales ratio. The reason is that our framework explicitly recognizes that an inventory good is an asset and its value falls during an expansion as expected growth goes together with a higher discount rate. Consequently, firms optimally adjust the price and quantity produced to economize on inventory accumulation leading to a countercyclical inventory-sales ratio.

Keywords: Inventory-sales ratio, search friction, goods and service sector, TFP and monetary-policy shocks.

JEL Classification: E32, G31.

Acknowledgement: We would like to thank ChihHan Hsueh, Greg Kaplan, Michael McMahon, Pascal Michaillat, Ben Moll, Valerie Ramey, Petr Sedláček, and Tim Willems 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 explained with a scale effect, that is, inventory levels would scale up and down with aggregate activity. There is more to it, however, because 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 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]

[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] Blinder (1981) writes *"Inventory fluctuations are important in business cycles; indeed, to a great extent, business cycles are inventory fluctuations."*

[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.

[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).

In this paper, we develop a new framework to model inventories that can replicate these 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 differs from the traditional inventory 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.[12]

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 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

[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. This turns out to be important for the ability of the model to generate observed inventory facts in response to TFP shocks.

[12] The literature is well aware that the interest rate is an important part of the cost of holding inventories. See, for example Deaton and Laroque (1992). New in this paper is that we have a framework in which we show that the key variable is the (asset) value of an inventory good. The discount rate plays a key role in determining that value, but is not the only relevant factor.

sell. 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 procyclical. The customer-finding rate is the same as the sell fraction, that is, the ratio of goods sold relative to the number of 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. Thus, sales are a constant fraction of the sum of newly produced goods and the inventory stock. 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 search effort levels which has a direct effect on the customer-finding rate. What about productivity shocks? An increase in productivity leads to an increase in supply, which would – by itself – lead to a counterfactual *countercyclical* customer-finding rate. In general equilibrium, it would go together with an increase in income which in turn leads

[13] See equation 1.

[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.

to an increase in demand and search effort. This opposite effect will lead – at best – to an acyclical 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 to economize on the level of inventories relative to that of sales. 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. 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.[16]

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 restaurant tables. The last section concludes.

2 Empirical Findings

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.[17] Inventories

[16] There are ways to get the desired hump-shaped consumption Impulse Response Function (IRF) 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 more realistic non-stationary process, because it 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 provided in this paper is that the way TFP is modeled not only affects an asset price, the value of an inventory good, but also key quantities.

[17] 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

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.[18] 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. An important aspect of our theoretical model is that not all goods available for sale are sold because of a goods-market friction. The following equation shows that the fraction sold is a simple transformation of the inventory-sales ratio, a well-measured and popular variable in the inventories literature.[19]

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

Here, $y_{g,t}$ 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 second equality uses that the amount produced, $y_{g,t}$, equals sales plus the investment in inventories, that is, $y_{g,t} = s_{g,t} + x_t - (1-\delta_x)x_{t-1}$. We will also refer to the fraction sold as the customer-finding rate. The remainder of this section discusses its empirical properties together with those of other key inventory variables.

Empirical properties of the customer-finding rate. The top panel of table 1, documents 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.[20] That is, quarterly sales, quarterly newly produced goods, and the stock of inventories

irregular fluctuations observed during the pandemic.

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

[19]The superscript f indicates that the fraction sold is viewed from the point of view of firms.

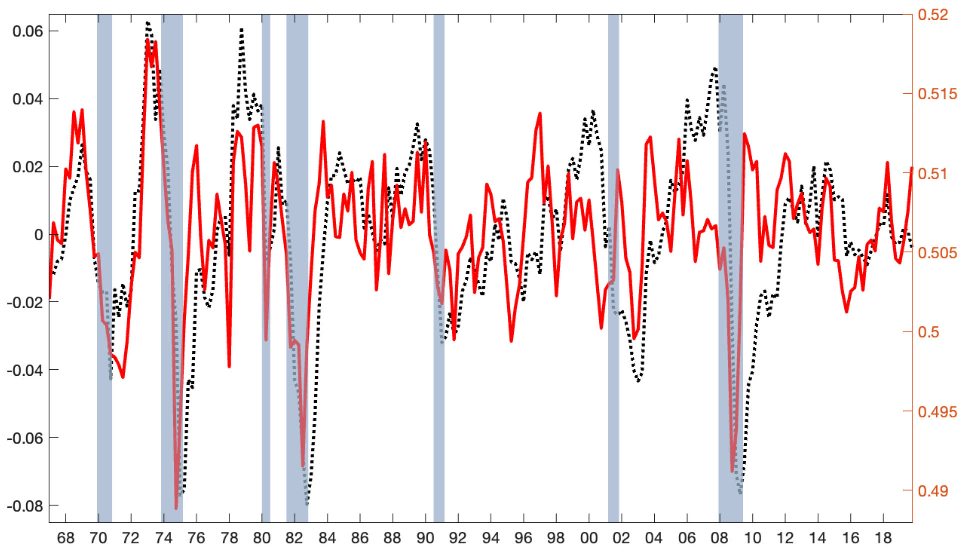
[20]The BEA and the media report the inventory-sales ratio based on monthly sales, which is three times bigger than the one used here based on quarterly sales.

Table 1: Inventory stylized facts and model predictions

	DATA	MODEL $\nu_g, \nu_s, \Gamma_y, \sigma_R/\sigma_A$ based on moment matching			MODEL $\nu_g, \nu_s, \Gamma_y, \sigma_R/\sigma_A$ estimated		
		TFP&R	TFP	R	TFP&R	TFP	R
$\mathbb{E}[f_g^f]$	0.506	=	=	=	=	=	=
	(0.002)						
$\mathbb{E}[\frac{x}{s_g}]$	0.976	=	=	=	=	=	=
	(0.077)						
$\frac{\sigma_{y_g}}{\sigma_{s_g}}$	1.124	1.175	1.212	1.080	1.109	1.192	1.005
	(0.026)	(0.030)	(0.034)	(0.020)	(0.027)	(0.033)	(0.019)
$\frac{\sigma_{f_g^f}}{\sigma_{y_g}}$	0.170	0.099	0.066	0.164	0.147	0.073	0.218
	(0.157)	(0.009)	(0.005)	(0.005)	(0.012)	(0.005)	(0.008)
$\rho(f_g^f, y_g)$	0.514	0.594	0.445	0.935	0.682	0.559	0.901
	(0.109)	(0.067)	(0.023)	(0.005)	(0.062)	(0.029)	(0.008)
$\rho(x, y_g)$	0.630	0.839	0.867	0.977	0.771	0.869	0.685
	(0.095)	(0.026)	(0.018)	(0.001)	(0.044)	(0.019)	(0.043)
$\rho(\Delta \ln x, y_g)$	0.442	0.839	0.558	0.496	0.472	0.543	0.294
	(0.104)	(0.026)	(0.046)	(0.038)	(0.053)	(0.047)	(0.056)
$\rho(f_g^f, x_{-1})$	-0.223	-0.018	-0.222	0.646	0.071	-0.063	0.391
	(0.105)	(0.115)	(0.078)	(0.049)	(0.124)	(0.096)	(0.047)

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 (1). 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 model 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 alternative based on explicit moment matching consists of $\nu_g = \nu_s = 0.565$, $\Gamma_y = 0.03$, and $\sigma_R/\sigma_A = 0.5921$. 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 (-) and output (:)



Notes. This figure plots the HP-filtered values of the customer-finding rate (red/solid and scale on the right axis), i.e., the fraction of available goods sold, calculated according to equation (1) 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. It also plots (the log of) production in the goods and structures sector (black/dashed and scale on the left axis).

are roughly equal to each other.

The customer-finding rate is strongly procyclical.[21] 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 recessions. 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 the 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 coefficient of the customer-finding rate with the level of inventory goods brought to the market and the one with newly produced goods have different signs. And this moments turns out to be helpful when choosing parameter values.

[21]We use HP-filtered data to evaluate business-cycle fluctuations. The results in the table are based on the constructed production series for goods and structure, but the same procyclicality is observed when GDP is used. 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.

Traditional inventory facts. The bottom panel of table 1 focuses on traditional inventory statistics. First, the inventory-sales ratio is countercyclical which must be true given that the customer-finding rate is procyclical. Second, inventories of finished goods are procyclical, that is, recessions are *not* periods when sellers are stuck with increased stocks of unsold goods. Third, a related – but even more intriguing – observation is the well-known fact that production is *more* volatile than sales.[22] For our sample, output is 12% more volatile. 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.

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.[23] If the increase in the customer-finding rate would be bigger than 0.249 basis points, then sales would be more volatile than output on impact.

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.

Importance of investment in inventories for cyclical fluctuations aggregates. 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

[22]See Ramey and West (1999) for a discussion and early empirical evidence. As pointed out by the authors, the identity $y_{g,t} = s_{g,t} + x_t - (1-\delta)x_{t-1}$ 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.

[23]Also, the relative standard deviations reported in the table are based on HP-filtered data whereas this simple numerical example focuses on raw numbers.

the sum of the two expenditure components usually modeled explicitly in business-cycle models which are consumption and investment. We look both at HP-filtered levels and first-differences.[24] When the aggregate considered is equal to the sum of the components considered, then the variance decomposition gives numbers that add up to 1.[25]

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.

Table 2: Variance decomposition

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

Notes. This table consists of two panels. The panel on the left focuses on business-cycle frequencies and the one on the right on first differences. The numbers in each block document the relative importance of the component listed in the top row for fluctuations in the aggregates listed in the left column. 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 parentheses and these are calculated using the VARHAC procedure of Den Haan and Levin (1997) which corrects for serial correlation and heteroskedasticity.

Customer-finding rate for services Firms providing services are also likely to face a sell friction. For example, a restaurant will fill more tables – but a smaller fraction – if it increases the number of tables. Whereas, the customer-finding rate for the goods sector can be constructed from the observed inventory-sales ratio using equation (1), this is not an option for services since there are no inventories. Although only for a short sample, 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 is discussed in appendix A.2.

[24] Since CIPI is at times negative, we cannot take logs to get scale free statistics. Therefore, we first scale the 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.

[25] 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.

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. 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 firms that sell services and these 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 retail sector; 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.[26]

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, investment goods, and 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.[27]

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,

$$n_t = \left(\int_{h=0}^{1} n_{h,t}^{\frac{\varepsilon_n - 1}{\varepsilon_n}} dh \right)^{\frac{\varepsilon_n}{\varepsilon_n - 1}}, \tag{2}$$

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

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

$$n_{h,t} = \left(\frac{W_{h,t}}{W_t} \right)^{-\varepsilon_n} n_t, \tag{3}$$

[26]We do not model inventories in the form 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).

[27]Eurostat reports that the time spend on shopping and enjoying personal 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.

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

$$W_t \equiv \left(\int_{h=0}^{1} W_{h,t}^{1-\varepsilon_n} dh \right)^{\frac{1}{1-\varepsilon_n}}. \tag{4}$$

Adjusting the nominal wage incurs a utility cost given by[28]

$$\frac{1}{2}\eta_W \left(\frac{W_{h,t}}{W_{h,t-1}} - 1 \right)^2 n_t, \tag{5}$$

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,

$$c_{h,t} + i_{h,t} \leq s_{h,t} \equiv \left(\int_{i=0}^{1} s_{i,h,t}^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{6}$$

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 aggregated goods 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:

$$s_{i,h,t} = f_{i,t}^{b} e_{i,h,t}. \tag{7}$$

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 obtain one unit of good i and households take this as given.[29] The superscript b indicates that the friction is viewed from the buyer's 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. We assume that this acquisition cost is a perfect substitute for consumption goods.[30]

[28]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.

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

[30]In the full model, this acquisition cost can be in terms of both services and goods. As shown in

Household problem. The household problem is given by

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

subject to

$$\int_{i=0}^{1} \frac{P_{i,t}}{P_t} s_{i,h,t} di + \frac{b_{h,t}}{P_t} \leq \frac{W_{h,t}}{P_t} n_{h,t} + r_{k,t} k_{h,t-1} + d_{h,t} + \frac{1 + R_{t-1}}{P_t} b_{h,t-1}, \tag{8a}$$

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

$$k_{h,t} = (1 - \delta_k) k_{h,t-1} + i_{h,t} \left(1 - \frac{\eta_i}{2} \left(\frac{i_{h,t}}{i_{h,t-1}} - 1 \right)^2 \right), \tag{8c}$$

$$s_{i,h,t} = f_{i,t}^b e_{i,h,t}, \tag{8d}$$

$$e_{h,t} = \int_{i=0}^{1} e_{i,h,t} di, \tag{8e}$$

$$n_{h,t} = \left(\frac{W_{h,t}}{W_t} \right)^{-\varepsilon_n} n_t. \tag{8f}$$

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 $\overline{\xi}_e$ is used as a normalization to set the effort term, $\xi_e e_{h,t} - \overline{\xi}_e$, equal to zero in the steady state. Alternatively, one can interpret $\overline{\xi}_e$ as home production and $\xi_e e_{h,t}$ as the effort cost.

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 (8a) is denoted by $\lambda_{h,t}$, the one associated with the purchases-allocation constraint (8b) by $\psi_{h,t}\lambda_{h,t}$, and the one associated with the capital accumulation equation (8c) by $\lambda_{k,h,t}\lambda_{h,t}$. That is, we ex-

appendix D, the assumption of perfect substitutability 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).

press 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

$$\psi_{h,t}\lambda_{h,t} = \xi_c \left(c_{h,t} - \xi_e(e_{h,t} - \overline{\xi}_e)\right)^{-\gamma}, \tag{9a}$$

$$\lambda_{h,t}\frac{1}{P_t} = \beta\mathbb{E}_t\left(\lambda_{h,t+1}\frac{1+R_t}{P_{t+1}}\right), \tag{9b}$$

$$\lambda_{h,t}\lambda_{k,h,t} = \beta\mathbb{E}_t\left(\lambda_{h,t+1}r_{k,t+1} + \lambda_{h,t+1}\lambda_{k,h,t+1}\left(1-\delta_k\right)\right), \tag{9c}$$

$$\begin{aligned}\psi_{h,t} &= \lambda_{k,h,t}\left(1 - \frac{\eta_i}{2}\left(\frac{i_{h,t}}{i_{h,t-1}} - 1\right)^2 - \eta_i\frac{i_{h,t}}{i_{h,t-1}}\left(\frac{i_{h,t}}{i_{h,t-1}} - 1\right)\right)\\ &\quad + \beta\mathbb{E}_t\left(\frac{\lambda_{h,t+1}}{\lambda_{h,t}}\lambda_{k,h,t+1}\eta_i\left(\frac{i_{h,t+1}}{i_{h,t}}\right)^2\left(\frac{i_{h,t+1}}{i_{h,t}} - 1\right)\right),\end{aligned} \tag{9d}$$

$$\psi_{h,t}\lambda_{h,t}\left(\frac{s_{h,t}}{s_{i,h,t}}\right)^{\frac{1}{\varepsilon}} f_{i,t}^b = \xi_c\left(c_{h,t} - \xi_e(e_{h,t} - \overline{\xi}_e)\right)^{-\gamma}\xi_e + \lambda_{h,t}\frac{P_{i,t}}{P_t}f_{i,t}^b, \tag{9e}$$

$$\begin{aligned}&\left(\varepsilon_n\xi_n\left(\frac{W_{h,t}}{W_t}\right)^{-\varepsilon_n-1} + \lambda_{h,t}w_t\left(1-\varepsilon_n\right)\left(\frac{W_{h,t}}{W_t}\right)^{-\varepsilon_n}\right)\\ &+ \beta\mathbb{E}_t\left(\eta_W\frac{n_{t+1}}{n_t}\left(\frac{W_{h,t+1}}{W_{h,t}} - 1\right)\frac{W_tW_{h,t+1}}{W_{h,t}^2}\right)\\ &= \eta_W\left(\frac{W_{h,t}}{W_{h,t-1}} - 1\right)\frac{W_t}{W_{h,t-1}},\end{aligned} \tag{9f}$$

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

$$\begin{aligned}\left(\varepsilon_n\xi_n n_t + \lambda_t w_t\left(1-\varepsilon_n\right)\right) + \beta\mathbb{E}_t\left(\eta_W\frac{n_{t+1}}{n_t}\left(\frac{W_{t+1}}{W_t} - 1\right)\frac{W_{t+1}}{W_t}\right)&\\ = \eta_W\left(\frac{W_t}{W_{t-1}} - 1\right)\frac{W_t}{W_{t-1}}.&\end{aligned} \tag{10}$$

Combining equations (9a) and (9e) and using that $\psi_{h,t} = \psi_t$ gives

$$\frac{\xi_e}{f_{i,t}^b} + \frac{P_{i,t}}{P_t}\frac{1}{\psi_t} = 1. \tag{11}$$

Aggregate price index. The aggregate price index, P_t, is defined to satisfy[31]

$$P_t = \left(\int_{i=0}^{\infty} \left(P_{i,t} + \frac{P_t \xi_e}{f_{i,t}^b} \right)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}. \tag{12}$$

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 (11) implies that $\psi_t = 1$ in each period.[32] In a symmetric equilibrium, we get

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

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.[33]

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. Equations (9a) and (9e) imply that the demand for good i is given by

$$s_{i,t} = \int_{h=0}^{\infty} s_{i,h,t} dh = \left(\frac{\xi_e}{f_{i,t}^b} + \frac{P_{i,t}}{P_t} \right)^{-\varepsilon} s_t, \tag{14}$$

where we have used that all household are identical so that the h subscript is no longer needed. 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, $\xi_e / f_{i,t}^b$.

[31] 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. See equation (25).

[32] 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. Thus, another motivation for the price index definition is to turn the reasoning around and start with the condition that $\psi_t = 1$, which then results in our price index.

[33] In Kryvtsov and Midrigan (2013), there is also a wedge between $P_{i,t}$ and P_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.

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 good.[34]

Selling process and the goods-market friction. We assume that the total amount of goods sold, $s_{i,t}$, depends on buyers' search effort and the firm's supply. Specifically, it is given by

$$s_{i,t} = \mu e_{i,t}^{1-\nu}(y_{i,t} + (1-\delta_x)x_{i,t-1})^{\nu}, \text{ with } 0 < \nu < 1, \tag{15}$$

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.[35] 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 success of a match is taken as given.[36]

There are different ways to motivate this approach. 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 collection and 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.

[34] In appendix B, 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.

[35] 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 measurement for GDP.

[36] 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.

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.[37] 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 (15) 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

$$f^b_{i,t} = f^b(\theta_{i,t}) = \mu\theta_{i,t}^{-\nu} = \mu\left(\frac{e_{i,t}}{y_{i,t} + (1-\delta_x)x_{i,t-1}}\right)^{-\nu}, \tag{16a}$$

$$f^f_{i,t} = f^f(\theta_{i,t}) = \mu\theta_{i,t}^{1-\nu} = \mu\left(\frac{e_{i,t}}{y_{i,t} + (1-\delta_x)x_{i,t-1}}\right)^{1-\nu}, \tag{16b}$$

where $\theta_{i,t}$ represents tightness in this market.[38]

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

$$s_{i,t} \leq \left(\frac{\xi_e}{\mu\left(\frac{e_{i,t}}{y_{i,t}+(1-\delta_x)x_{i,t-1}}\right)^{-\nu}} + \frac{P_{i,t}}{P_t}\right)^{-\varepsilon} s_t. \tag{17}$$

[37] 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 units. 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.

[38] 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.

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.[39] This does not only give 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

$$x_{i,t} \;=\; \left(1 - f^f(\theta_{i,t})\right)\left(y_{i,t} + (1-\delta_x)x_{i,t-1}\right). \tag{19}$$

Cost minimization. The production technology is given by

$$y_{i,t} = A_t \left(n_{i,t}\right)^{\alpha} k_{i,t}^{1-\alpha}, \tag{20}$$

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:

$$\ln\left(\frac{A_t}{A_{t-1}}\right) = \rho_A \ln\left(\frac{A_{t-1}}{A_{t-2}}\right) + \varepsilon_{A,t}, \tag{21}$$

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 literature also indicates that there are TFP shocks that are associated with (further) expected growth.[40] Whether A_t is stationary or not does not matter for standard business-cycle properties.[41] But a key result of this paper is that it does matter for implied inventory properties.

The cost of producing $y_{i,t}$ is given by $w_t n_{i,t} + r_{k,t} k_{i,t}$ and minimizing the cost subject to the constraint that $y_{i,t} \leq A_t \left(n_{i,t}\right)^{\alpha} k_{i,t}^{1-\alpha}$ gives the usual condition that $w_t n_{i,t} / r_{k,t} k_{i,t} =$

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

$$s_{i,t} = \tilde{\gamma}_t (y_{i,t} + (1-\delta_x)x_{i,t-1})^{\phi} \left(P_{i,t}/P_t\right)^{\varepsilon}, \tag{18}$$

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 $d_t(\cdot)$.

[40]See Beaudry and Portier (2006).

[41]See Christiano and Eichenbaum (1990).

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

Firm $i \in [0,1]$ solves the following optimization problem:[42]

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

s.t.

$$s_{i,t} \le \left(\frac{\xi_e}{f^b(\theta_{i,t})} + \frac{P_{i,t}}{P_t}\right)^{-\varepsilon} s_t, \tag{22a}$$

$$s_{i,t} \le f^f(\theta_{i,t})(y_{i,t} + (1-\delta_x)x_{i,t-1}), \tag{22b}$$

$$x_{i,t} \le \left(1 - f^f(\theta_{i,t})\right)\left(y_{i,t} + (1-\delta_x)x_{i,t-1}\right). \tag{22c}$$

When discussing the first-order conditions, we focus on the symmetric equilibrium. The Lagrange multipliers of the demand constraint, the sales constraint, and the inventories accumulation constraint are denoted by $\lambda^f_{d,t}$, $\lambda^f_{s,t}$, and $\lambda^f_{x,t}$, respectively. In appendix, C, we present the full set of first-order conditions, provide an interpretation, and show how they can be combined to a more concise system that is easier to interpret. That system is given by

$$1 = \frac{\xi_e}{f^b(\theta_t)} + \frac{P_{i,t}}{P_t}, \tag{23a}$$

$$\lambda^f_{d,t} = \frac{P_{i,t}}{P_t} + \frac{1-f^f(\theta_t)}{f^f(\theta_t)}\lambda^f_{x,t} - \frac{MC_t}{f^f(\theta_t)} \tag{23b}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = \varepsilon\lambda^f_{d,t}\frac{\nu}{1-\nu}\xi_e\theta_t, \tag{23c}$$

$$\lambda^f_{x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\begin{array}{c} f^f(\theta_{t+1})\lambda^f_{s,t+1} \\ +(1-f^f(\theta_{t+1}))\lambda^f_{x,t+1}\end{array}\right)\right], \tag{23d}$$

$$1 - \varepsilon\lambda^f_{d,t} = \eta_P\frac{P_t}{P_{i,t}}\left(\begin{array}{c} \left(\frac{P_{i,t}}{P_{i,t-1}}-1\right)\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \\ -\beta\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\frac{P_{i,t+1}}{P_{i,t}}-1\right)\left(\frac{P_{i,t+1}}{P_{i,t}}\right)\frac{s_{t+1}}{s_t}\right]\end{array}\right). \tag{23e}$$

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 (23a) 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.

In the standard New-Keynesian model, the Lagrange multiplier of the demand

[42]In appendix B, 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 environment with inventories.

constraint, $\lambda^f_{d,t}$, is equal to the markup, which is equal to $1 - MC_t$ when scaled by the aggregate price index, P_t. Equation (23b) indicates that the expression for $\lambda^f_{d,t}$ is a bit more complicated here. The revenue term is not equal to 1 but equal to $P_{i,t}/P_t \leq 1$ because search costs drive a wedge between the price the firm receives and the aggregate price. The marginal cost term is also different. To understand the modification suppose that $f^f_t = 1/4$. A one-unit reduction in sales means that output can be lowered by $1/f^f_t = 4$ units so costs drop by $4 \times MC_t$ and not by $1 \times MC_t$. This reduction of output with four units causes not only sales to be one unit less, but also lowers end-of-period inventories with three ($= (1-f^f_t)/f^f_t$) units.

Equation (23c) is a rewritten version of the firm's first-order condition for θ_t. 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, which is related to the gap between marginal costs and the value of an inventory good. Equations (23d) 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. Finally, equation (23e) is the first-order condition related to $P_{i,t}$. This equation gives us our modified New-Keynesian Phillips Curve, which we discuss next.

New-Keynesian (NK) Phillips Curve for our model with inventories. In the remainder of this section, we focus on the symmetric equilibrium.[43] The firm's first-order condition for $P_{i,t}$, i.e., equation (23e), is *identical* to the standard New-Keynesian Phillips Curve, however, the expression for $\lambda^f_{d,t}$ is different. When we use the expression for $\lambda^f_{d,t}$ from equation (23b) in (23e), then we get

$$1 - \varepsilon \left(\frac{P_{i,t}}{P_t} + \frac{1 - f^f(\theta_t)}{f^f(\theta_t)} \lambda^f_{x,t} - \frac{MC_t}{f^f(\theta_t)} \right) =$$
$$\eta_P \frac{P_t}{P_{i,t}} \left(\begin{array}{c} \left(\frac{P_{i,t}}{P_{i,t-1}} - 1 \right) \left(\frac{P_{i,t}}{P_{i,t-1}} \right) \\ -\beta \mathbb{E}_t \left[\left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(\frac{P_{i,t+1}}{P_{i,t}} - 1 \right) \left(\frac{P_{i,t+1}}{P_{i,t}} \right) \frac{s_{t+1}}{s_t} \right] \end{array} \right). \tag{24}$$

If there is no goods-market friction, then $P_{i,t} = P_t$ and $\lambda^f_{d,t} = 1 - MC_t$, which means that we get the standard New-Keynesian Phillips curve which contains only marginal costs, current inflation, and expected inflation. With our goods-market friction, it also includes the gap between $P_{i,t}$ and P_t, the customer-finding rate, $\lambda^f_{d,t}$, and the value of carrying a good into the next period as inventory, $\lambda^f_{x,t}$, which is an NPV and depends on market discount rates.

[43] 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.

A direct role for the real interest rate in the monetary transmission mechanism. 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 Phillips Curve, since it is the inverse of the marginal rate of substitution 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:

$$R_t = -\ln\beta(1-\Gamma_{\text{lag}}) + \Gamma_{\text{lag}}R_{t-1} + \Gamma_\pi\left(\frac{P_t}{P_{t-1}}-1\right) + \Gamma_y\left(\frac{Y_t}{\widetilde{Y}_t}-1\right) + \varepsilon_{R,t}, \tag{25}$$

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 Why the model can replicate observed inventory facts with either demand or supply shocks

Our model is a dynamic model and determining how shocks affect model variables requires a numerical solution taking into account expectations of future developments. However, the condensed sub-system (23) 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 (23a), (23b), and (23c) determine θ_t, MC_t, and $P_{i,t}/P_t$ and also the customer-finding rate as it is a function of tightness, θ_t, only.[44]

This subsystem allows us to derive the following two propositions.[45]

Proposition 1 $\frac{\partial f^f(\theta_t)}{\partial\lambda^f_{d,t}} < 0$. *That is, an increase in inflationary pressure (relative to expected future inflation), which leads to a decrease in* $\lambda^f_{d,t}$ *according to equation (23e),*

[44]Having such convenient analytical expressions 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.

[45]Proofs are provided in appendix D.

is associated with an increase in the customer-finding rate.[46]

Proposition 2 $\frac{\partial f^f(\theta_t)}{\partial \lambda^f_{x,t}} < 0$. *That is, an increase in the value of carrying an unsold good into the future as inventory is associated with a reduction in the customer-finding rate.*

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. The reason is that an increase in the customer-finding rate may be a protective measure. 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 and explain why the model can replicate observed joint behavior of inventories, production, and sales for *both* demand and supply shocks.

A successful 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. Since there is a lot of debate on the importance of demand versus supply shocks, it would be helpful if the model can replicate these empirical findings for both types of shocks.[47] If the first prediction is satisfied, then the model also correctly predicts that the inventory-sales ratio is countercyclical. These four properties are related to each other and for the model to generate the last three it is important to get the first one right.

A procyclical customer-finding rate in response to demand shocks. Proposition 1 shows that this is true in our model. Why? A decrease in $\lambda^f_{d,t}$ represents 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

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

[47]This is quite ambitious as demand and supply shocks clearly do affect the economy differently. And it is, of course, possible to match unconditional moments with a model in which moments generated by the different types of shocks are quite different. Another motivation for this additional challenge is that appendix A.4 provides some indication that the customer-finding rate is indeed procyclical in response to both types of shocks.

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 real terms must increase which in turn increases output. This is welfare improving for the economy as a whole because the expansion is due to a reduction in 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 efforts. When restricted to reducing sales by increasing prices, firms increase tightness by restricting the increase in output somewhat which increases search costs for customers which in turn would dampen the increase in demand, just as an increase in the price would. Important for us is that this implies a procyclical customer-finding rate in response to nominal demand shocks in the presence of sticky prices. Although the flexibility to affect tightness dampens the increase in output, output still increases 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.[48]

As discussed below in section 3.5, this channel is similar to the "stockout-avoidance" recognized in the literature. Although stockouts never happen with our representative-firm approach, the increase in the customer-finding rate does push the level of sales closer to the level of goods that firms bring to the market, that is, closer to a stockout.

A procyclical customer-finding rate in response to supply shocks. One would think that a positive supply shock would lower the customer-finding rate, since the associated increase in the amount of goods that firms bring to the market has a direct negative impact. One general-equilibrium mechanism going in the opposite direction is that higher productivity leads to higher income levels which induces buyers to increase effort. For this indirect effect to overturn the direct negative effect of increased supply, it must be the case that firms set output levels and prices such that supply of available goods relative to effort would fall. This definitely does not happen in our model. As indicated by proposition 2, however, there is another mechanism that affects the cyclicality of the customer-finding rate. If the value of carrying a good into the next period as inventory, $\lambda_{x,t}$, falls, then this leads to an increase in the customer finding rate. The intuition for proposition 2 is relatively simple. A drop in $\lambda^f_{x,t}$ 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. With proposition 2 in place, the remaining question is whether $\lambda^f_{x,t}$ is countercyclical. If it is, then the customer-finding rate will be procyclical.

[48]Appendix D proves that marginal costs increase together with tightness when $\lambda^f_{d,t}$ falls for small changes around the steady-state. This implies that output must increase. Numerical results indicate that this is a robust finding for larger changes as well.

By manipulating the firm's first-order conditions, we get that

$$\lambda^f_{x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\frac{\lambda_{t+1}}{\lambda_t}MC_{t+1}\right]. \tag{26}$$

That is, the value of bringing a good as inventory into the next period is equal to the expected value of producing one good less next period, taking into account depreciation. This equation makes clear that an unsold good is an asset with a future payoff, namely not having to produce the good in the future. And key in determining its value is the marginal rate of substitution, i.e., the inverse of the discount rate.[49] 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. It is this mechanism that makes it possible to generate a procyclical customer-finding rate in response to TFP shocks so that our model can match observed unconditional inventory facts for both demand *and* supply shocks.

The remaining question is whether the marginal rate of substitution, that is, the discount factor, is countercyclical. Key in generating a countercyclical discount factor is that consumption is expected to increase following a shock. Ramey (2016) provides empirical support for this. That is, the IRF of consumption should be either hump-shaped before it declines when TFP is stationary or continue to increase and settle at a higher level when TFP is non-stationary. The discount factor is robustly countercyclical in our model in response to TFP shocks, because we have adopted a realistic $I(1)$ process for TFP.[50]

The other three key inventory properties in response to both types of shocks. What about the other three inventory facts listed above? Recall that the level of sales, s_t, is equal to the customer-finding rate, f^f_t, times the amount of goods made available for sale, $y_t + (1-\delta)x_{t-1}$. Since inventories, x_{t-1}, are a stock they will respond more slowly than newly produced output, y_t. Consequently, sales will automatically be less volatile than output and (investment in) inventories will be procyclical *as long as* the fluctuations of the customer-finding rate are – as in the data – not too large. The variability of the customer-finding rate depends crucially on the

[49]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 D shows that this must be true locally around the steady state, but we find it to be true numerically in our simulations as well.

[50]Appendix H.1 documents that our model can generate hump-shaped consumption IRFs (and thus a countercyclical $\lambda_{x,t}$) for *some* parameter values when A_t is stationary. Our utility function is very basic and this makes it harder to generate a hump-shaped consumption IRF with a stationary TFP processes. Hump-shaped consumption IRFs with stationary TFP processes become possible, however, by enriching the model. For example, by introducing habits in the utility function as shown by Fuhrer (2000).

value of ν which determines the curvature of f_t^f as a function of tightness; the lower the value of ν the larger the fluctuations in f_t^f. Although our model robustly predicts a procyclical f_t^f for both demand and supply shocks, the responses of f_t^f are larger for demand shocks. In principle, this could indicate that the appropriate value for ν depends on whether fluctuations are driven by demand or supply shocks. Fortunately, there is a range of values for ν such that the model can replicate key empirical findings for both demand and supply shocks using the same value of ν.

Appendix E discusses properties of the goods-only model in more detail. Specifically, it discusses how the magnitude of model responses to shocks depend on features such as price stickiness, wage stickiness, investment adjustment costs, and monetary policy. Two aspects are worth mentioning here.

The role of 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 $\lambda_{x,t}^f$ and/or $\lambda_{d,t}^f$? 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.[51] The situation is a bit more complicated in our setup, but still follows the logic of the standard 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, $\lambda_{x,t}^f$. In that situation, overall activity is higher because of sticky wages, but the optimal level of tightness will only be different if $\lambda_{x,t}^f$ or $\lambda_{d,t}^f$ take on different values. Consequently, inventory accumulation will also only be different if wage stickiness affects the behavior of $\lambda_{x,t}^f$ or $\lambda_{d,t}^f$.

The role of systematic monetary policy. Another important lesson from the discussion in appendix E is that the response of monetary policy to the output gap is quite important. This is controlled by the parameter Γ_y. This is quite intuitive. A positive productivity shock leads to a nontrivial positive output gap in the full model in which (approximate) divine coincidence no longer holds. If $\Gamma_y > 0$, then this means that the positive productivity shock is accompanied by a negative nominal demand response of the central bank. If this response is too large, then it could undo the desirable properties that our model generates for supply shocks. This is the reason why Γ_y, together with ν, plays a key role.

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

3.5 Comparison with the literature

The two propositions of section 3.4 describe 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 and its value changes in response to changes in the discount rate. 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 novel in terms of explaining the cyclical behavior of inventories.[52] 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.[53] 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 there are price-adjustment costs *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 in the value of inventory goods which means that the cost of oversupply relative to the cost of 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.[54]

At first sight, this setup with firm-specific idiosyncratic shocks looks quite different

[52] In fact, the discount rate, a key aspect of the value of an inventory good, is typically assumed to be constant in the inventory literature. Of course, it is well known that the discount or interest rate is an important aspect of the maintenance or carrying costs of holding inventories. See, for example, Richardson (1995) for a quantitative assessment and Deaton and Laroque (1992) for a model in which holding inventories are costly because of a positive (constant) interest rate. de Sousa Rodrigues and Willems (2024) highlight the empirical relevance of variations in the interest rate and document how the level of housing inventory (fraction of unoccupied houses) interacts with the interest rate in terms of how monetary policy shocks affect housing costs. Our model makes clear that it is not just the discount rate that matters, but the *discounted* value of next period's marginal costs, although changes in the discount rate turn out to be the most relevant component quantitatively.

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

[54] 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.

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 presence of 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 an unexpected stockout or an unexpected increase in 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.[55] 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 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.[56] 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.[57] 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 and when they are not.[58]

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

[56]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.

[57]Chen (2017) considers a general-equilibrium model with TFP shocks that can replicate key inventory facts, but only considers the case with flexible prices. It adopts the timing assumption of KM regarding price setting and production and also relies on good-specific idiosyncratic demand shocks. In addition, it introduces a search friction. Specifically, search effort by the households affects the variety of goods consumed, v, and this introduces an increasing returns to scale aspect to the model because utility of consumption is assumed to be a function of cv^{ρ} with $\rho > 1$.

[58]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. 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 delay ensures that (i) output is more

There are two reasons for this. The first is that monetary policy follows an exogenous monetary-supply rule in the KM model 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.

4 Model with goods and services

Whereas there are no inventories in the service sector, providers of services are also likely to face frictions in finding buyers. To adapt our sell-friction mechanism to the service sector, one simply sets 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.

The previous section documented the importance of fluctuations in the value of unsold goods, $\lambda^f_{x,t}$, to match observed fluctuations in the customer-finding rate. But the value of unsold services is fixed (at zero). Thus, the model's predictions for the cyclical behavior of the customer-finding rate for services may differ from those for the goods sector.[59] Moreover, the presence of a service sector might affect the cyclical properties of inventory, production, and sales in the goods sector which are the main focus of this paper. We want to answer this question in a transparent insightful way by 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.

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.

[59]Whereas we have a direct measurement for the goods-sector customer-finding rate (a simple transformation of the inventory-sales ratio), no such measure is available for services, although we provide some possibly relevant observations for the cyclical behavior of the service-sector customer-finding rate in appendix A.2.

4.1 Key assumptions and updated demand functions

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. Nevertheless, there are some changes and in particular, the relative price of goods versus services matters.

Households obtain utility from a homogeneous consumption good, but we interpret it as a mixture of durables and non-durables. Under the assumption that the benefit flow 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 a mixture of both durable and non-durable goods. This complication does not affect the properties that we are interested in like the behavior of customer-finding rates in the two sectors or inventory facts. But it does ensure that – despite the Leontief structure – *expenditures* on consumer goods are more volatile than purchases of consumer services, as is observed in the data.[60]

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

The following set of equations captures our setup:

$$c_t \leq \min\left\{\frac{\bar{c}_{g,t}}{\omega_{g,c}}, \frac{c_{s,t} - \Upsilon_s(\xi_e e_t - \overline{\xi}_e)}{\omega_{s,c}}\right\}, \tag{27a}$$

$$i_t \leq \min\left\{\frac{i_{g,t}}{\omega_{g,i}}, \frac{i_{s,t}}{\omega_{s,i}}\right\}, \tag{27b}$$

$$\bar{c}_{g,t} = (1 - \delta_c)\,\bar{c}_{g,t-1} + c_{g,t} - \Upsilon_g(\xi_e e_t - \overline{\xi}_e). \tag{27c}$$

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, and $i_{s,t}$ investment services.

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, search effort is associated with a cost in terms of services when $\Upsilon_s > 0$. The model allows for both coefficients to be positive.

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

Demand functions. From the household problem, we can derive the following demand functions:[61]

$$s_{i,g,t} = \left(\left(\Upsilon_g + \Upsilon_s \frac{P_{s,t}}{P_{g,t}}\right) \frac{\xi_e}{f^b\left(\theta_{i,g,t}\right)} + \frac{P_{i,g,t}}{P_{g,t}}\right)^{-\varepsilon_g} s_{g,t}, \tag{28}$$

$$s_{i,s,t} = \left(\left(\Upsilon_g \frac{P_{g,t}}{P_{s,t}} + \Upsilon_s\right) \frac{\xi_e}{f^b\left(\theta_{i,s,t}\right)} + \frac{P_{i,s,t}}{P_{s,t}}\right)^{-\varepsilon_s} s_{s,t}, \tag{29}$$

where the g and s subscripts denote that the variable is related to the goods and service sector, respectively.

Despite the simplicity of the Leontief structure, the demand functions are a bit more complicated than the one of the model with only a goods sector. But the structure is the same. 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. In our extended model, the cost of effort can be in the form of goods or services or both. And this means that the relative price of goods versus services shows up in at least one demand equation and in both when search requires giving up both goods and services, i.e., when both Υ_g and Υ_s are positive. 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 (16), 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.

As discussed in appendix F, the (sector-specific) price indices resemble the one of the previous section, but the search term has become a bit more complicated.

Interaction between the goods and service sector. Section 3 showed that the model is capable of matching key observations stressed in the inventory literature for both monetary policy and TFP shocks. The key question is whether the favorable results for the goods sector are affected by the presence of the service sector. After all, the two sectors are quite different given that unsold goods are stored as inventory (and valued at a variable $\lambda^f_{x,t}$) and unsold services are not.

Equation (23) specified a system with which we could derive the properties of the goods-sector customer-finding rate as a function of two forward-looking variables. Although a bit more complicated (and substantially bigger), there is a similar system for the two-sector model. Appendix F.2 provides the equations of this system and how this system can be used to obtain insights in the properties of the customer-finding rates in the two sectors and specifically it provides a detailed analysis on the interaction between the responses of the two sectors. Since the results are fairly intuitive, we summarize the results here.

In response to a monetary policy shock, the two sectors respond in a similar way. This is not very surprising. A positive monetary policy shock increases the demand for both goods and services which induces buyers to increase search effort leading to a strong upswing in the customer-finding rate in both sectors.

[61] The household problem resembles the one from section 3 and is discussed in appendix F.

The results following a TFP shock are more intriguing. Recall that the previous section showed that key in generating a procylical response for the customer-finding rate is the countercyclical behavior of $\lambda_{x,t}$, the value of an unsold good. Such a mechanism is not present for the goods-sector. As discussed in more detail in appendix F.2 using a graphical analysis, the interaction between the two sectors provides a mechanism for a procyclical customer-finding rate in the service sector. The intuition is as follows. The reduction of $\lambda_{x,t}$ induces goods producers to respond to the increase in productivity with a smaller increase in production. This leads to a relative increase (decrease) of the goods (services) price. This will dampen the increase in the customer-finding rate of the goods-sector (induced by the fall in $\lambda_{x,t}$), but provides upward pressure on the customer-finding rate in the service sector. This – together with increased effort (induced by increased income) – goes against the direct negative effect of increased supply on the customer-finding rate. We find that the customer-finding rate in the service sector can increase or decrease in response to a positive TFP shock, but the responses are always small. So a better way to describe the cyclical response of the customer-finding rate of the service sector in response to TFP shocks is that it is acyclical.

However, there is one exception. A robust procyclical customer-finding rate response is possible in both sectors when TFP of the service sector is less responsive than TFP in the goods sector to an aggregate shock. Now the relative price of goods versus services falls and the associated relative demand shift out of services induces producers of services to let potential output respond by less. A graphical representation is given in appendix F.2. And appendix H.1 documents this possibility by plotting the IRFs for a particular numerical example.

4.2 Parameter estimation and calibration

In this section, we first discuss the five parameters that are most important for the time-series behavior of inventories, production, and sales. Those are estimated with a full-information estimation method. Next, we discuss how other parameter values are chosen using standard calibration.

Estimation of the five key parameters. The most important parameter to match observed inventory, production, and sales facts is the curvature of the matching function, ν_g. Generating a procyclical (countercyclical) customer-finding rate (inventory-sales) ratio comes quite naturally in response to demand shocks. Section 3 showed that this is also possible in response to TFP shocks when the value of an unsold good, $\lambda_{x,t}$, is countercyclical. The magnitude of ν_g is important for two reasons. First, a lower value for ν_g implies a more volatile customer-finding rate.[62] But if the customer-finding rate is too volatile, then sales will be more volatile than output, which is the opposite of what is observed in the data. The exact value of ν_g is also important if we want

[62] Recall that the customer-finding rate, $f^f_{g,t}$, is equal to $\mu_g \theta_{g,t}^{1-\nu_g}$.

the model to replicate observed inventory facts for demand as well as supply shocks which we consider a desirable property given that there is no consensus on the relative importance of the two types of shocks. Although the customer-finding rate is robustly procyclical in our model, it responds more strongly to a demand shock because there is an offsetting effect in the opposite direction in response to a productivity shock due to the change in the quantity of goods brought to the market. As pointed out above, we can control the volatility of the customer-finding rate with the parameter ν_g, but the value that is suitable for monetary policy shocks may not be appropriate for TFP shocks. Similarly, the corresponding parameter for the matching function for the service sector, ν_s, matters for the behavior of the services-sector customer-finding rate. This would affect the price of services relative to goods and could – in principle – affect inventory behavior in the goods sector as the relative price affects the demand for goods.

Our numerical work shows that the responsiveness of monetary policy to the output gap, Γ_y, is important for model properties. This is fairly intuitive.[63] In the richer version of our model with wage stickiness and investment-adjustment costs, the deviations from divine coincidence are no longer negligible. Consequently, a positive productivity shock leads to positive output gap. If $\Gamma_y > 0$, then the positive TFP shock – which we know can generate a positive customer-finding rate by itself – is accompanied by a negative demand force due to the monetary tightening. If Γ_y is large enough, then the negative demand effect could dominate and result in a countercyclical customer-finding rate.

When we want to compare model properties with the observed unconditional moments, then we will have to take a stand on the volatility of the monetary policy and TFP shock, which are controlled by σ_R and σ_A.

To estimate Γ_y, ν_g, ν_s, σ_R, and σ_A, we implement a full-information Bayesian strategy using data for the growth rates of inventories and sales.[64] The posterior modes for these five parameter values are given in the top block of table 3. Details are given in appendix G. 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.

[63] See appendix E for a detailed discussion.

[64] 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.

Table 3: Benchmark calibration and estimation

estimated using inventory and sales data		
curvature search goods: $\nu_g = 0.3469$	estimated	-
curvature search services: $\nu_s = 0.6713$	estimated	-
standard deviation $\varepsilon_{A,t}$: $\sigma_A = 0.0038$	estimated	-
standard deviation $\varepsilon_{R,t}$: $\sigma_R = 0.0036$	estimated	-
$\frac{\partial R}{\partial \text{output gap}}$: $\Gamma_y = 0.0120$	estimated	-
commonly used values	target	
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	
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: $\delta_c = 0.6936$	Cao et al. (2022) and $\overline{c_{\text{durables}}/c}$	O
inventory depreciation: $\delta_x = 0.0040$	$\overline{\Delta x/y}$	O
inventory maintenance: $\eta_x = 0.0694$	Richardson (1995)	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: $\mu_g = 0.5060$	$\overline{f}^f_g$	M
scaling services search friction: $\mu_s = 0.2295$	$\overline{f}^f_s$	M
based on normalization	normalization	
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

Notes. An upper bar indicates the corresponding 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.

Calibration of the remaining parameters. The second block of table 3 contains the parameters for which we use values that are common in the literature.[65]

The third panel contains parameter values that are pinned down by empirical observations using standard calibration arguments. The second column lists 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 Leontief weights, $\omega_{g,c}$, $\omega_{g,s}$, $\omega_{g,i}$, and $\omega_{g,i}$, are pinned down by observed averages for the ratios of purchases of goods relative to purchases of services, 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 Leontief weights.[66]

Our consumption good is assumed to be durable. This makes it possible to match the observed relative volatility of goods and services expenditures despite having a Leontief structure. The value of the depreciation rate, δ_c, is such that it matches the average of the 100% depreciation of non-durable goods and the observed depreciation of durable consumption goods. Cao et al. (2022) estimate the latter to be equal to 16%. Using the observed ratio of durable versus non-durable consumption expenditures, which is equal to 0.68, we obtain a quarterly depreciation rate of 0.6936 for our composite consumption good.

Average 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 (or carrying) costs, η_x, on Richardson (1995), but exclude two of the paper's inventory cost components. We do not include the costs related to the cost of money as this is explicitly captured by the discount rate (marginal rate of substitution) in our model and its endogeneity is key in our inventory-valuation channel. Also, we exclude his depreciation estimate, since we want depreciation to be such that average gross investment in inventories in our model is consistent with national accounting data.[67] 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

[65]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_n, is set to be the same as ε_g and ε_s. This value for ε_n 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 H.2, we consider results when the three Taylor-rule coefficients are based on estimates from Mazelis et al. (2023).

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

[67]As discussed in footnote 35, 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.

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 procyclical customer-finding rate in response to TFP shocks is the countercyclical fluctuation in the value of unsold goods. Quantitatively, this 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 this value.[68]

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).[69] 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.[70] The price- and wage-adjustment-cost parameters, η_P and η_W, are chosen such that a monetary-policy shock leads to a plausible outcome for the aggregate real

[68]One could argue that we should not include taxes, since we abstract from taxes in our model. On the other hand, the numbers in Richardson (1995) imply a higher depreciation rate than the one implied by GDP accounting. Thus, an alternative calibration strategy would be to exclude taxes, but include obsolescence as well as deterioration and pilferage from Richardson (1995). If we would use the midpoint estimates, then we get a value equal to 7.44% for $\eta_x + \delta_x$, which is virtually identical to the value using our benchmark parameter values.

[69]Specifically, the estimated auto-correlation of the annual growth rate of adjusted TFP is equal to 0.112. We are interested in replicating observed business cycle characteristics, but those depend on the properties of the *raw* TFP series. Thus, we should not feed the model the business cycle component of TFP. Annual TFP growth data display a trend. It is very minor, but raises a technical issue. Our model is already more realistic – and thus more complicated – than other business cycle models as we assume that TFP is non-stationary. Introducing non-stationary TFP *growth* would further complicate the analysis. This does not seem worth it, given that the trend is very minor. Instead, we correct for this minor trend in the growth rate using an HP filter with a very high smoothing coefficient, namely $(10{,}000{,}000/4)^4$. The correlation coefficient drops to 0.095 after we have extracted this trend in the growth rate. The value for ρ_A is chosen such that the auto-correlation of the implied annual growth rate of our quarterly TFP series matches this estimate.

[70]For the goods sector, the average customer-finding rate is equal to 0.501 and – as indicated by equation (1) – 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 still be sold at some future date, whereas this is not the case for services. As pointed out in appendix A.2, the length of the data series is very short and clearly not as ideal as what we have for the goods sector. Fortunately, appendix H.4 documents that the value for the target average customer-finding rate turns out to be not important.

economy.[71,72] 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.[73]

The bottom block of table 3 summarizes the calibration of parameters whose values are pinned down by normalizations. Model properties would not be affected if other targets are chosen.

Replacing the estimation with moment matching. One disadvantage of full-information estimation is that the procedure is a (bit of a) black box.[74] This matters 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. Moreover, we would like to investigate whether our model is consistent with key inventory, production, and sales facts for monetary policy *as well as* TFP shocks and whether that can be accomplished using the same values for the structural parameters.[75] But full-information as well as standard method-of-moments estimation procedures would consider all shocks simultaneously.[76] Given these potential drawbacks, we want to make sure that we also consider an alternative procedure to set parameter values. It turns out not to matter for the properties we are interested in, except for the IRF of inventories in response to a monetary policy shock. And this is where choosing parameters by matching moments leads to a better outcome.

In Den Haan and Sun (2024), the extended working-paper version of this paper,

[71] 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%. 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. But, our theoretical responses are consistent with the data. For example, 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. Moreover, standard-error bands are quite large.

[72] 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.

[73] As discussed in appendix E, setting $\eta_i > 0$ is necessary to ensure that the initial investment response to a TFP shock is not negative.

[74] Specifically, full-information methods like Maximum Likelihood – or the Bayesian version that we use – find parameters to minimize residuals. But how particular parameter values affect the values of those residuals is often not intuitive and the estimated values may be heavily affected by outliers. By contrast, the mapping between parameter values and model moments is typically more intuitive.

[75] 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.

[76] Also, full-information methods are quite ambitious, since they require that the model is correctly specified which – of course – is not the case. Den Haan and Drechsel (2021) show that even slightly misspecified empirical models can lead to large biases in parameter estimates, which in turn are associated with biased predictions of the theoretical model evaluated using estimated parameters.

we discuss an elaborate procedure to obtain a *range* of parameter values that are consistent with key inventory facts for both types of shocks, taking into account sampling uncertainty.[77] As an alternative to the parameters estimated with the Bayesian full-information procedure, we also document model properties for the preferred combination of parameter values in this "admissible range" from Den Haan and Sun (2024).

The motivation for that alternative set of parameter values is the following. A key advantage of our approach to model inventories is that a TFP shock can also generate a procyclical customer-finding rate, $f_{g,t}^{f}$, because $\lambda_{x,t}$ is countercyclical. However, if this is accompanied by a tightening, i.e., $\Gamma_y > 0$, then the procyclicality would be dampened and would even be overturned if Γ_y is large enough. The estimated correlation coefficient of $f_{g,t}^{f}$ and output is significantly positive, and using the 95% confidence interval means that the model-generated correlation should be at least 0.335. This implies an upper bound for Γ_y of around 0.06 and this bound is fairly insensitive to changes in other parameter values. The natural lower bound for Γ_y is equal to zero. For our alternative parameter set, we use the middle value, that is, 0.03. The curvature of the search friction, ν_g, and the relative importance of the two shocks, i.e., ${}^{\sigma_R}/_{\sigma_A}$, are chosen to ensure the following. First, output should be more volatile than sales in response to monetary policy shocks. If this is true, then this desirably model property is also true for TFP shocks because the customer-finding rate is less responsive to TFP shocks. This will ensure that our model can generate this key relative-volatility property from the inventory literature for both types of shocks. Second, we want parameters to be such that the model with both shocks to generate a value for ${}^{\sigma_{y_g}}/_{\sigma_{s_g}}$ in the estimated 95% confidence interval.[78] Den Haan and Sun (2024) show that these these two conditions are satisfied for values of ν_g (and ν_s) in between 0.5006 and 0.6574. To turn a range of admissible parameter values into a specific number, we add the following restriction. Third, we want the model to generate a value for the correlation between the customer-finding rate and beginning-of-period inventories that is in the estimated 95% confidence interval. This moment is highly useful in identifying the relative role of monetary-policy and TFP shocks because it is negative for TFP shocks and positive for monetary-policy shocks. Imposing these three restrictions implies $\nu_g = 0.565$ and ${}^{\sigma_R}/_{\sigma_A} = 0.5921$.[79] If a lower value for ν_g would be chosen, then the gap between the volatility of output and sales would no longer be sufficiently large in the economy with monetary policy shocks. At higher values for ν_g, the tension between the second and

[77] That is, it differs from standard method-of-moment estimation in that it does not provide particular parameter estimates that generate the best match with observed values of unconditional target moments, but provides a range of values that are consistent with these targets for both types of shocks taking into account confidence bands. This way we learn that a nontrivial range of values for ν_g are consistent with observed facts.

[78] But because of the first requirement, this does not happen because a high value for ${}^{\sigma_{y_g}}/_{\sigma_{s_g}}$ in response to TFP shocks compensates for ${}^{\sigma_{y_g}}/_{\sigma_{s_g}}$ being less than one in response to monetary-policy shocks.

[79] Strictly speaking a range of values around these numbers is admissible, but the range is very narrow. In this exercise, we set $\nu_s = \nu_g$, but in appendix H.5 we show robustness of our results to alternatives.

third restriction can no longer be resolved. See Den Haan and Sun (2024) for a more detailed discussion.

4.3 Predictions of the full model with goods and services

In this section, we compare model properties with their empirical counterparts.

4.3.1 Impulse response functions

Figures 2 and 3 plot IRFs in response to a monetary policy and a TFP shock, respectively.[80] The size of the initial shock is set to generate a peak 1 percent increase in GDP for both types of shocks and for both parameter sets.

The two parameter sets considered are quite different. 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 explicit moment matching. Nevertheless, the IRFs display a very similar shape and even the magnitudes of the responses are often similar. A key outcome is that the customer-finding rate is procyclical, which in turn implies that the inventory-sales ratio is countercyclical. And this is true for both types of shocks.

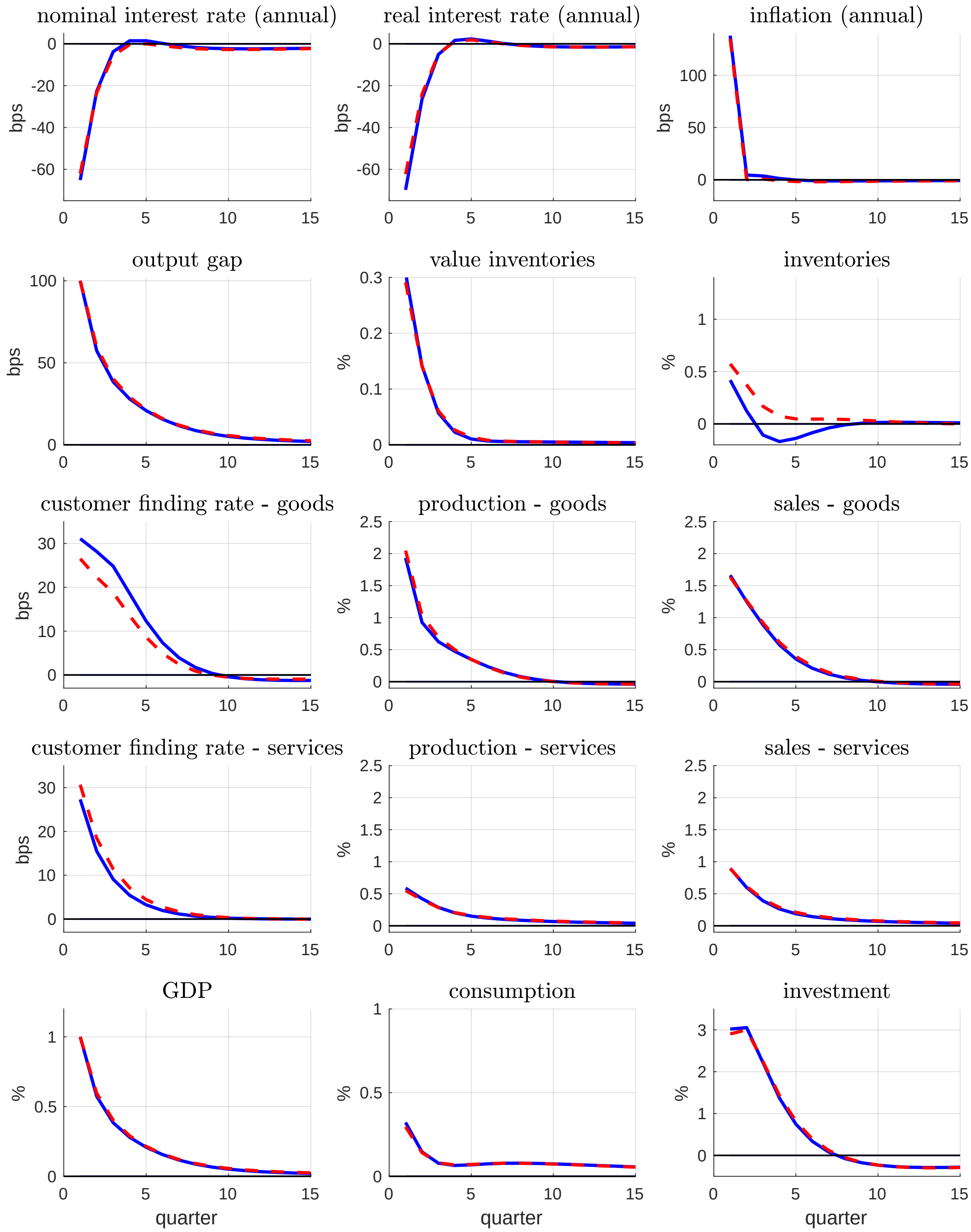
The difference in the value for ν_g does matter for some model properties. There is a minor and a more substantial difference. The minor difference is the following. When using the higher value obtained by matching moments, we find that the output response exceeds the sales response at all horizons for both types of shocks. When using the estimated lower value for ν_g, we find that this is also true for a TFP shock and initially for a monetary policy shock. For the latter, however, the sales response is slightly higher in subsequent periods.

The main difference between the two parameter sets is the response of the inventory stock following a monetary-policy shock. It quickly turns negative when estimated parameter values are used whereas the response remains uniformly positive for the parameter values based on matching moments.

Both the smaller and the bigger difference can be explained by the fact that a lower value of ν_g increases the volatility of the customer-finding rate. A stronger response of the customer-finding rate makes it more likely that sales responds stronger than output and that inventories move in the opposite direction. The moment-matching procedure has, of course, an advantage in matching the observed procyclicality of inventories as it was designed to ensure that $\sigma_{y_g} > \sigma_{s_g}$ for *both* types of shocks.

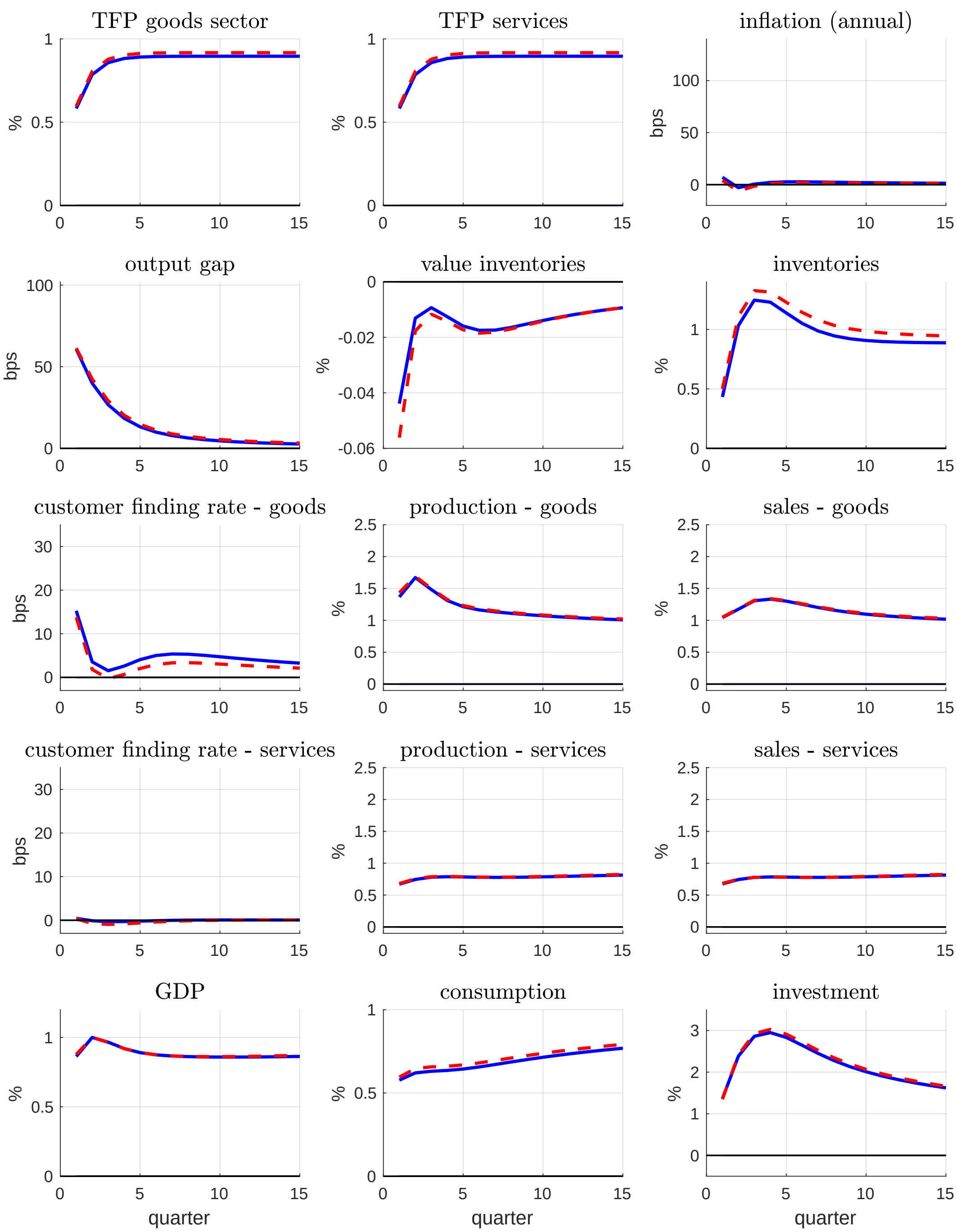
[80] Model properties are based on a first-order perturbation numerical approximation.

Figure 2: Monetary-policy shock; benchmark parameters



Notes. The blue/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 red/dashed lines correspond to the case when $\nu_g = \nu_s = 0.565$ and $\Gamma_y = 0.03$, which are the calibrated values. Shock size is calibrated to ensure a peak 1 percent increase in GDP in both cases.

Figure 3: TFP shock; benchmark parameters



Notes. The blue/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 red/dashed lines correspond to the case when $\nu_g = \nu_s = 0.565$ and $\Gamma_y = 0.03$, which are the calibrated values. Shock size is calibrated to ensure a peak 1 percent increase in GDP in both cases.

Predicted model responses for services. The reasons the model can match key observed inventory, production, and sales facts are really the same as the ones given in section 3 for the economy with 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 very limited empirical data on sell frictions in the service sector.[81]

A monetary-policy shock stimulates demand which in turn directly increases the customer-finding rate in both sectors. 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. There is no such dampening effect in the service sector since the value of "unsold" services is fixed (and equal to zero). The goods-sector customer-finding rate increase is more persistent. 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.[82]

As discussed in section 4.1 and appendix F.2, 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 can flip even for relatively small changes in parameter values. 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 H.1, we show that the customer-finding rate in the service sector can display a robust sizable procyclical response *if* productivity in the service sector lags the increase in the goods sector, consistent with the theoretical analysis of the discussion in appendix F.2.[83]

Relative volatility of usual expenditure components. The IRFs also document that the model generates the usual relative volatility for GDP, consumption, and investment in response to both types of shocks. Moreover, sales of goods are more volatile than their counterparts of the service sector.[84] 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 more volatile than consumption of services

[81]When considering the results, it is important to take into consideration that both the estimation and the moment-matching exercise are based on available data from the goods sector only. Whereas the inventory-sales ratio provides direct evidence on the ability to sell available goods, that is, not possible for the service sector as the level of "potential" sales is not measured.

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

[83]It does not seem unreasonable that productivity in the service sector does not respond one-for-one with productivity in the goods sector. To ensure balanced growth, however, we have to impose that the long-run responses in the two sectors are equal. The experiment in which service-sector TFP responds with a lag gives a plausible prediction of what would happen if service-sector TFP is less volatile. The reason is that responses in the customer-finding rate are short-lived for both types of shocks.

[84]The same is true for output levels using the production function to determine output in both sectors. But note that output is potential output in the service sector and actual output in the goods sector because unsold goods are not lost, but end up in inventories.

because they are partially durable and the Leontief structure links consumption services to the *stock* of consumption goods.

4.3.2 Comparing model moments with observed counterparts

The key business cycle facts regarding inventory-related data are that the cyclical component of output is more volatile than the cyclical component of sales and that the cyclical components of the customer-finding rate, the level of inventories, and investment in inventories are all procyclical. Here we address the question whether the model can match these observations and do so not only in an economy with both types of shocks, but also in economies with only one type of shock.[85] The conclusion is that the model is quite successful, even quantitatively. Table 1 presents the values of key moments together with their empirical counterparts. 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.[86]

Comovement statistics. Table 1 documents that the cyclical components of the customer-finding rate, the inventory level, and investment in inventories are all procyclical and that is true for both parameter sets and for both types of shocks. The latter implies that it is also true for economies with both types of shocks. Since the inventory-sales ratio is an inverse function of the customer-finding rate, this implies that the customer-finding rate is countercyclical. These results are quite intuitive given the responses documented by the IRFs, except perhaps for the positive correlation of the investment in inventories with output in an economy with only monetary policy shocks. Following a monetary policy shock, the IRF for investment displays a large

[85] We use the Hodrick-Prescott filter to extract business cycle frequencies.

[86] 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). The two approaches lead to the same answer for first-order moments. For higher-order moments, however, the average of a statistic across short-sample replications does not have to be equal to the population moment. For example, if a variable is very persistent, then the average of a set of variances calculated using short samples will be lower than the unconditional variance, since the means over the shorter samples adjust which reduces the variance. This actually turns out to be of minor importance for our model. But, the Kydland-Prescott approach has an additional advantage. Even if the underlying model is the true data-generating process, then the outcomes for a statistic of interest could still vary substantially across model 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. As documented in the table, some model moments display very little variation across Monte Carlo replications, whereas others do. Thus, a proper evaluation of the model takes into account both the standard errors of the estimated moment, and – following Kydland and Prescott (1982) – the standard deviations across replications.

increase followed by subsequent gradual decreases. The former corresponds to an increase in the investment in inventories and the latter with decreases. However, the initial increase dominates the subsequent gradual decreases. 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 0.558 when the moment-matching parameter set is used. Consistent with the IRFs, it is smaller and only equal to 0.294 when the estimated posterior modes are used.

The magnitude of a correlation coefficient is typically not that insightful in economies with only one type of shock.[87] And as pointed out above, the model gets the sign of the three correlation coefficients of interest right for both parameter sets and for the economy with only monetary policy and for the one with only TFP shocks. In terms of the sign of comovement, there is one intriguing outcome and that is related to the sign of the correlation between the customer-finding rate with the beginning-of-period inventory stock, $\text{COR}(f_g^f, x_{-1})$. It is positive for monetary-policy shocks and negative for TFP shocks. And this is a very robust result that we found to hold for a wide range of parameter value values.[88] Since monetary and TFP shocks have different implications for this particular moment, it is useful in identifying the role of the two types of shocks which is exploited when key parameters are set using explicit moment matching.

Next, we turn to the question whether the model with both shocks can quantitatively match observed cyclicality for the three key inventory variables taking into account sampling uncertainty. For this exercise, it is not only the values of ν_g, ν_s, and Γ_y that matter, but also the relative importance of monetary-policy and TFP shocks, i.e., the value of ${}^{\sigma_R}/_{\sigma_A}$. This ratio is pinned down by our full-information estimation procedure. Using the mode of the estimated parameters, the model generates values for the three correlation coefficients with output that are inside the 95% confidence intervals.[89] At these model parameter values, however, the model's prediction for $\text{COR}(f_g^f, x_{-1})$ is positive (equal to 0.071) which is outside the 95% confidence region of its estimated counterpart which is quite wide, but contains only negative values.[90]

[87]Unless the dynamic responses to the shock are quite different, then then the correlation coefficient would be close to minus or plus 1.

[88]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 returns 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. By contrast, the responses of the two series are both temporary following a monetary-policy shock.

[89]This simply focuses on the average of the generated correlated coefficients across Monte Carlo replications. Of course, we stand even stronger if we take into account the standard deviation observed across model replications.

[90]Recall that this is satisfied by construction when parameters are set using the explicit moment-matching exercise. However, we are quite strict in evaluating the results based on the full-information estimation procedure. The reason is that this is an example, where taking the variation for generated moments across model replications does make a difference as the standard deviation of $\text{COR}(f_g^f, x_{-1})$ across Monte Carlo replications is quite high.

Our moment-matching procedure determines parameters simultaneously, but the key statistic affecting the relative importance of the two shocks is $\text{COR}(f_g^f, x_{-1})$. So the calibration procedure does get this number right and does so even if we ignore variability across Monte Carlo model replications. In terms of the three main correlation coefficients, however, model performance is a bit worse for the moment-matching parameter set. The correlation of the customer-finding rate with output is a good match with its empirical counterpart. The same is true for the correlation of the level of inventories with output *if* we take into account the sampling variation across Monte Carlo replications and almost fine if we do not.[91] With this set of parameter values, however, the model does predict a value for the correlation of *investment* in inventories that is too high. Namely, the average across model replications is equal to 0.839 whereas the upper bound of the 95% confidence interval is equal to 0.646. As documented in table 1, this is also true in the two economies with only one shock. An easy way to get this moment closer to its empirical counterpart is to lower the value of ν_g, i.e., set it closer to the estimated value. But this would worsen the ability of the model to generate the property that output is more volatile than sales in response to monetary policy shocks.

Relatve volatility of output versus sales. A key inventory fact that has received a lot of attention in the inventory literature is that output is more volatile than sales. Here we address the question whether the quantitative predictions of the model are consistent with observed values of this ratio at business cycle frequencies.

As shown in table 1, the value of $\sigma_{y_g}/\sigma_{s_g}$ for the monetary-policy-shock economy is equal to 1.080 for the parameter values based on moment matching. This value is inside the 95% confidence band. By contrast, the value is only equal to 1.005 when estimated parameter values are used, which is substantially below the lower bound of the 95% confidence interval which is equal to 1.073.[92] This does not mean that the model with the estimated 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 the model with only TFP shocks, the implied value for $\sigma_{y_g}/\sigma_{s_g}$ is somewhat above the upper bound of the 95% confidence interval. This is true for both parameter sets considered. Again, this does not mean that the model is not consistent with this statistic. It just means that some monetary-policy shocks are needed for the model to match the observed unconditional value.

Thus, the model can generate a value for $\sigma_{y_g}/\sigma_{s_g}$ that is inside the 95% confidence interval for both sets of parameter values when fluctuations are driven by both types of shocks.

[91] The upper bound of the estimated confidence interval is equal to 0.816 and the average across Monte Carlo replications is equal to 0.839. Thus, even if we do not take into account variation across Monte Carlo replications the model prediction is not that far outside the 95% confidence interval.

[92] Also, taking into account the standard deviation across Monte Carlo replications does not help because its value, 0.019, is quite small.

4.3.3 Summary of model properties and alternative specifications

The model has only two types of shocks, we do not rely on measurement error, and it has few bells and whistles, 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 quite successful in key dimensions, not only qualitatively, but also quantitatively.

In the appendix, we discuss alternative specifications of our model. In appendix H.1, we focus on 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. Section 3 made clear that a non-stationary process is helpful in generating a robust procylical discount rate response, which implies a countercyclical value for the value of an unsold good, which in turn is key in getting a procyclical customer-finding rate in response to TFP shocks. But a stationary TFP process is quite popular in the business-cycle literature.[93] We also consider the case when TFP in the service sector lags TFP in the goods sector.

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 H.2.

In our benchmark specifications, we assume that the relative importance of goods and services for search costs is equal to the observed ratio of the corresponding expenditures. But it does not seem implausible that services are more important for acquiring purchases. In appendix H.3, we show that our results are robust when services are more important. In fact, model outcomes are very similar to our benchmark results 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 H.4, we consider the case when the means are the same across the two sectors. In appendix H.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 H.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 really matters is the persistence of the TFP process. When deviations in

[93] 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). This paper 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_x^f. 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_x^f. As documented in Bansal and Yaron (2004), this is important for asset prices.

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 observations regarding the behavior of inventory, production, and sales data; both when fluctuations are driven by monetary policy and when they are driven by 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.

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.[94] 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.

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 H.1, however, we show that stronger interaction effects are present 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.

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

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.S., 1981. Retail inventory behavior and business fluctuations. Brookings Papers on Economic Activity 12, 443–520.

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.

Chen, L., 2017. Asymmetric inventory dynamics and product market search. Unpublished manuscript, Duke University.

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.

Deaton, A., Laroque, G., 1992. On the behaviour of commodity prices. Review of Economic Studies 59, 1–23.

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.

Den Haan, W.J., Sun, T., 2024. The role of sell frictions for inventories and business cycles. CEPR Discussion Paper DP19422.

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).

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.

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.

de Sousa Rodrigues, D., Willems, T., 2024. Inventories matter for the transmission of monetary policy: Uncovering the cost-of-carry channel. Unpublished manuscript.

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 and some more business cycle facts

Data are from the Bureau of Economic Analysis (BEA).[95]

- GDP and its components are from table 1.1.6: Real Gross Domestic Product; chained dollars, seasonally adjusted at annual rates.

[95] 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.

- Data series for wholesale trade inventories, retail trade inventories, and final sales of goods and structures of domestic business are from tables 5.8.6A & 5.8.6B: Real Private Inventories and Real Domestic Final Sales by Industry; chained dollars, seasonally adjusted at annual rates.

- Data series for finished-goods inventories for the manufacturing sector are from tables 4AU3 & 4BU3: Real Manufacturing Inventories, by Stage of Fabrication (Finished goods); chained dollars, seasonally adjusted.

Consistent production series. A key stylized fact in the inventory literature is the volatility of sales relative to the volatility of output. Therefore, we need production series for the exact same universe of firms that these inventory and sales data are based on.[96] A production series can be easily constructed using the equation: $y_{g,t} = s_{g,t} - (1-\delta_x)x_{t-1}$ for a given value of the depreciation rate, δ_x. A natural choice is to use 0.4% which is consistent with average investment in inventories as a fraction of GDP.[97]

Additional business-cycle facts. Table 4 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.

A.2 Customer-finding-rate measure for the service sector.

As shown in equation (1), we can construct a measure for the customer-finding rate for the goods sector using the observed inventory-sales ratio. That is not an option for the service sector. 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? ⋯%."*

The answers are used to construct a capacity-utilization measure.[98] Figure 4 dis-

[96] As discussed in section 4.3, 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. It will not be possible for the model to match the properties of both statistics if the empirical production data used are based on a different universe of firms with different volatility.

[97] Kryvtsov and Midrigan (2013) use a value equal to zero, but this value is not that different from 0.4% and such changes in the depreciation rate only have a minor effect on the statistics reported.

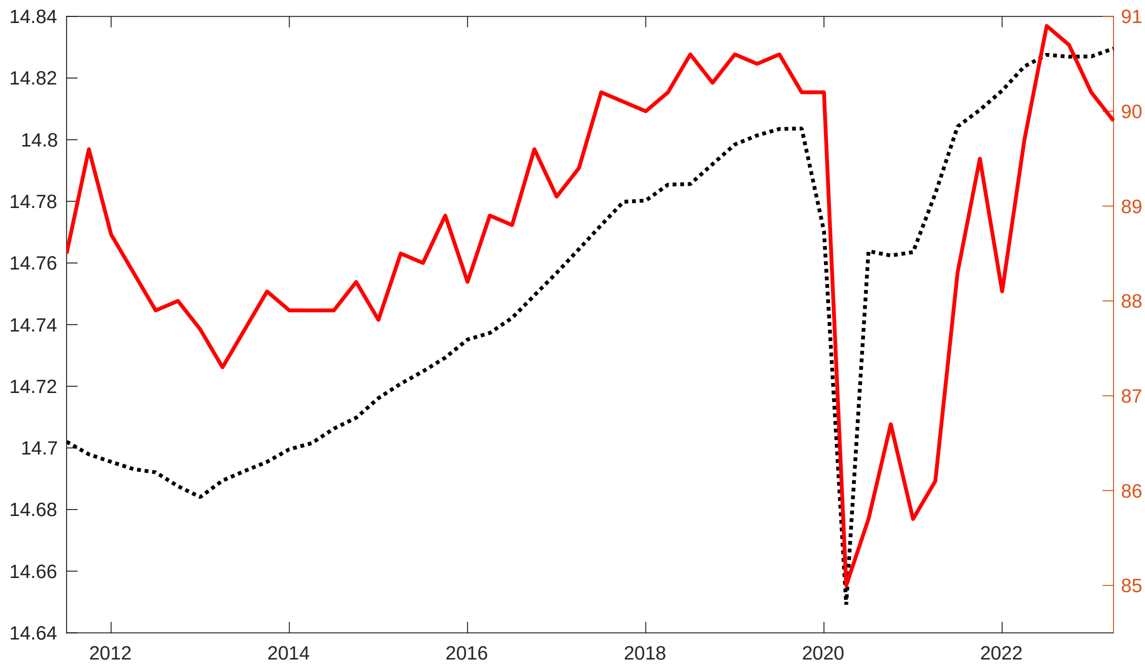
[98] Let x denote the reported percentage with which the firm's volume of activity could be increased. Then, the capacity utilization rate is defined in percentage as $^{100}/_{(1+x/100)}$. Thus, if the firm reports that x is equal to 0, then capacity utilization is equal to 100%. And if the $x = 100$, which means that

Table 4: 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.

Figure 4: 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 (red/solid and scale on right axis) and the log of real GDP for the Euro Area (black/dashed and scale on right axis).

plays 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.

activity could be doubled, then capacity utilization is equal to 50%.

A.3 Variance Decomposition.

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

$$\mathbb{V}\text{ar}[X_t] = \mathbb{V}\text{ar}[X_{t,1}] + 2\mathbb{C}\text{ov}[X_{t,1}, X_{t,2}] + \mathbb{V}\text{ar}[X_{t,2}] = \mathbb{C}\text{ov}[X_t, X_{t,1}] + \mathbb{C}\text{ov}[X_t, X_{t,2}]. \quad (30)$$

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 used to decompose the fluctuations in total finished-goods inventories in the three sectoral components and to determine the quantitative importance of investment in inventories for GDP fluctuations.

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

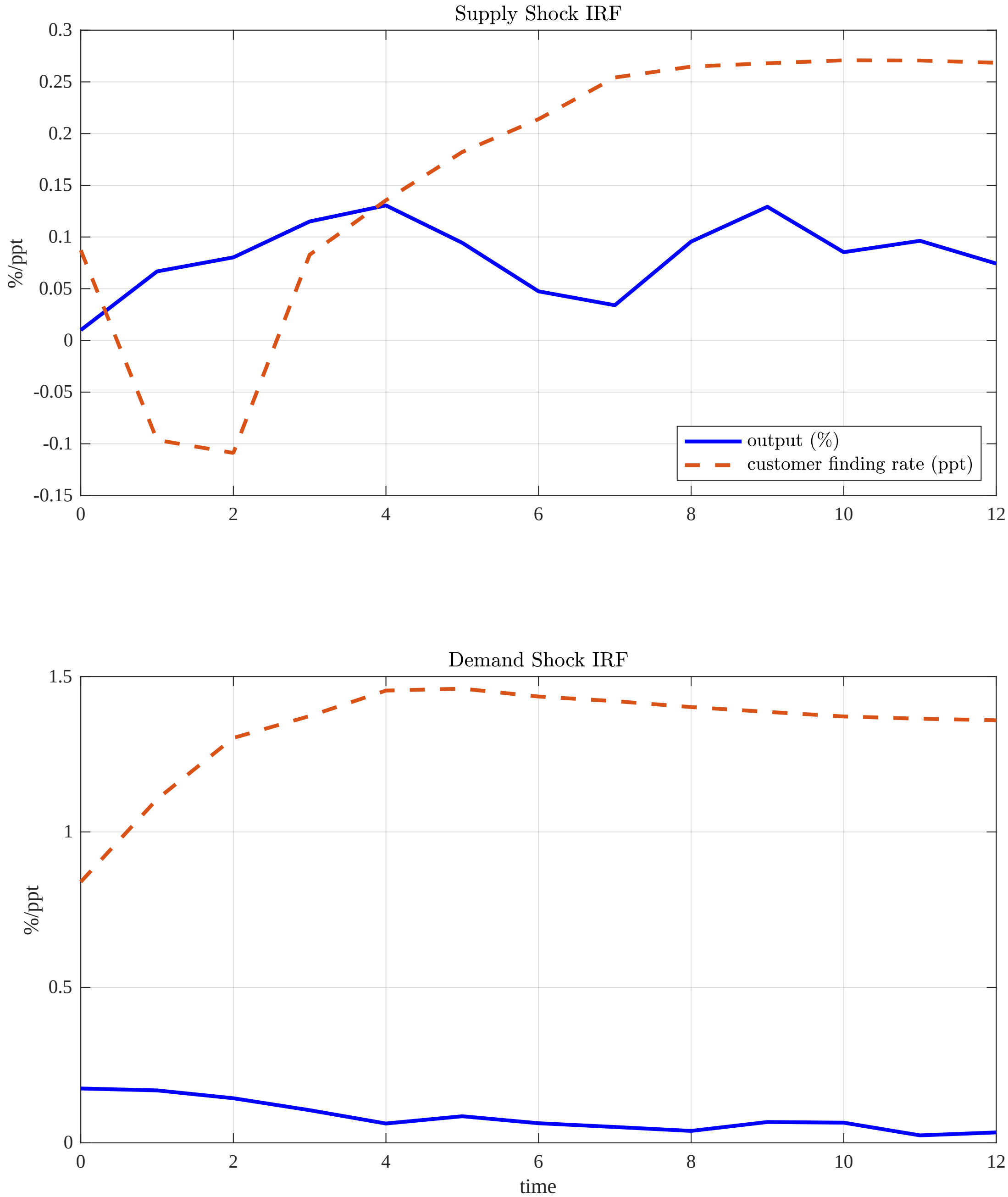
In section 2, it is shown that the correlation between the customer-finding rate and aggregate activity is positive. This means that there is positive comovement when averaged across *all* shocks and leaves open the possibility that there is a negative comovement in response to some types of shocks. 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 believed to have for business-cycle fluctuations, it would be helpful to know whether TFP driven fluctuations also imply a procyclical customer-finding rate.

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 5 plots the estimated 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, we find 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.

Figure 5: Customer-finding-rate IRFs: 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.

B 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

$$\max_{c_i,s_i,e_i} \ln(c_i)$$
$$\text{s.t.}$$
$$p_i s_i = \omega - \eta e_i, \tag{31a}$$
$$c_i = s_i, \tag{31b}$$
$$s_i = f_i^b e_i, \tag{31c}$$

where e_i stands for effort, p_i for the price of the good, c_i for consumption, s_i for sales, and $1/f_i^b$ the amount of effort needed to obtain 1 unit of good i. For simplicity, we assume that there is only one good, but we use the i subscript to be consistent with the model in the main text. Resources of the consumer are a fixed endowment, ω, but those are diminished if more effort is put into acquiring goods.[99]

Substituting out c_i and s_i, we get

$$\max_{e_i} \ln\left(f_i^b e_i\right)$$
$$\text{s.t.}$$
$$p_i f_i^b e_i = \omega - \eta e_i. \tag{32a}$$

The first-order conditions are the constraint and

$$\frac{1}{e_i} = (p_i f_i^b + \eta)\lambda. \tag{33}$$

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

$$s_i = \frac{\omega}{p_i + \frac{\eta}{f_i^b}}, \tag{34}$$

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

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

The firm problem is given by

$$\max_{s_i, y_i, p_i, \theta_i} p_i s_i - \alpha y^2$$

s.t.

$$s_i = \frac{\omega}{p_i + \frac{\eta}{f^b(\theta_i)}}, \tag{35a}$$

$$s_i = f^f(\theta_i) y_i, \tag{35b}$$

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

$$f_i^b = f^b(\theta_i) = \mu \left(\frac{e_i}{y_i}\right)^{-\nu}, \tag{36a}$$

$$f_i^f = f^f(\theta_i) = \mu \left(\frac{e_i}{y_i}\right)^{1-\nu}. \tag{36b}$$

One can obtain c_i from $c_i = s_i$ and by combining the household first-order condition with the one remaining constraint, equation (32a), one gets that $\lambda = 1/\omega$.

C Firm first-order conditions for the goods-only economy

The firm's first-order conditions are given by the following set of equations.

$$s_{i,t} \leq \left(\frac{\xi_e}{f^b(\theta_{i,t})} + \frac{P_{i,t}}{P_t}\right)^{-\varepsilon} s_t, \tag{37a}$$

$$s_{i,t} \leq f^f(\theta_{i,t})(y_{i,t} + (1-\delta_x)x_{i,t-1}), \tag{37b}$$

$$x_{i,t} \leq \left(1 - f^f(\theta_{i,t})\right)\left(y_{i,t} + (1-\delta_x)x_{i,t-1}\right), \tag{37c}$$

$$MC_t = \left(\frac{w_t}{\alpha}\right)^{\alpha}\left(\frac{r_{k,t}}{1-\alpha}\right)^{1-\alpha}\frac{1}{A_t} = \frac{w_t n_t}{\alpha y_t} \tag{37d}$$

$$MC_t = f^f(\theta_{i,t})\lambda^f_{i,s,t} + (1 - f^f(\theta_{i,t}))\lambda^f_{i,x,t}, \tag{37e}$$

$$\lambda^f_{i,s,t} = \frac{P_{i,t}}{P_t} - \lambda^f_{i,d,t}, \tag{37f}$$

$$\lambda_{i,x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\begin{array}{c} f^f(\theta_{i,t+1})\lambda^f_{i,s,t+1} \\ +(1 - f^f(\theta_{i,t+1}))\lambda^f_{i,x,t+1}\end{array}\right)\right], \tag{37g}$$

$$\begin{aligned}&\left(\frac{\partial f^f(\theta_{i,t})}{\partial \theta_{i,t}}\right)\left(y_{i,t} + (1-\delta_x)x_{i,t-1}\right)\left(\lambda^f_{i,s,t} - \lambda^f_{i,x,t}\right) \\ &= -\lambda^f_{i,d,t}\varepsilon\left(\frac{\xi_e s_t}{f^b(\theta_{i,t})^2}\right)\left(\frac{\partial f^b(\theta_{i,t})}{\partial \theta_{i,t}}\right)\left(\frac{\xi_e}{f^b(\theta_{i,t})} + \frac{P_{i,t}}{P_t}\right)^{-\varepsilon-1},\end{aligned} \tag{37h}$$

$$\begin{aligned}&s_{i,t} - \varepsilon\lambda^f_{i,d,t}\left(\frac{\xi_e}{f^b(\theta_{i,t})} + \frac{P_{i,t}}{P_t}\right)^{-\varepsilon-1} s_t = \eta_P\left(\frac{P_{i,t}}{P_{i,t-1}} - 1\right)\left(\frac{P_t}{P_{i,t-1}}\right)s_t \\ &+ \beta\eta_P\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\frac{P_{i,t+1}}{P_{i,t}} - 1\right)\left(\frac{P_{i,t+1}P_t}{P^2_{i,t}}\right)s_{t+1}\right].\end{aligned} \tag{37i}$$

The first three equations are the demand constraint, the sales constraint, and the inventories accumulation constraint and all three will be binding. The three associated Lagrange multipliers are denoted by $\lambda^f_{i,d,t}$, $\lambda^f_{i,s,t}$, and $\lambda^f_{i,x,t}$, respectively.[100] Equation (37d) is the usual expression that relates the marginal cost of producing an extra unit to the wage rate, the rental cost of capital, and the productivity level. Equation (37e) 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 (37f) 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 (37g) 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 (37h) presents the tradeoff when changing tightness. If the

[100] In equilibrium, the sales constraint (37b) is identical to equation (8d) 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 $f^f_{i,t}$ by choosing tightness.

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 (37i) is the first-order condition related to $P_{i,t}$. This equation provides us with our modified New-Keynesian Phillips Curve, which is discussed in detail in the main text.

In the main text, we focused on the following sub-set of equations:

$$1 = \frac{\xi_e}{f^b(\theta_t)} + \frac{P_{i,t}}{P_t}, \tag{38a}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = f^f(\theta_t)\left(\frac{P_{i,t}}{P_t} - \lambda^f_{d,t} - \lambda^f_{x,t}\right), \tag{38b}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = \varepsilon\lambda^f_{d,t}\frac{\nu}{1-\nu}\xi_e\theta_t, \tag{38c}$$

$$\lambda^f_{x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\begin{array}{c} f^f(\theta_{t+1})\lambda^f_{s,t+1} \\ +(1-f^f(\theta_{t+1}))\lambda^f_{x,t+1}\end{array}\right)\right], \tag{38d}$$

$$1 - \varepsilon\lambda^f_{d,t} = \eta_P\frac{P_t}{P_{i,t}}\left(\begin{array}{c}\left(\frac{P_{i,t}}{P_{i,t-1}} - 1\right)\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \\ -\beta\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\frac{P_{i,t+1}}{P_{i,t}} - 1\right)\left(\frac{P_{i,t+1}}{P_{i,t}}\right)\frac{s_{t+1}}{s_t}\right]\end{array}\right). \tag{38e}$$

This system can be derived from the first-order conditions as follows. From the first-order condition for the output level, equation (37e), we get that

$$MC_t - \lambda^f_{x,t} = f^f(\theta_t)(\lambda^f_{s,t} - \lambda^f_{x,t}). \tag{39}$$

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 (37f), 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 (39) gives equation (38b).

Equation (38c) is a rewritten version of the firm's first-order condition for θ_t where we have also used equation (39). Equation (38d) is identical to equation (37g). If we combine equations (13), (37e), (37f), and (37i), then we get equation (38e).

D 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, $P_{i,t}/P_t$,[101] 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 $\lambda^f_{d,t}$.

$$1 = \frac{\xi_e}{f^b(\theta_t)} + \frac{P_{i,t}}{P_t} \tag{40a}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = \left(\frac{P_{i,t}}{P_t} - \lambda^f_{d,t} - \lambda^f_{x,t}\right) f^f(\theta_t) \tag{40b}$$

$$\left(MC_t - \lambda^f_{x,t}\right) = \varepsilon\lambda^f_{d,t}\frac{\nu}{1-\nu}\xi_e\theta_t \tag{40c}$$

$$\lambda^f_{x,t} = \beta(1-\delta_x)\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\begin{array}{c} f^f(\theta_{t+1})\lambda^f_{s,t+1} \\ +(1-f^f(\theta_{t+1}))\lambda^f_{x,t+1}\end{array}\right)\right], \tag{40d}$$

$$1-\varepsilon\lambda^f_{d,t} = \eta_P\frac{P_t}{P_{i,t}}\left(\begin{array}{c}\left(\frac{P_{i,t}}{P_{i,t-1}}-1\right)\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \\ -\beta\mathbb{E}_t\left[\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(\frac{P_{i,t+1}}{P_{i,t}}-1\right)\left(\frac{P_{i,t+1}}{P_{i,t}}\right)\frac{s_{t+1}}{s_t}\right]\end{array}\right), \tag{40e}$$

Recall from equation (1) 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 $\frac{\partial f^f(\theta_t)}{\partial\lambda^f_{d,t}} < 0$. *That is, an increase in inflationary pressure (relative to expected future inflation), which leads to a decrease in* $\lambda^f_{d,t}$ *according to equation (40e), is associated with an increase in the customer-finding rate.*

Proposition 2 $\frac{\partial f^f(\theta_t)}{\partial\lambda^f_{x,t}} < 0$. *That is, an increase in the value of carrying an unsold good into the future as inventory is associated with a reduction in the customer-finding rate.*

Proofs. Using equations (40a) and (40b) we can substitute out $\frac{P_{i,t}}{P_t}$ and MC_t and rewrite equation (40c) as

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

[101] 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.

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

$$1 = \lambda_{x,t}^f + \lambda_{d,t}^f + \xi_e \left(1 + \varepsilon \lambda_{d,t}^f \frac{\nu}{1-\nu}\right) \frac{\theta_t^{\nu}}{\mu}, \tag{41}$$

which shows that θ_t is an implicit function of $\lambda_{x,t}^f$ and $\lambda_{d,t}^f$. Rewriting and taking the relevant partial derivative immediately gives the desired results, namely that $\frac{\partial \theta_t}{\partial \lambda_{x,t}^f} < 0$, and $\frac{\partial \theta_t}{\partial \lambda_{d,t}^f} < 0$. 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 $\lambda_{x,t}^f$ and $\lambda_{d,t}^f$. ■

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

Proof. We first prove that MC_t increases with $\lambda_{x,t}^f$ locally around the steady state. Using equations (40a) and (40c) to substitute out $\frac{P_{i,t}}{P_t}$ and $\lambda_{x,t}^f$ in equation (40b), we get

$$MC_t = \left(1 - \lambda_{d,t}^f\right) + \xi_e \left(\varepsilon \lambda_{d,t}^f \frac{\nu}{1-\nu}\theta_t - \left(1 + \varepsilon \lambda_{d,t}^f \frac{\nu}{1-\nu}\right) \frac{\theta_t^{\nu}}{\mu}\right). \tag{42}$$

At the steady state, $\lambda_d^f = \frac{1}{\varepsilon}$ 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

$$\left.\frac{\partial MC(\theta_t)}{\partial \theta_t}\right|_{\theta_t = \theta_{\text{ss}}} = \frac{\nu}{1-\nu} - \left(1 + \frac{\nu}{1-\nu}\right) \frac{\nu \theta_{\text{ss}}^{\nu-1}}{\mu} = \frac{\nu}{1-\nu}\left(1 - \frac{1}{\mu \theta_{\text{ss}}^{1-\nu}}\right) < 0. \tag{43}$$

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

Next, we prove that MC_t decreases with $\lambda_{d,t}^f$ locally around the steady state. Recall that equation (41) shows that θ_t is an implicit function of $\lambda_{x,t}^f$ and $\lambda_{d,t}^f$. Holding $\lambda_{x,t}^f$ constant and differentiating equation (41) locally around the steady state gives

$$\left.\frac{d\theta_t}{d\lambda_{d,t}^f}\right|_{\theta_t = \theta_{\text{ss}}} = \frac{\theta_{\text{ss}}\varepsilon}{\nu} \left(\frac{-\xi_e \left(1 + \frac{\nu}{1-\nu}\right)}{\left(\xi_e \left(\frac{\nu}{1-\nu}\right) + \frac{\mu \theta_{\text{ss}}^{-\nu}}{\varepsilon}\right)}\right)^{-1} < 0. \tag{44}$$

In addition, holding $\lambda_{x,t}^f$ constant and differentiating equation (40c) locally around the steady state gives

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

$$= \frac{\xi_e\nu}{1-\nu}\left(1+\frac{-\xi_e\left(1+\frac{\nu}{1-\nu}\right)}{\left(\xi_e\left(\frac{\nu}{1-\nu}\right)+\frac{\mu\theta_{\rm ss}^{-\nu}}{\varepsilon}\right)}\nu\right), \quad (45)$$

which shows a positive relationship between MC_t and θ_t because

$$1+\frac{-\xi_e\left(1+\frac{\nu}{1-\nu}\right)}{\left(\xi_e\left(\frac{\nu}{1-\nu}\right)+\frac{\mu\theta_{\rm ss}^{-\nu}}{\varepsilon}\right)}\nu = \frac{\frac{\mu\theta_{\rm ss}^{-\nu}}{\varepsilon}}{\left(\xi_e\left(\frac{\nu}{1-\nu}\right)+\frac{\mu\theta_{\rm ss}^{-\nu}}{\varepsilon}\right)} > 0. \quad (46)$$

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

E Model properties for the goods-only version

This appendix investigates how typical model features such as price/wage stickiness and investment adjustment costs as well as modifications of the monetary policy rule affect the results.[102] We focus mainly on TFP shocks because – as explained in the main text – the ability of the model to generate a countercyclical inventory-sales ratio

[102]The size of the shock is set such that it generate a peak 1% increase in GDP in the version with all features added. Other parameter values are set equal to the moment-matching set of the model with both sectors, except for ν, which controls the curvature of the function characterizing the search friction. When the same value for ν is used, then the responses of the customer-finding rate following a monetary-policy shock are too strong in the model with only a goods sector leading to undesirable properties regarding inventories, production, and sales. To make the comparison with the full model transparent, we set ν equal to 0.851 which ensures that the two versions generate an identical value for the response of the customer-finding rate relative to the response of goods-sector production level following a monetary-policy shock. Recall that a higher value of ν implies more curvature and thus a less volatile customer-finding rate. Why is the response of the customer-finding rate lower in the full model with both sectors when the same value for ν is used? In the economy with only a goods sector, the goods sector produces all consumption and all investment goods. In the full model, however, production is shared with the service sector and the goods sector is responsible for a larger share of investment goods, consistent with the data. This means that production of the goods-sector will be more volatile in the complete model. What about the volatility of the customer-finding rate. As explained in section 3.4 and appendix F, the customer-finding rate in both models is determined in a subsystem and the included variables are only affected by inflationary pressure and the value of an unsold inventory good. The output level is not part of this subsystem. That is, the scale of operations does not have a direct effect. And a more volatile output level would only affect the volatility of the customer-finding if it does so indirectly. And that turns out not to be the case because inflationary pressure and the value of an unsold good are not that different in the two versions of our model.

and other desirable inventory properties in response to TFP shocks is surprising.[103] And we want to document robustness of the main results.

E.1 Responses to a TFP shock

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.[104]

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 (23), 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.[105] 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.[106] Consequently, the customer-finding rate (inventory-sales ratio) is procyclical (countercyclical) as observed in the data.[107]

Sticky prices. Figure 6 displays the results when prices are sticky, there are no investment adjustment costs, wages are not sticky, and the central bank does not

[103]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 under the standard assumption of sticky prices which is counterfactual.

[104]See section 4.2 and in particular footnote 69 for details and motivation. In appendix H.1, we discuss model predictions when TFP is instead assumed to be a stationary process.

[105]The response of the aggregate price level, P_t, depends on the rest of the model and in particular on whether monetary policy responds to de/inflationary pressure.

[106]As shown in equation (26), 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 (23), the determination of marginal costs is a bit more complicated here 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 D 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 $\lambda^f_{x,t}$.

[107]As discussed in the main text, key is that expected consumption growth is positive following a positive shock which ensures that the discount rate is procyclical which in turn implies a countercyclical $\lambda_{x,t}$.

respond to the output gap, i.e., $\Gamma_y = 0$.[108] 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 the output-gap panel.[109]

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 mainly driven by the fall in $\lambda^f_{x,t}$ which in turn is driven by the expected increase in consumption. The increase in inflationary pressure only has a marginal effect since it is so small, but it does reinforce the procyclical response of the customer-finding rate.

The transition to new permanent levels (for non-stationary variables) or steady-state levels (for stationary variables) occurs relatively fast for most variables. This is also true for production levels. But consumption and investment dynamics take a long time to settle down. This is a robust outcome and also present when we add additional model features.

There is one prediction that is not satisfactory and that is that investment actually drops on impact.[110] Investment-adjustment costs and sticky wages will push up the initial investment response. How these two standard model features affect the inventory 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.[111] The results are shown in figure 7. The response of the customer-finding rate is now stronger which 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.

[108] The IRFs and all other model properties are based on a first-order perturbation approximation.

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

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

[111] 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.

Adding sticky wages. As explained in the main text – and contrary to our expectations – sticky wages do not play a key role for fluctuations in inventories. Figure 8 displays the results when wages are sticky. Consistent with the standard NK model, the initial response of the aggregate economy to a TFP shock is stronger in the presence of sticky wages. Moreover, the output gap is now substantially positive. Also, adding sticky wages to the model reduces the magnitude of the increase in the customer-finding rate. But – as pointed out in the main text – wage stickiness only affects the customer finding rate indirectly by changing the responses of $\lambda^f_{x,t}$ and $\lambda^f_{d,t}$. This dampened response of the customer-finding rate is beneficial because it ensures that the output response is quite a bit stronger than the sales response as is observed in the data. There are two reasons why the customer-finding rate increase is dampened. Inflationary pressure is reduced with sticky wages. Moreover, there is initially a sharper increase in real activity with sticky wages, 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 in section 3.4.

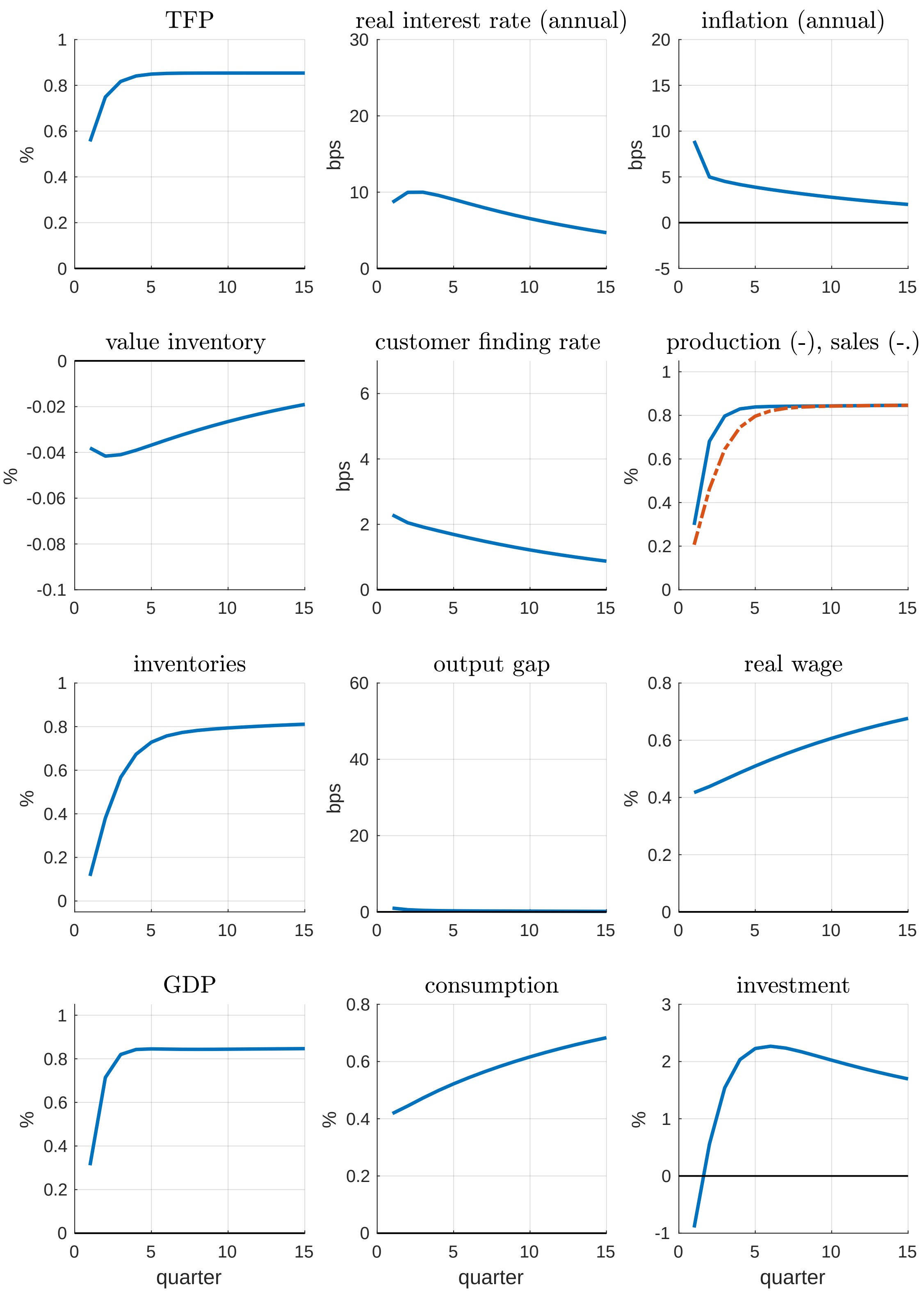
With sticky wages, the deviation from divine coincidence increases. Inflationary pressure is still small. Except for a 14 basis points increase on impact, it is less than (a bit more than) three basis points during the transition. However, there is now a nontrivial output gap which starts out at roughly 0.5% of flexible-price output on impact. Since $\Gamma_y = 0$ for 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 9. 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 4.8 on impact instead of 5.2 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.[112] At this higher value of Γ_y, the customer-finding rate is still clearly procyclical with a peak response of 4.1 basis points, but the tightening of the central bank in response of a positive TFP

[112]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 7.0 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 2.4 basis points. So it seems reasonable to conclude that the central bank "roughly" keeps inflation at target with this value of Γ_y.

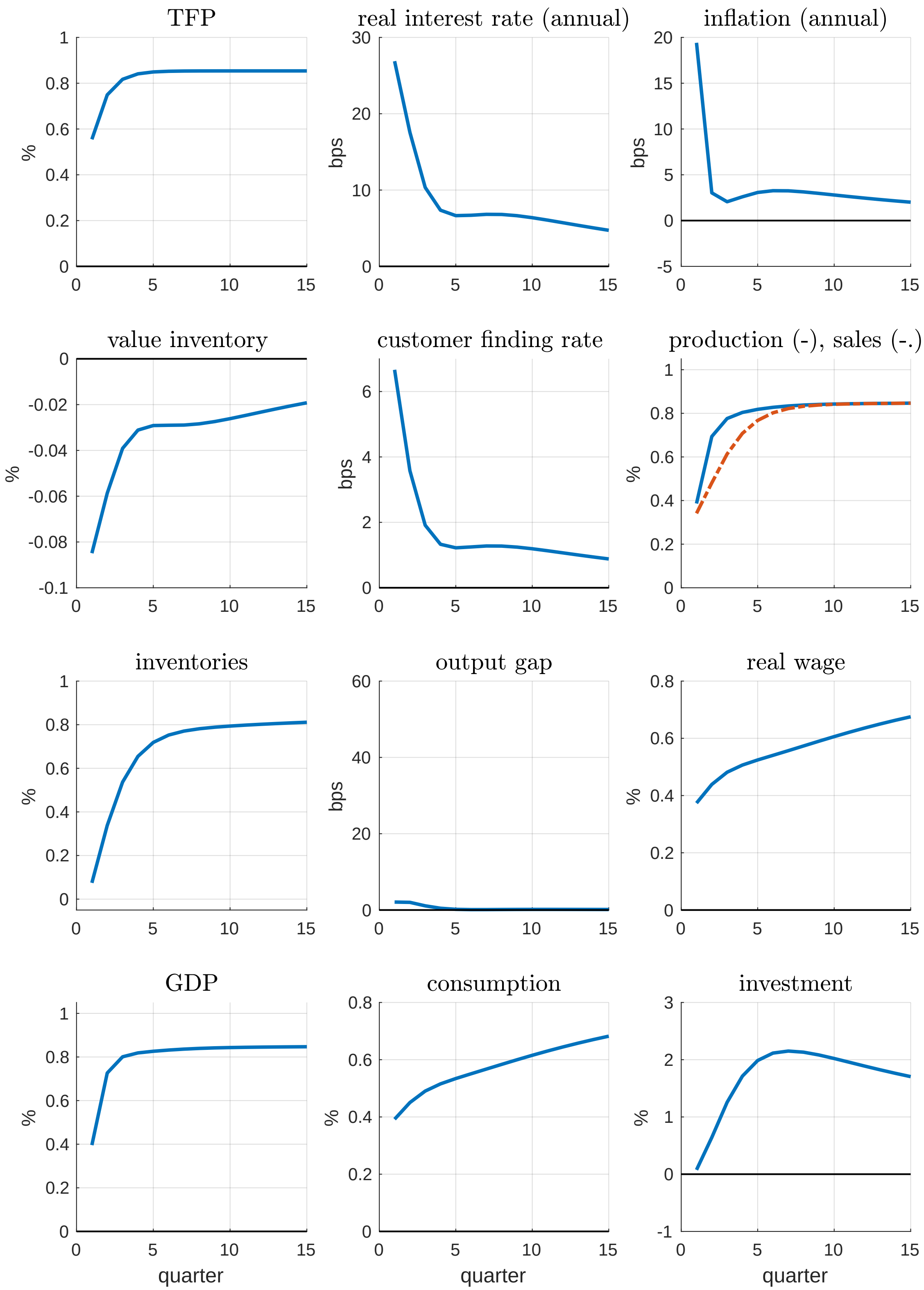
shock could in principle be so strong that the customer-finding rate falls.

Figure 6: TFP shock; $\eta_P > 0, \eta_i = 0, \eta_W = 0, \Gamma_y = 0$



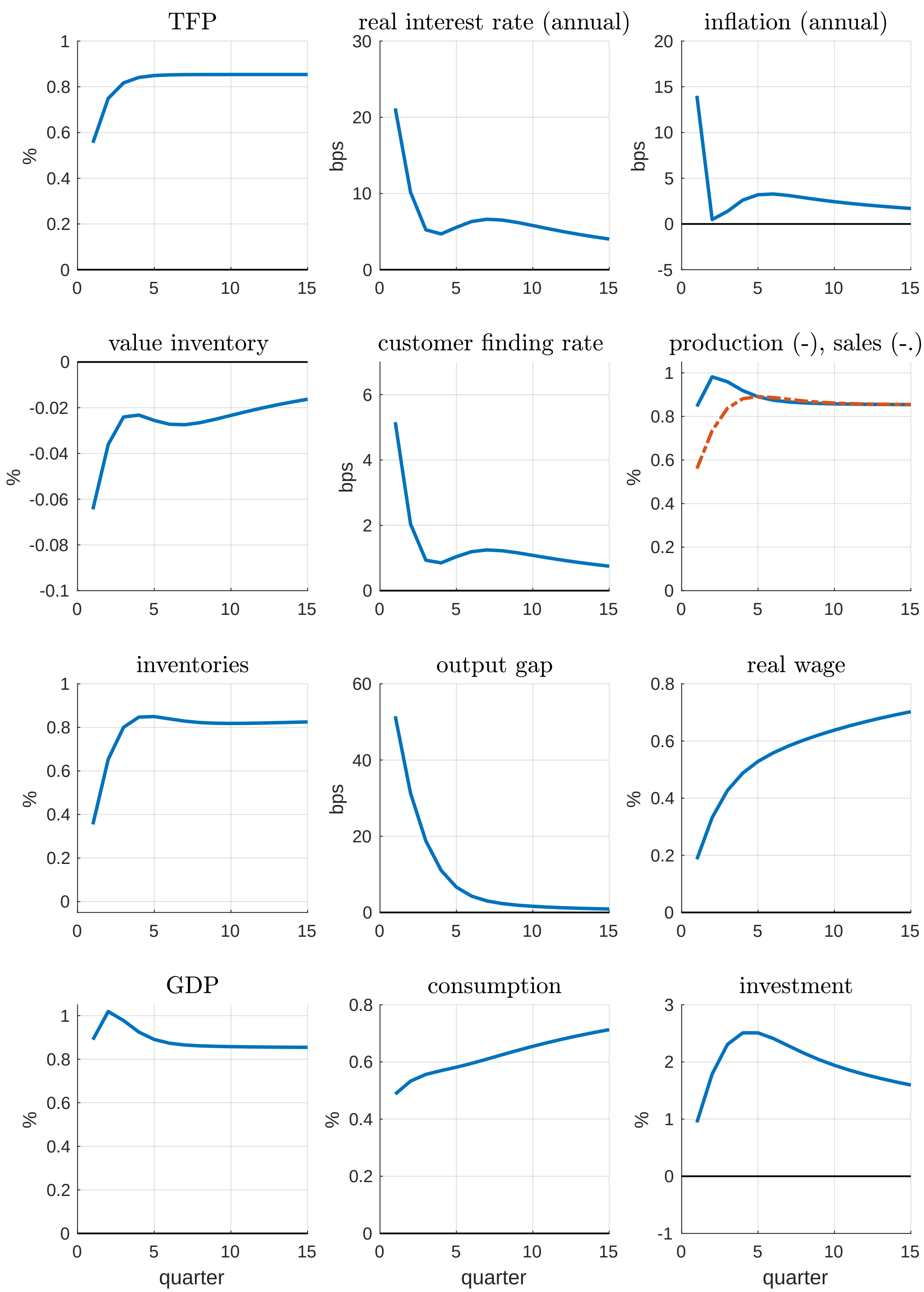
Notes. Impact of a TFP shock with sticky prices ($\eta_P = 10$), but no sticky wages nor investment adjustment costs. The value of Γ_y is of very little importance here, since the output gap is approximately zero.

Figure 7: TFP shock: plus investment adjustment costs; $\eta_P > 0, \eta_i > 0, \eta_W = 0, \Gamma_y = 0$



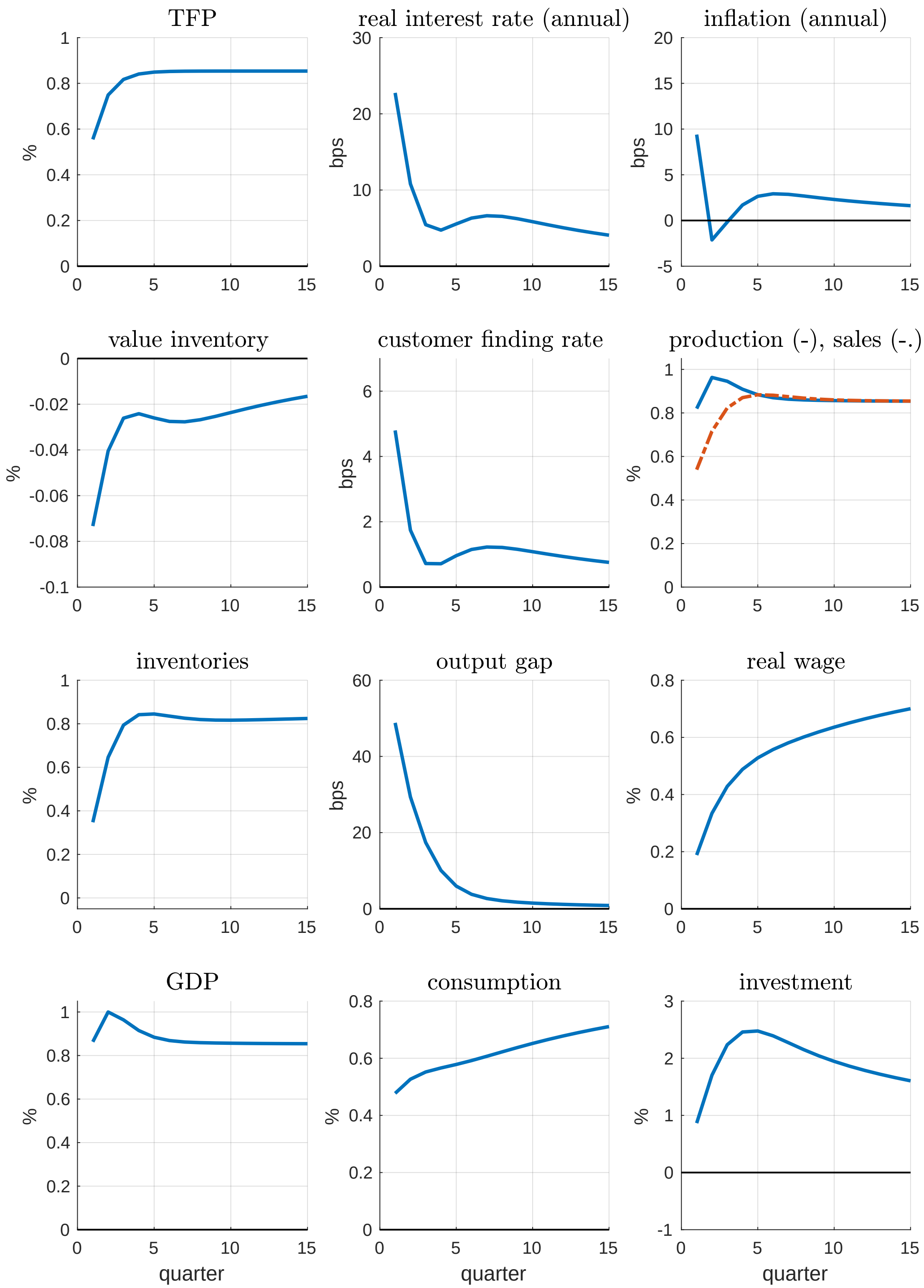
Notes. Impact of a TFP shock with sticky prices ($\eta_P = 10$) and investment adjustment costs ($\eta_i = 0.1$). No sticky wages. The value of Γ_y is of very little importance here, since the output gap is approximately zero.

Figure 8: TFP shock: plus sticky wages; $\eta_P > 0, \eta_i > 0, \eta_W > 0, \Gamma_y = 0$



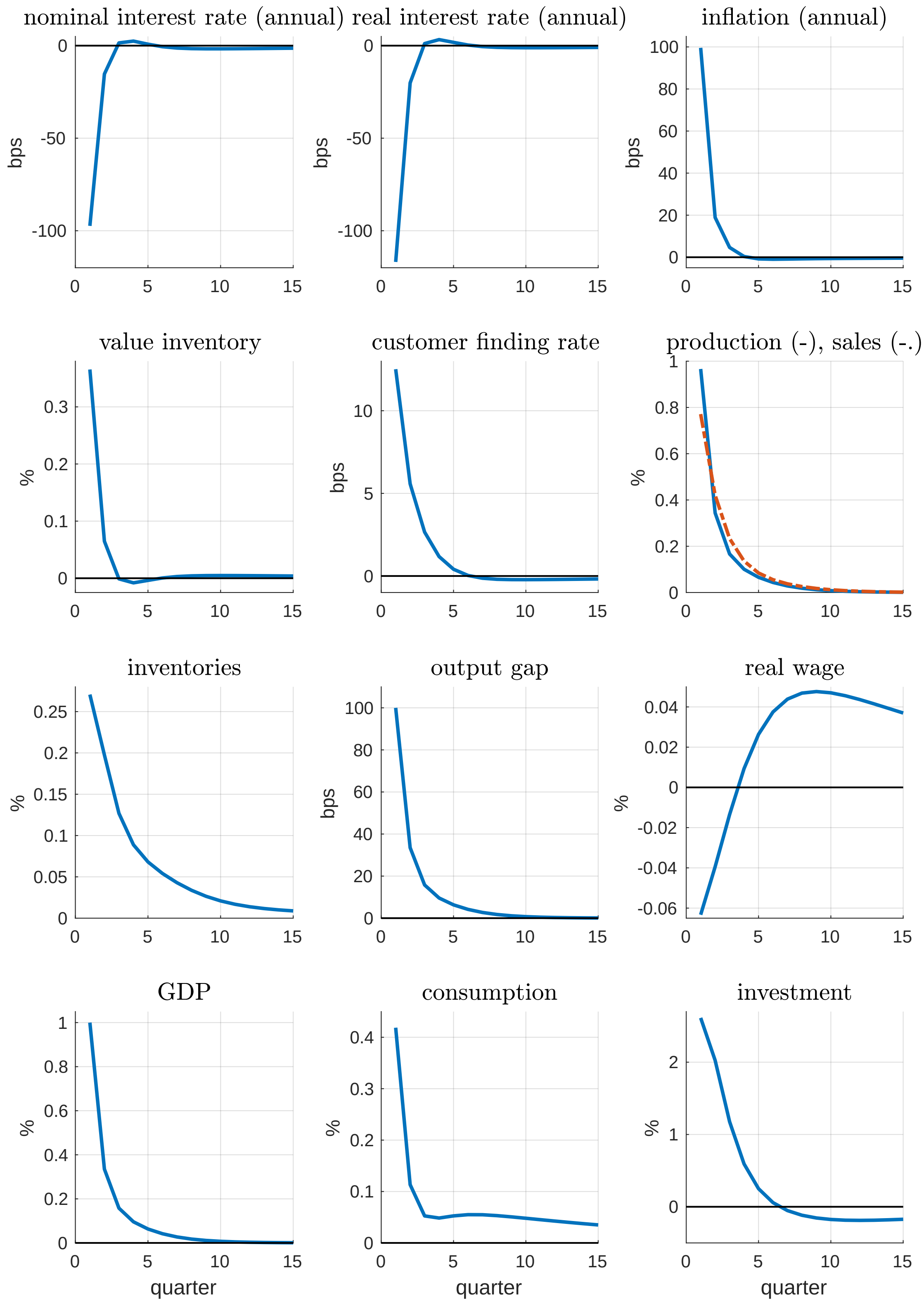
Notes. Impact of a TFP shock with sticky prices ($\eta_P = 10$), sticky wages ($\eta_W = 10$), and investment adjustment costs ($\eta_i = 0.1$).

Figure 9: 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 ($\Gamma_y = 0.03$).

Figure 10: 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 ($\Gamma_y = 0.03$).

E.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. For these parameter values, the response of the customer-finding rate is not too strong which ensures that the response of the inventory stock is positive. Moreover, the output response is substantially stronger than the sales response initially; after a while the responses become quite similar. Investment in inventories is also procyclical, because the sharp initial increase dominates the subsequent gradual decreases observed for the inventory level.

F Additional details for the complete model with services

F.1 Household problem

The household problem is given by the following optimization problem:[113]

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

subject to

[113]To economize on notation, we drop the h subscript which was used in section 3 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 $f^b_{i,g,t}$ and $f^b_{i,s,t}$ 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 when we incorporate services into the model.

$$
\begin{aligned}
\int_{i=0}^{1} P_{i,g,t}s_{i,g,t}di + \int_{i=0}^{1} P_{i,s,t}s_{i,s,t}di + b_t & \\
\leq W_t n_t + R_{k,t}k_{t-1} + d_t + (1+R_{t-1})b_{t-1}, & \quad (47a)\\
c_t \leq \min\left\{\frac{\bar{c}_{g,t}}{\omega_{g,c}}, \frac{c_{s,t} - \Upsilon_s(\xi_e e_t - \overline{\xi}_e)}{\omega_{s,c}}\right\}, & \quad (47b)\\
i_t \leq \min\left\{\frac{i_{g,t}}{\omega_{g,i}}, \frac{i_{s,t}}{\omega_{s,i}}\right\}, & \quad (47c)\\
\bar{c}_{g,t} = (1-\delta_c)\,\bar{c}_{g,t-1} + c_{g,t} - \Upsilon_g(\xi_e e_t - \overline{\xi}_e), & \quad (47d)\\
c_{g,t} + i_{g,t} = \left(\int_{i=0}^{1} s_{i,g,t}^{\frac{\varepsilon_g - 1}{\varepsilon_g}}\, di\right)^{\frac{\varepsilon_g}{\varepsilon_g - 1}}, & \quad (47e)\\
c_{s,t} + i_{s,t} = \left(\int_{i=0}^{1} s_{i,s,t}^{\frac{\varepsilon_s - 1}{\varepsilon_s}}\, di\right)^{\frac{\varepsilon_s}{\varepsilon_s - 1}}, & \quad (47f)\\
k_t = (1-\delta_k)\,k_{t-1} + i_t\left(1 - \frac{\eta_i}{2}\left(\frac{i_t}{i_{t-1}} - 1\right)^2\right), & \quad (47g)\\
e_t = \int_{i=0}^{1}\left(\frac{s_{i,g,t}}{f_{i,g,t}^b} + \frac{s_{i,s,t}}{f_{i,s,t}^b}\right) di. & \quad (47h)
\end{aligned}
$$

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 type-i goods, $P_{i,s,t}$ the price of type-i services, 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

$$
\begin{aligned}
\bar{c}_{g,t} &= \omega_{g,c}c_t, & (48a)\\
c_{g,t} &= \omega_{g,c}\left(c_t - c_{t-1}\left(1-\delta_c\right)\right) + \Upsilon_g(\xi_e e_t - \overline{\xi}_e), & (48b)\\
c_{s,t} &= \omega_{s,c}c_t + \Upsilon_s(\xi_e e_t - \overline{\xi}_e), & (48c)\\
i_{g,t} &= \omega_{g,i}i_t, & (48d)\\
i_{s,t} &= \omega_{s,i}i_t. & (48e)
\end{aligned}
$$

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:

$$s_{i,g,t} = \left(\left(\Upsilon_g + \Upsilon_s \frac{P_{s,t}}{P_{g,t}}\right) \frac{\xi_e}{f^b\left(\theta_{i,g,t}\right)} + \frac{P_{i,g,t}}{P_{g,t}}\right)^{-\varepsilon_g} s_{g,t}, \tag{49}$$

$$s_{i,s,t} = \left(\left(\Upsilon_g \frac{P_{g,t}}{P_{s,t}} + \Upsilon_s\right) \frac{\xi_e}{f^b\left(\theta_{i,s,t}\right)} + \frac{P_{i,s,t}}{P_{s,t}}\right)^{-\varepsilon_s} s_{s,t}. \tag{50}$$

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 (16), 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[114]

$$P_{g,t} = \left(\int_0^1 \left(\frac{(\Upsilon_g P_{g,t} + \Upsilon_s P_{s,t})\xi_e}{f^b\left(\theta_{i,g,t}\right)} + P_{i,g,t}\right)^{1-\varepsilon_g} di\right)^{\frac{1}{1-\varepsilon_g}}, \text{ and} \tag{51}$$

$$P_{s,t} = \left(\int_0^1 \left(\frac{(\Upsilon_g P_{g,t} + \Upsilon_s P_{s,t})\xi_e}{f^b\left(\theta_{i,s,t}\right)} + P_{i,s,t}\right)^{1-\varepsilon_s} di\right)^{\frac{1}{1-\varepsilon_s}}. \tag{52}$$

The aggregate price for consumption goods is given by

$$P_t = \omega_{g,c} P_{g,t} + \omega_{s,c} P_{s,t}. \tag{53}$$

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

F.2 Interaction between goods and service sector

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.[115] This subsystem resembles

[114] As pointed out in footnote 31, 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.

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

$$\frac{P_{g,t}}{P_t} = \frac{\tilde{\psi}_t \xi_e}{f_g^b\left(\theta_{g,t}\right)} + \frac{P_{i,g,t}}{P_t}, \tag{54a}$$

$$MC_{g,t} - \lambda_{x,t}^f = \left(\frac{P_{i,g,t}}{P_t} - \lambda_{x,t}^f - \frac{P_{g,t}}{P_{g,t}}\lambda_{d,g,t}^f\right) f_g^f\left(\theta_{g,t}\right), \tag{54b}$$

$$MC_{g,t} - \lambda_{x,t}^f = \varepsilon_g \lambda_{d,g,t}^f \xi_e \tilde{\psi}_t \frac{\nu_g}{1-\nu_g}\theta_{g,t}, \tag{54c}$$

$$\frac{P_{s,t}}{P_t} = \frac{\tilde{\psi}_t \xi_e}{f_s^b\left(\theta_{s,t}\right)} + \frac{P_{i,s,t}}{P_t}, \tag{54d}$$

$$MC_{s,t} = \left(\frac{P_{i,s,t}}{P_t} - \frac{P_{s,t}}{P_{s,t}}\lambda_{d,s,t}^f\right) f_s^f\left(\theta_{s,t}\right), \tag{54e}$$

$$MC_{s,t} = \varepsilon_s \lambda_{d,s,t}^f \xi_e \tilde{\psi}_t \frac{\nu_s}{1-\nu_s}\theta_{s,t}, \tag{54f}$$

$$1 = \omega_{g,c}\frac{P_{g,t}}{P_t} + \omega_{s,c}\frac{P_{s,t}}{P_t}, \tag{54g}$$

$$MC_{g,t} = \frac{A_{s,t}}{A_{g,t}} MC_{s,t}, \tag{54h}$$

$$\widetilde{\psi}_t = \Upsilon_g \frac{P_{g,t}}{P_t} + \Upsilon_s \frac{P_{s,t}}{P_t}, \tag{54i}$$

where $\lambda_{d,g,t}^f$, $\lambda_{d,s,t}^f$, and $\lambda_{x,t}^f$ are given by

$$1 - \varepsilon_g \lambda_{d,g,t}^f = \left(\eta_{P,g}\frac{P_t}{P_{i,g,t}}\right)\left(\begin{array}{c} \left(\frac{P_{i,g,t}}{P_{i,g,t-1}} - 1\right)\frac{P_{i,g,t}}{P_{i,g,t-1}} \\ -\beta \mathbb{E}_t\left[\frac{\lambda_{t+1}}{\lambda_t}\left(\frac{P_{i,g,t+1}}{P_{i,g,t}} - 1\right)\frac{P_{i,g,t+1}}{P_{i,g,t}}\frac{s_{g,t+1}}{s_{g,t}}\right]\end{array}\right), \tag{55a}$$

$$1 - \varepsilon_s \lambda_{d,s,t}^f = \left(\eta_{P,s}\frac{P_t}{P_{i,s,t}}\right)\left(\begin{array}{c} \left(\frac{P_{i,s,t}}{P_{i,s,t-1}} - 1\right)\frac{P_{i,s,t}}{P_{i,s,t-1}} \\ -\beta \mathbb{E}_t\left[\frac{\lambda_{t+1}}{\lambda_t}\left(\frac{P_{i,s,t+1}}{P_{i,s,t}} - 1\right)\frac{P_{i,s,t+1}}{P_{i,s,t}}\frac{s_{s,t+1}}{s_{s,t}}\right]\end{array}\right), \tag{55b}$$

$$\lambda_{x,t}^f = \beta\left(1-\delta_x\right)\left(1-\eta_x\right)\mathbb{E}_t\left[\frac{\lambda_{t+1}}{\lambda_t}MC_{g,t+1}\right]. \tag{55c}$$

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

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.

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.[116]

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. Equations (54a) and (54d) are rewritten versions of the demand equations for a type-i good and type-i service. It differs in in two ways from equation (23a), which presents the demand equation for the goods-only economy. 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 (54b) and (54e) are the equivalent of equation (23b). 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., $\lambda_{d,j,t}^f > 0, j \in \{g, s\}$. These equations are basically the same as equation ((23b) except that the sector's relative price is included. The equivalent versions of Equation (23c), i.e., the firm's first-order condition for tightness, are equations (54c) and (54f). The latter two take into account that search costs depend on both goods and services and, thus, on their relative prices. Equation (54g) simply states that the weighted sum of the two relative sector prices have to add up to 1.[117] Equation (54i) 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 (55a) and (55b). Equation (55c) gives an expression for the value of bringing an inventory good into the next period and corresponds to equation (23d).

Do the customer-finding rates in the two sectors move in the same direction? In response to positive demand shocks, the customer-finding rate increases in both sectors which works through changes in $\lambda_{d,g,t}^f$ and $\lambda_{d,s,t}^f$, exactly as in the goods-only model. Since this discussion is (again) quite intuitive, we focus on TFP shocks in this section.

In section 3, we learned that the countercyclical behavior of $\lambda_{x,t}^f$ 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 "unsold" services 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

[116]But recall from footnote 35, 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.

[117]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.

are fully flexible. Under flexible prices, we have that $\lambda^f_{d,g,t} = \lambda^f_{d,s,t} = 0$ which simplifies the sub-system given above considerably.[118]

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 set of equations that determine the outcomes for the goods sector on the one side and the set of equations that do this for the service sector on the other side. 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 (54g). Second, marginal costs satisfy the following relationship

$$MC_{g,t} = MC_{s,t}\frac{A_{s,t}}{A_{g,t}}. \tag{56}$$

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, $\lambda^f_{d,g,t}$ and $\lambda^f_{d,s,t}$ 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.

We will now discuss what the subsystem tells us about model outcomes when $\lambda^f_{x,t}$ falls and $A_{g,t}/A_{s,t}$, $\lambda^f_{d,g,t}$ and $\lambda^f_{d,s,t}$ remain constant. Using equations (54a), (54b), and (54c) 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, $P_{g,t}/P_t$.[119] 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

[118]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, $\lambda^f_{d,g,t}$ and $\lambda^f_{d,s,t}$ are always *exactly* equal to zero, which allows us to derive analytical results.

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

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 ${}^{P_{s,t}}/_{P_t} = (1 - \omega_{g,c}{}^{P_{g,t}}/_{P_t})/\omega_{c,s}$.

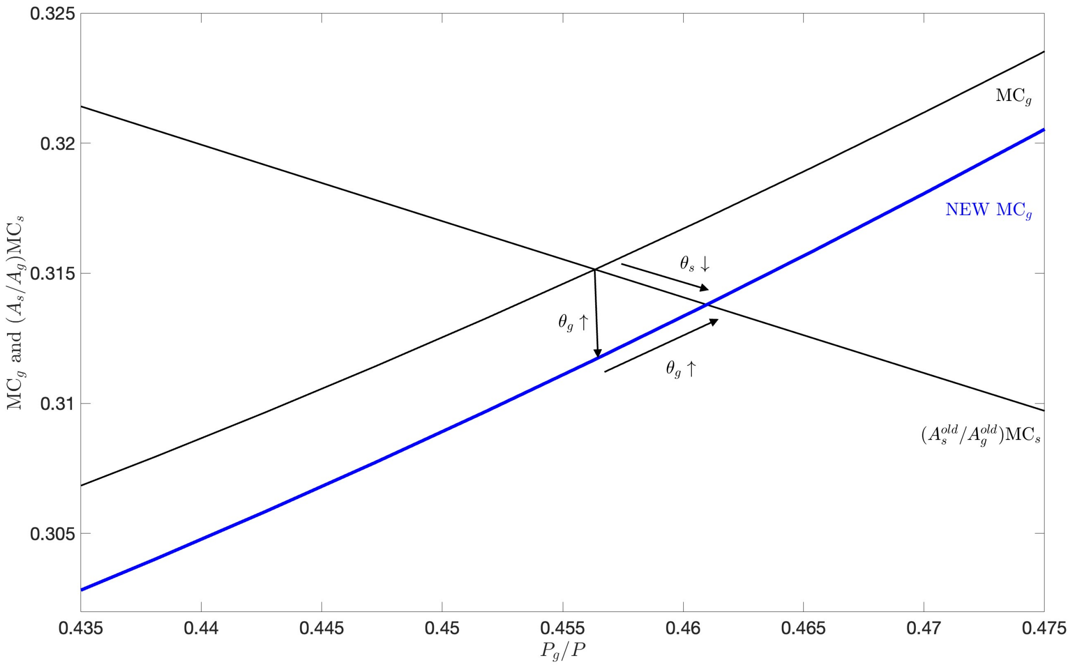
These two functions are plotted in figure 11, where we have scaled the marginal cost function for the service sector with ${}^{A_{g,t}}/_{A_{s,t}}$ consistent with equation (56). The solution of the subsystem is given by the intersection of the two curves at which point equation (56) is satisfied.

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.[120] 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 prediction of a countercyclical response of the customer-finding rate in the service sector turns out to be not robust. The reason is that $\lambda^f_{d,g,t}$ and $\lambda^f_{d,s,t}$ are not constant in the full model as divine coincidence no longer holds. So the customer-finding rate in the service sector could be counter- or procyclical, although a better way to characterize our model prediction for this variable is that it is acyclical since the responses are always small as long as ${}^{A_{g,t}}/_{A_{s,t}}$ remains constant. The third case discussed below explains why the response of the customer-finding rate in the service sector is procyclical if productivity in the service sector does not respond one for one with productivity in the goods sector.

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

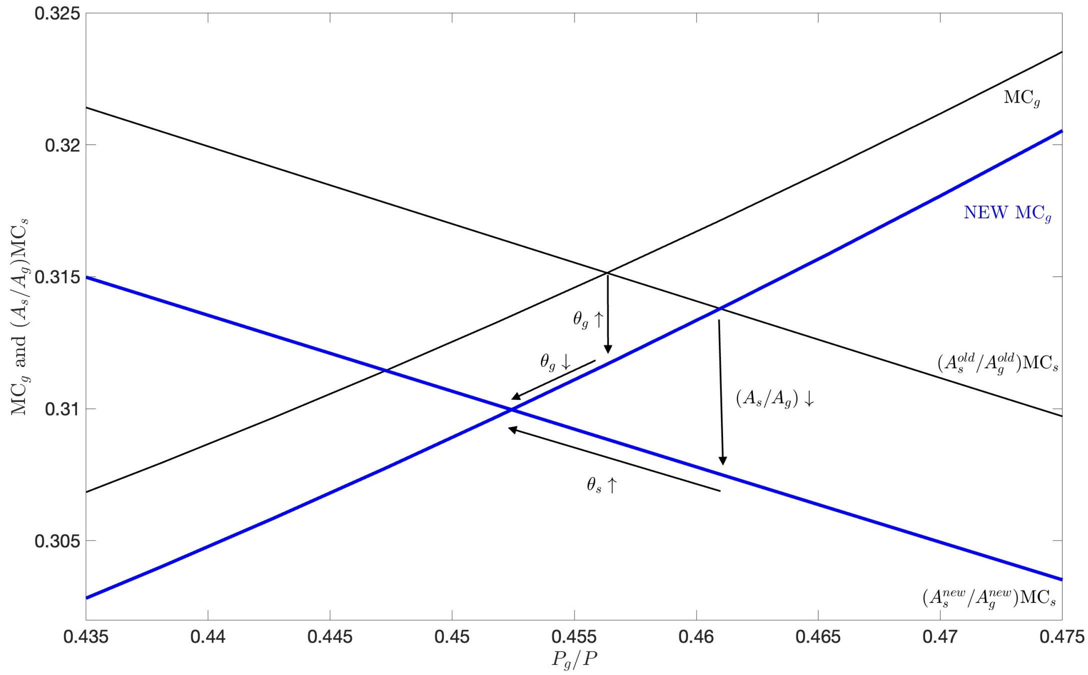
Figure 11: Impact of an increase in TFP and associated fall in $\lambda_{x,t}^f$ on both sectors; $\Delta\left(\frac{A_{g,t}}{A_{s,t}}\right) = 0$



Notes. As made clear in section 3, an increase in TFP leads to a decrease in the value of an unsold good, $\lambda_{x,t}^f$. Sub-system (54) implies that the drop in $\lambda_{x,t}^f$ 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 11 remain unchanged. However, there is *no* change in the customer-finding rate for the service sector, $\theta_{s,t}$ when – after the drop in $\lambda_{x,t}^f$ – 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 at this *corner* choice for Υ_g and Υ_s, again keeping $A_{g,t}/A_{s,t}$ constant.

Figure 12: 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 (54) 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 (56). 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 12. 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 for our parameter values. Consequently, customer-finding rates increase in both sectors. A quantitative illustration is given in appendix H.1.

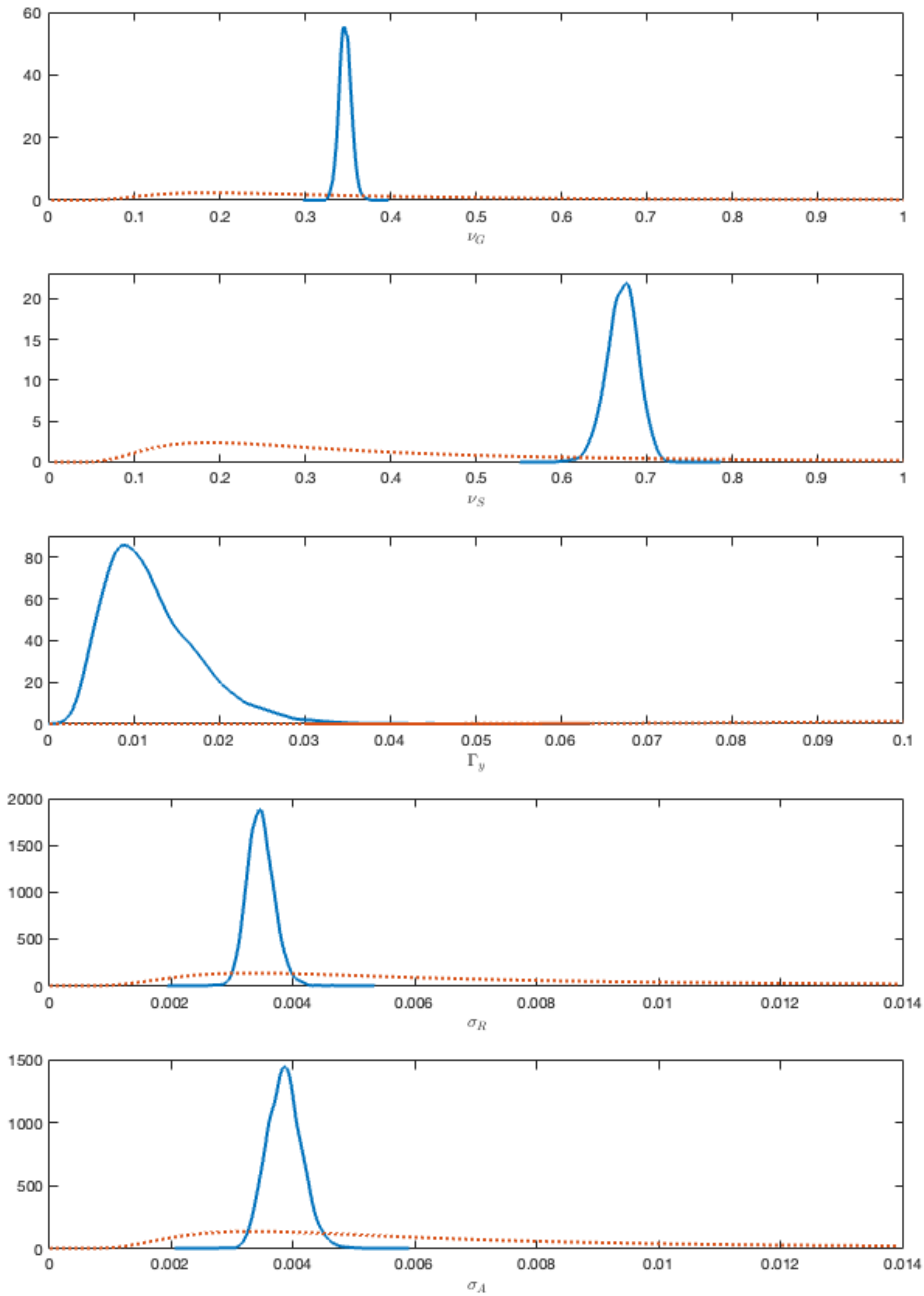
G Full-information estimation of key parameters

In the main text, we adopt 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 13 plots the prior and posterior densities and table 5 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 moment-matching exercise. 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.

Figure 13: 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 5. The posterior is obtained using five MCMC sequences with 10,000 observations each.

Table 5: 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.

H Additional results for the full model

H.1 Alternative TFP processes

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 process 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 convenient stationary (but persistent) process and the 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 variation. 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.[121] Thus, the interesting aspect of our model is that low-frequency properties of the driving process *do* matter for some real variables, namely inventories.

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 14 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 figures 2 and 3.[122] We see that the value of an inventory good, $\lambda^f_{x,t}$, increases on impact and the customer-finding rate drops. The reason $\lambda^f_{x,t}$ increases is that consumption is expected to fall following the initial increase. 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.[123]

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,

[121] 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 is likely to increase due to a rise in expected earnings.

[122] Throughout this appendix, we use the parameters based on the moment-matching exercise as the benchmark. Recall that the results based on the estimation are very similar, but that the IRF for inventories following a monetary policy shocks looks better for the moment-matching parameter set.

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

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 14. The figure shows that consumption now does display a minor – but long-lasting – hump and that the customer-finding rate is once again procyclical. Note that the two IRFs for consumption are very similar. Both are stationary processes and will eventually return to zero. But one displays a hump and the other starts a very gradual decline immediately. And this difference causes the customer-finding rate to behave very differently. 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 as this – on its own – is sufficient for the model to generate desirable inventory properties.

Productivity differences across sectors. In appendix F.2, 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.

$$\ln\left(\frac{A_{g,t}}{A_{g,t-1}}\right) = \rho_A \ln\left(\frac{A_{g,t}}{A_{g,t-1}}\right) + \varepsilon_{A,t} \tag{57a}$$

$$\ln\left(\frac{A_{s,t}}{A_{g,t}}\right) = \rho_{\text{gap}} \ln\left(\frac{A_{s,t-1}}{A_{g,t-1}}\right) - \rho_{\text{gap}} \ln\left(\frac{A_{g,t}}{A_{g,t-1}}\right) \tag{57b}$$

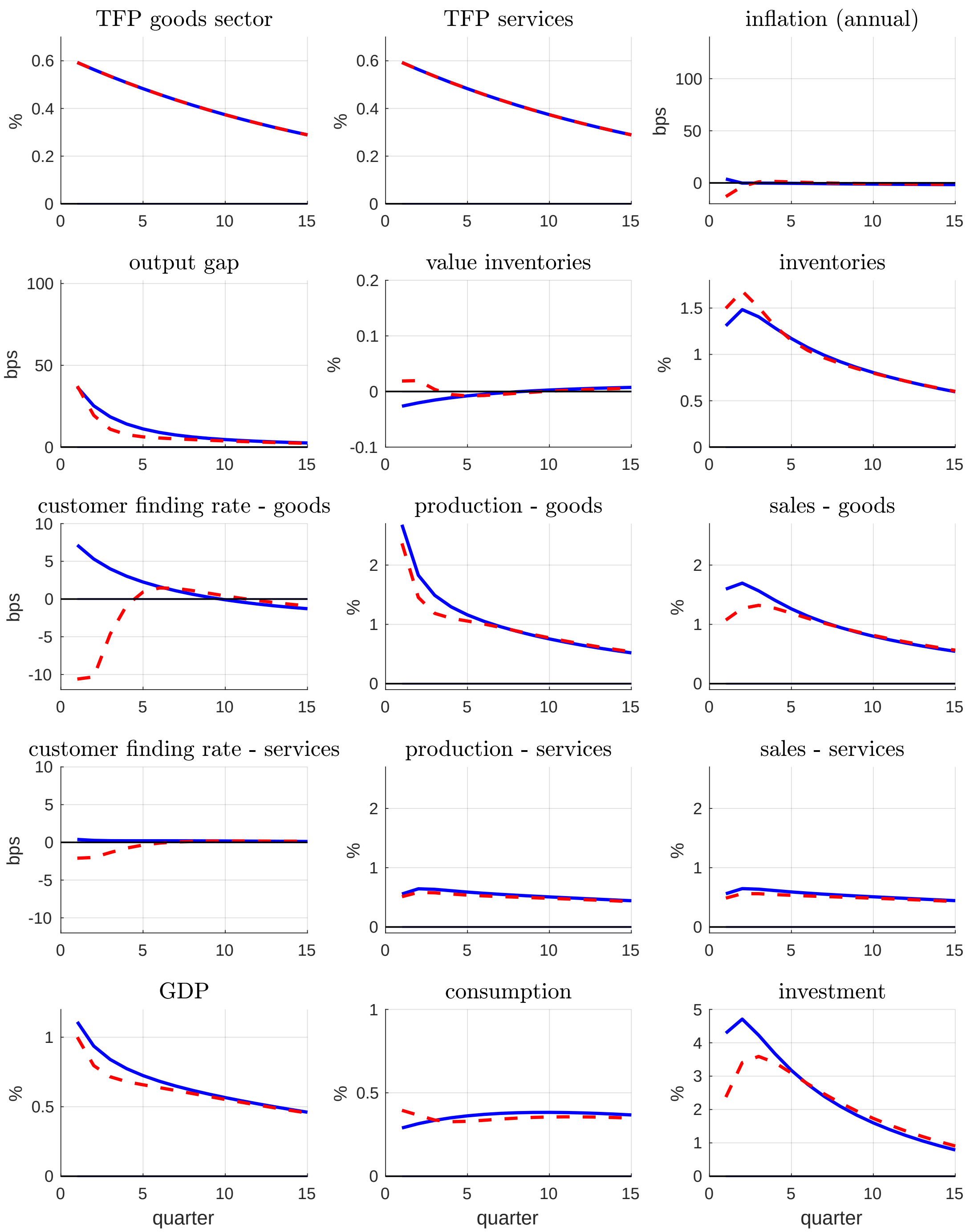
Thus, the law of motion for $A_{g,t}$ is unchanged. Following a shock, the change in $A_{s,t}$ is always less than the change in $A_{g,t}$ with the difference being the biggest on impact, but gradually going to zero.

Figure 15 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 appendix F.2, 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 six quarters after which there is only a very small positive one. But the customer-finding rate in the service sector still only increases by 5.68 basis points.[124]

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 available data discussed in section 2 is very short and only covers two recessions and demand factors are believed to have been important in both.

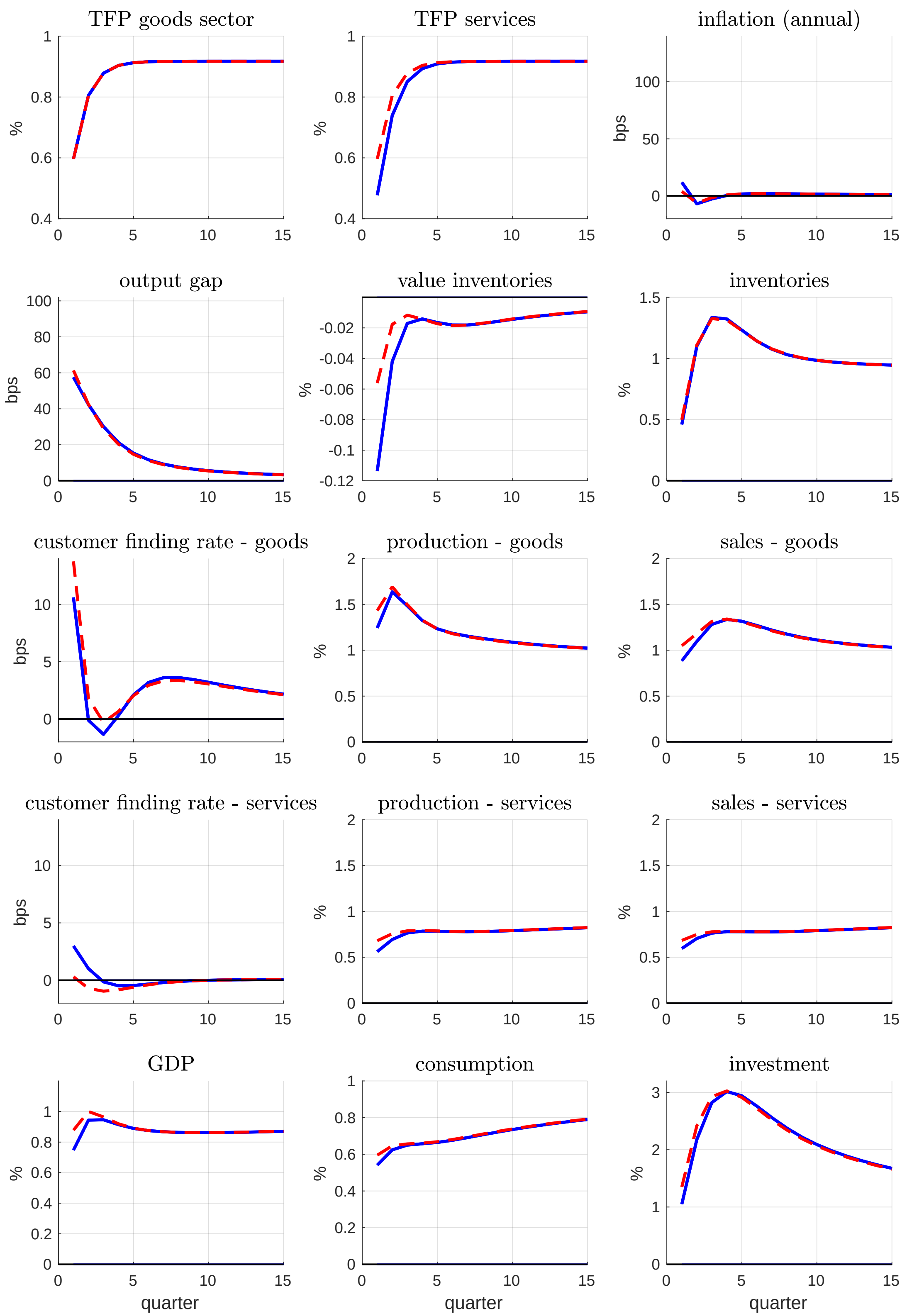
[124] A stronger response of $f^f_{s,t}$ is obtained when $A_{s,t} = A_{g,t-1}$, that is, the law of motion for $A_{s,t}$ is identical to the one of $A_{g,t}$ but with a one-period lag. Then the increase on impact is equal to 15.57 basis points. But again this comes at the cost of generating a negative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

Figure 14: Stationary TFP process; without (-) and with (- -) investment adjustment costs



Notes. TFP follows a stationary process with an auto-regressive coefficient equal to 0.95. Blue/solid lines display the case when there are no investment adjustment costs, i.e., $\eta_i = 0$, and the red/dashed lines the ones when $\eta_i = 0.2$. Other parameter values are set equal to the benchmark moment-matching values used to generate figures 2 and 3.

Figure 15: TFP shock; $A_{s,t}$ lags $A_{g,t}$



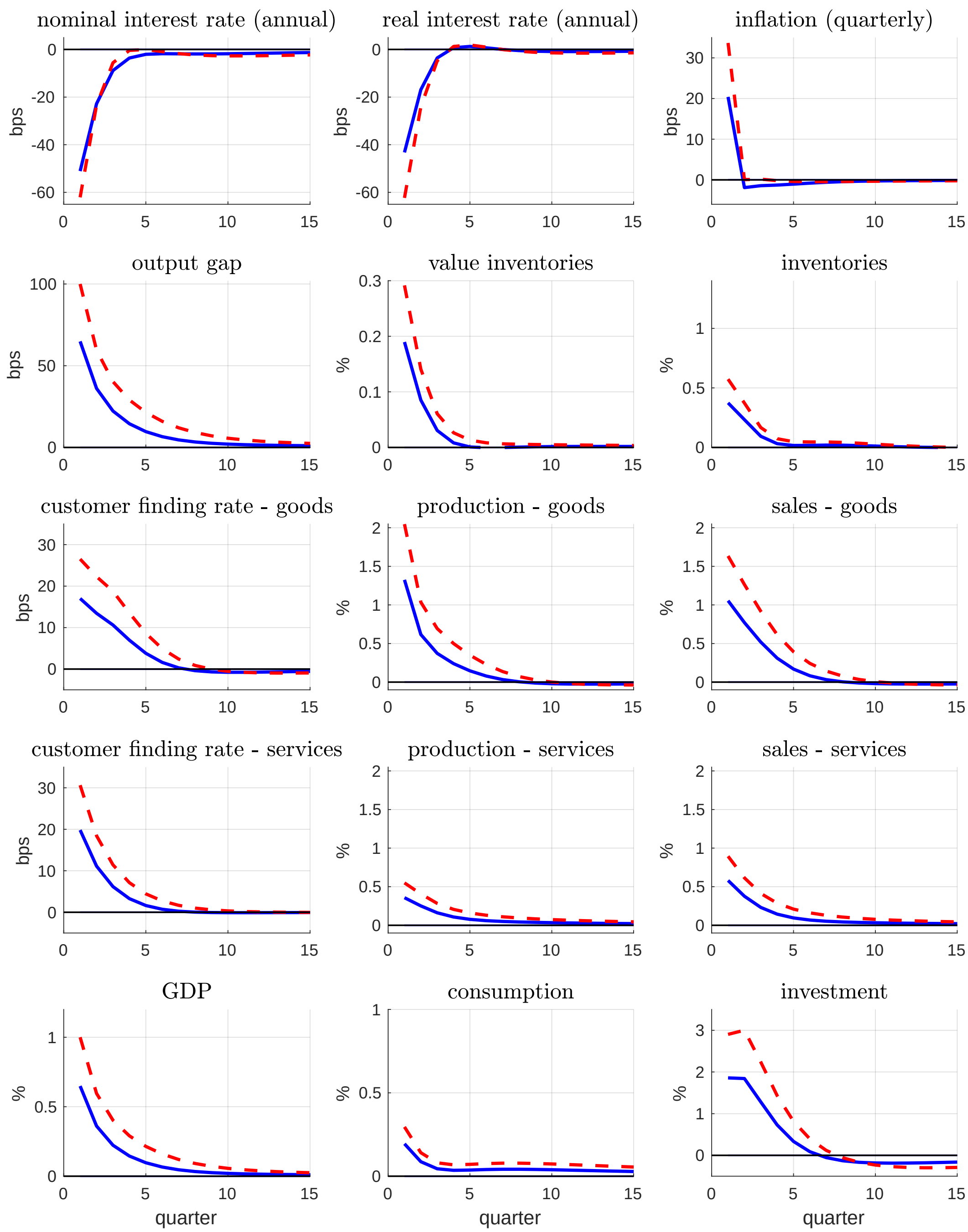
Notes. The blue/solid lines plot the IRFs when productivity in the service sector, $A_{s,t}$, lags productivity in the goods sector as in equation (57) with $\rho_{\text{gap}} = 0.2$. The red/dashed lines plot the benchmark results based on the calibration procedure. Other parameter values are set equal to the benchmark moment-matching values used to generate figures 2 and 3.

H.2 Alternative Taylor rule

Figure 16 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. 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 17.0 basis points and GDP increases with 0.65%. This corresponds to 26.2 basis points per percentage point increase in GDP. Thus, the alternative Taylor rule affects the overall impact of a monetary-policy shock, but not the *relative* responses.

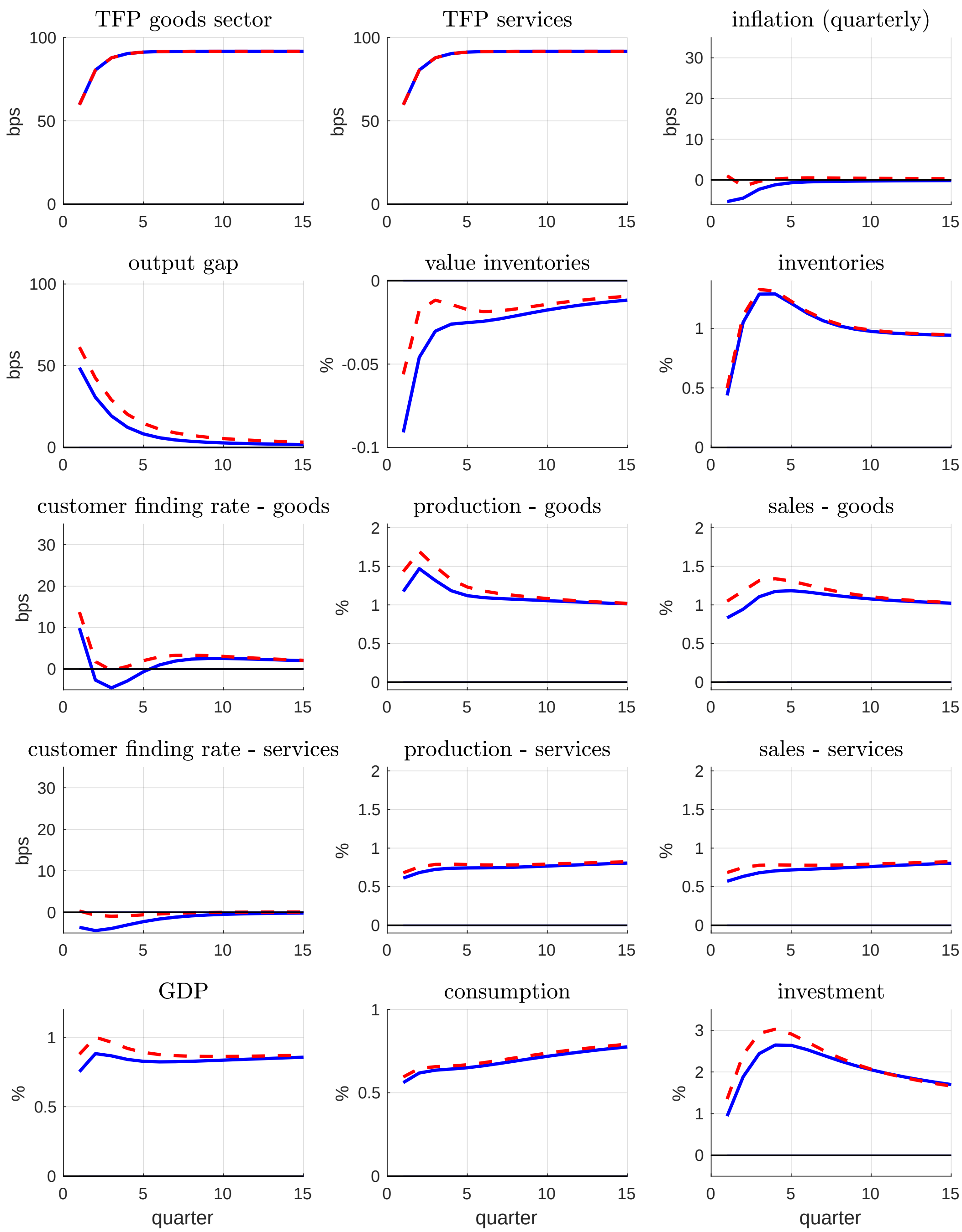
Figure 17 displays the corresponding IRFs for a TFP shock. For our benchmark Taylor rule, the central bank avoids both inflationary and deflationary pressure following this supply-side shock. For our more hawkish Taylor rule, a TFP shock is accompanied with some deflationary pressure. According to proposition 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. The response of the customer-finding rate in the service sector is now uniformly negative, but still quite small.

Figure 16: Monetary-policy shock; alternative Taylor Rule



Notes. The blue/solid 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 benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

Figure 17: TFP shock; alternative Taylor Rule



Notes. The blue/solid 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 benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

H.3 Alternative assumptions about search costs

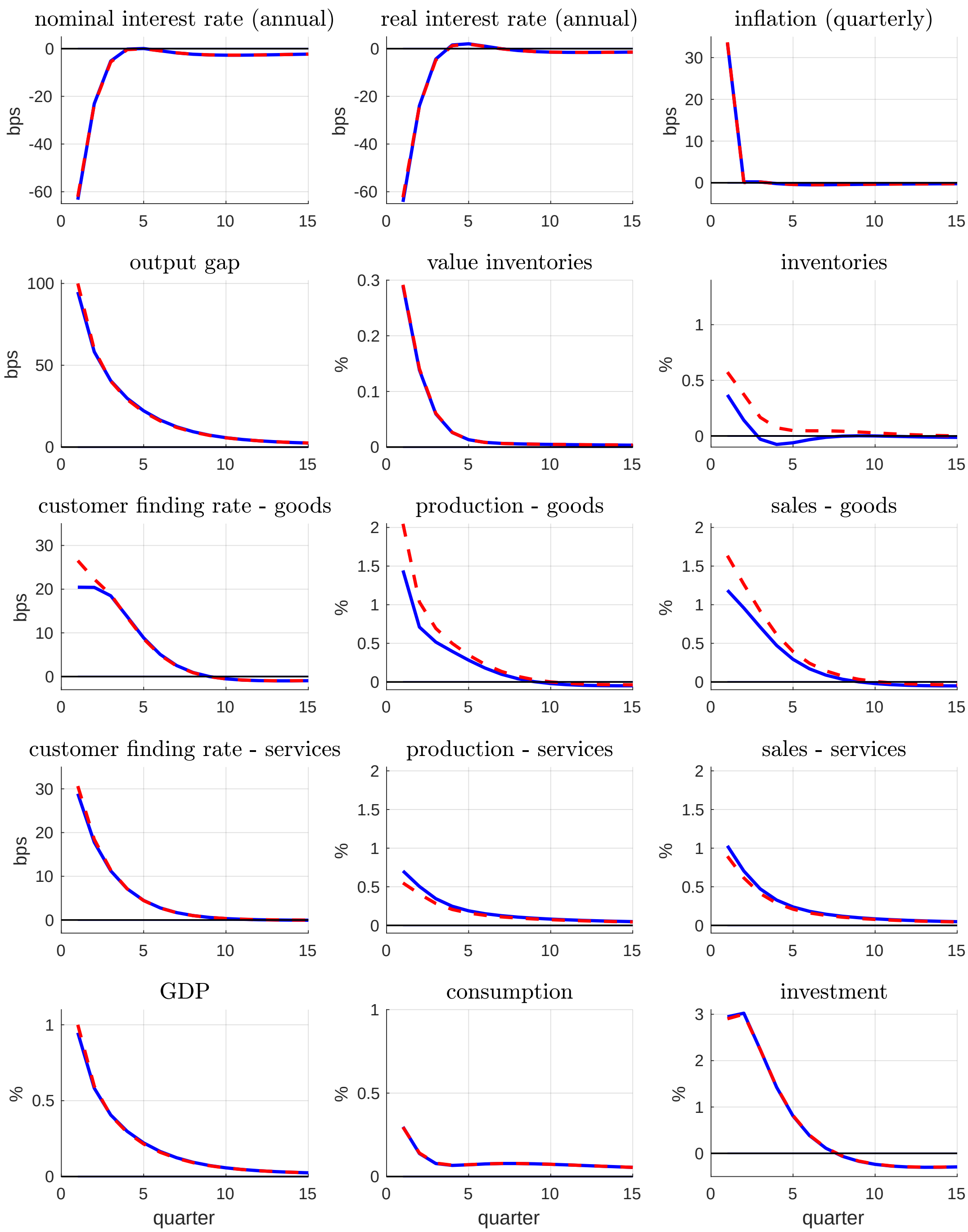
The results in the main text are based on the assumption that search costs consists of a mix of goods and services and the calibrated value of the fraction of services, Υ_s, was set equal to the fraction of services in the consumption bundle, $\omega_{s,c} = 0.5771$.[125] This may underestimate the services 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 18 and 19 display the associated IRFs for this alternative parameterization as well as the ones for our benchmark moment-matching parameter set as displayed in figures 2 and 3. Although the GDP response following a monetary expansion is very similar for the two cases, the expansion in the goods sector is less when goods are not needed for the acquisition of consumption and investment expenditures. This goes together with a somewhat smaller response of the customer-finding rate in the goods sector,namely 20.1 instead of 26.5 basis points. But the overall pattern of results for this alternative corner calibration is very similar to that of the benchmark.

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. Of course, the short-run responses could still be affected when the value of Υ_s changes. But this experiment indicates the relevance of having long-run restrictions that are equal by construction.

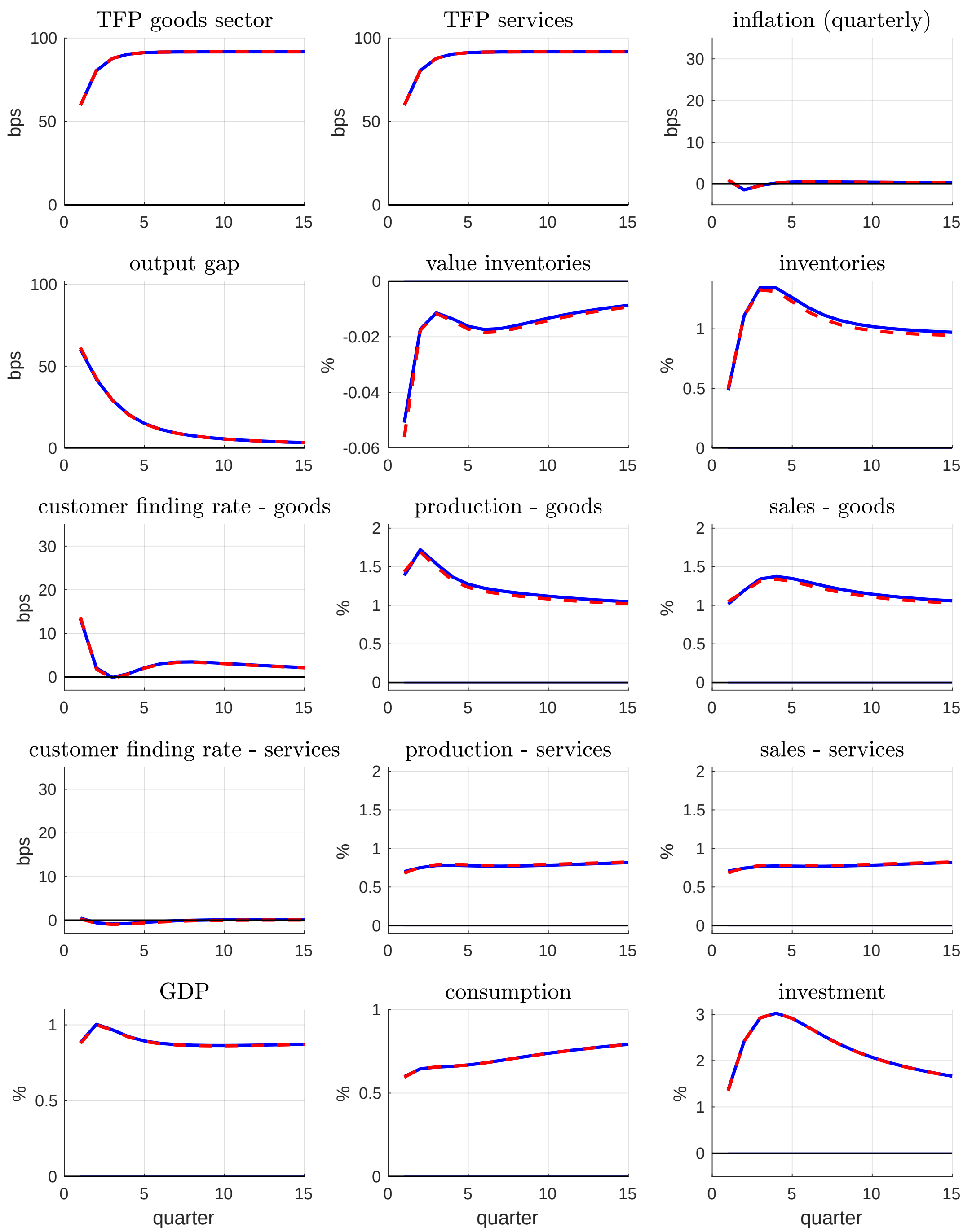
[125] And the fraction spend on goods, Υ_g is equal to $1 - \Upsilon_s$.

Figure 18: Monetary-policy shock; search only requires services - $\Upsilon_s = 1$



Notes. The blue/solid lines display IRFs of a monetary-policy shock when search costs are only in the form of services, that is, $\Upsilon_s = 1$. Other parameter values are set equal to the benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

Figure 19: TFP shock; Search only requires services - $\Upsilon_s = 1$



Notes. The blue/solid 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 benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

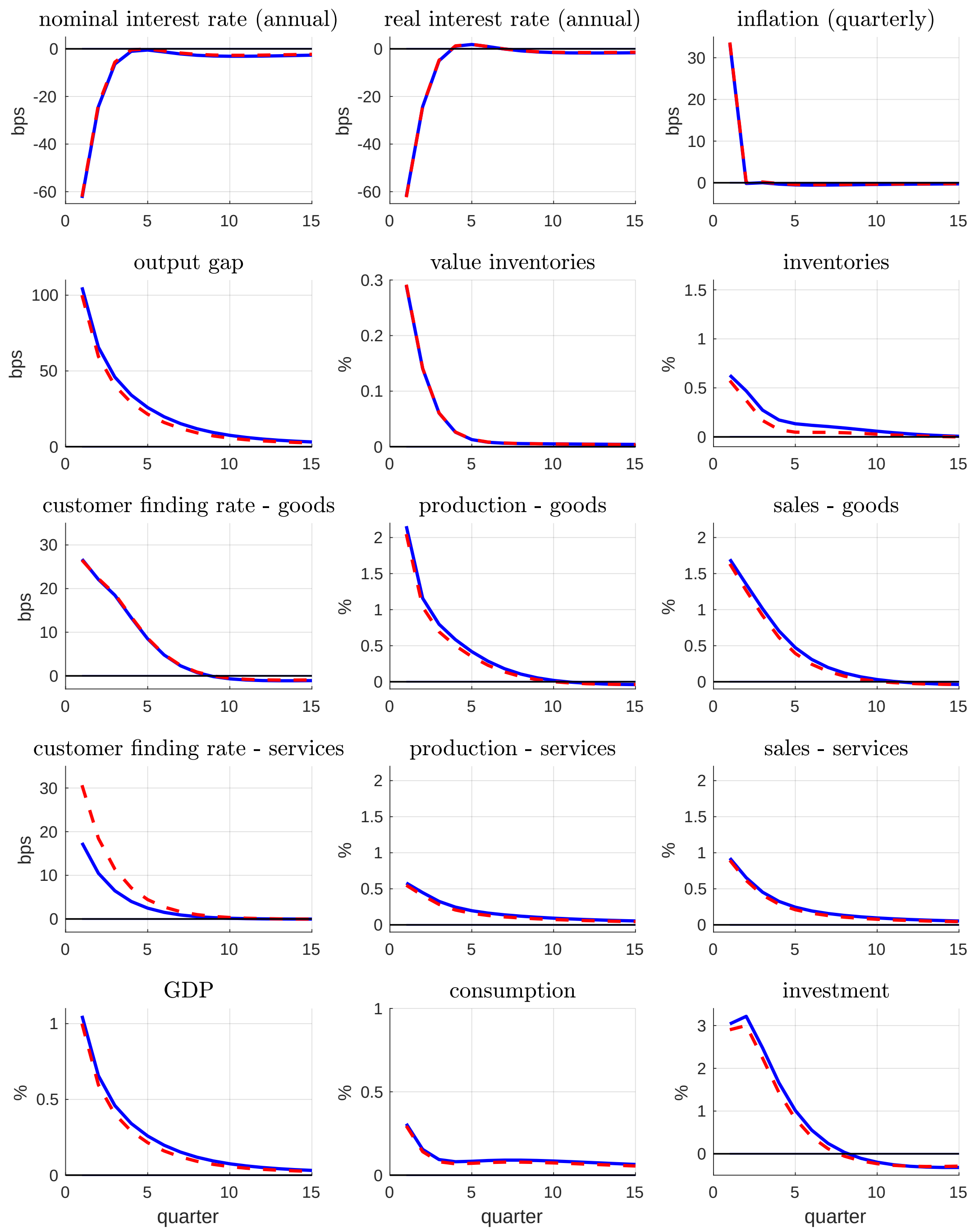
H.4 Alternative mean service-sector customer-finding rate

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 (1), 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. 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 20 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. Specifically, the service-sector customer-finding rate now goes up with only 17.4 instead of 30.6 basis points. But note that relative to the lower steady-state value – 0.506 instead of 0.888 - the response of the customer-finding rate is virtually the same in both cases as well.

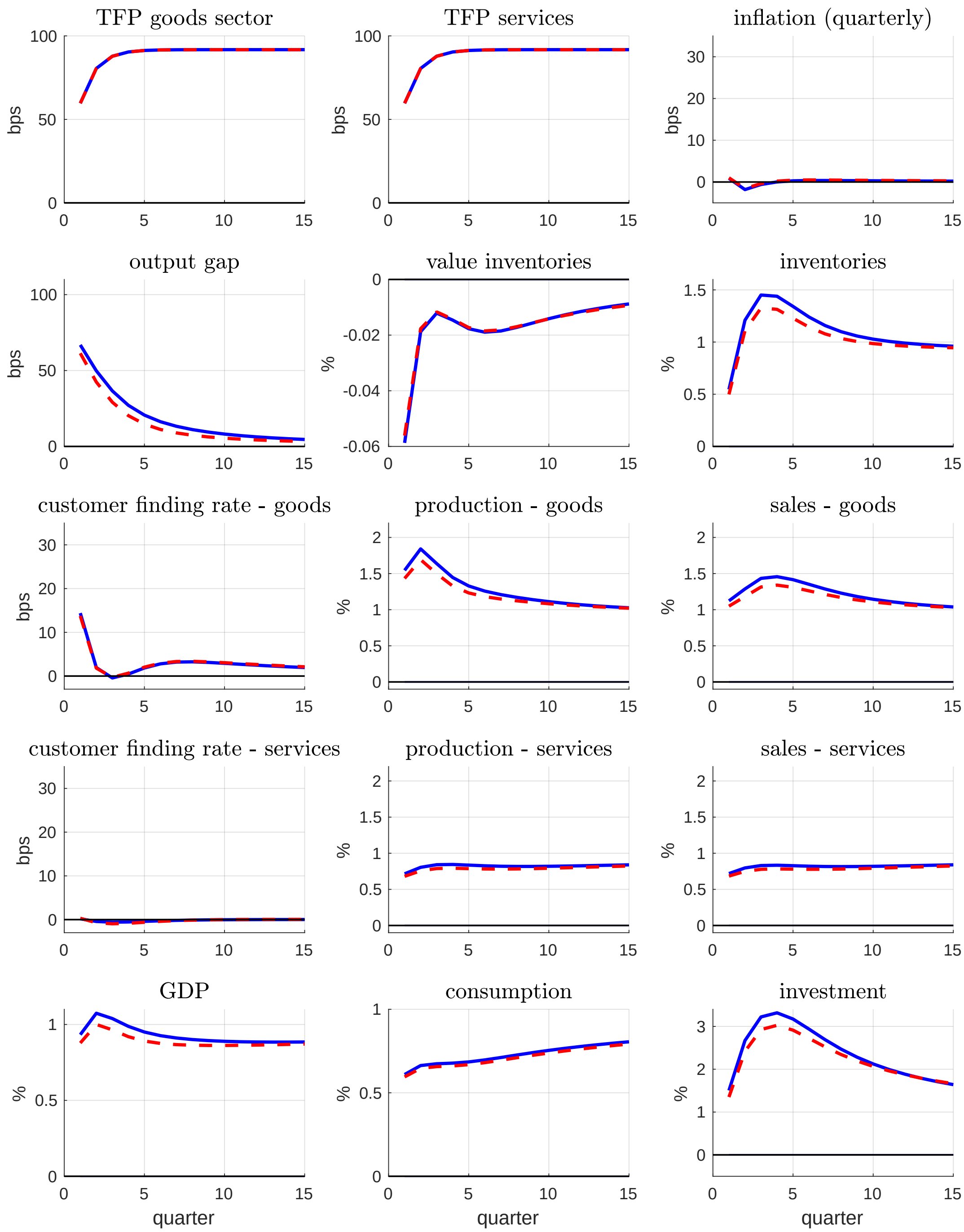
Although the effect is quantitatively small, the economy as a whole has become slightly more volatile. The reduction in the average sell fraction in the service sector has made the economy less efficient and the same employment level generates less value added. This is more noticeable for TFP shocks as documented in figure 21, although the quantitative impact of this change in parameter values is still very minor.

Figure 20: Monetary-policy shock; same average customer-finding rate across sectors



Notes. The blue/solid 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 benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

Figure 21: TFP shock; same average customer-finding rate across sectors



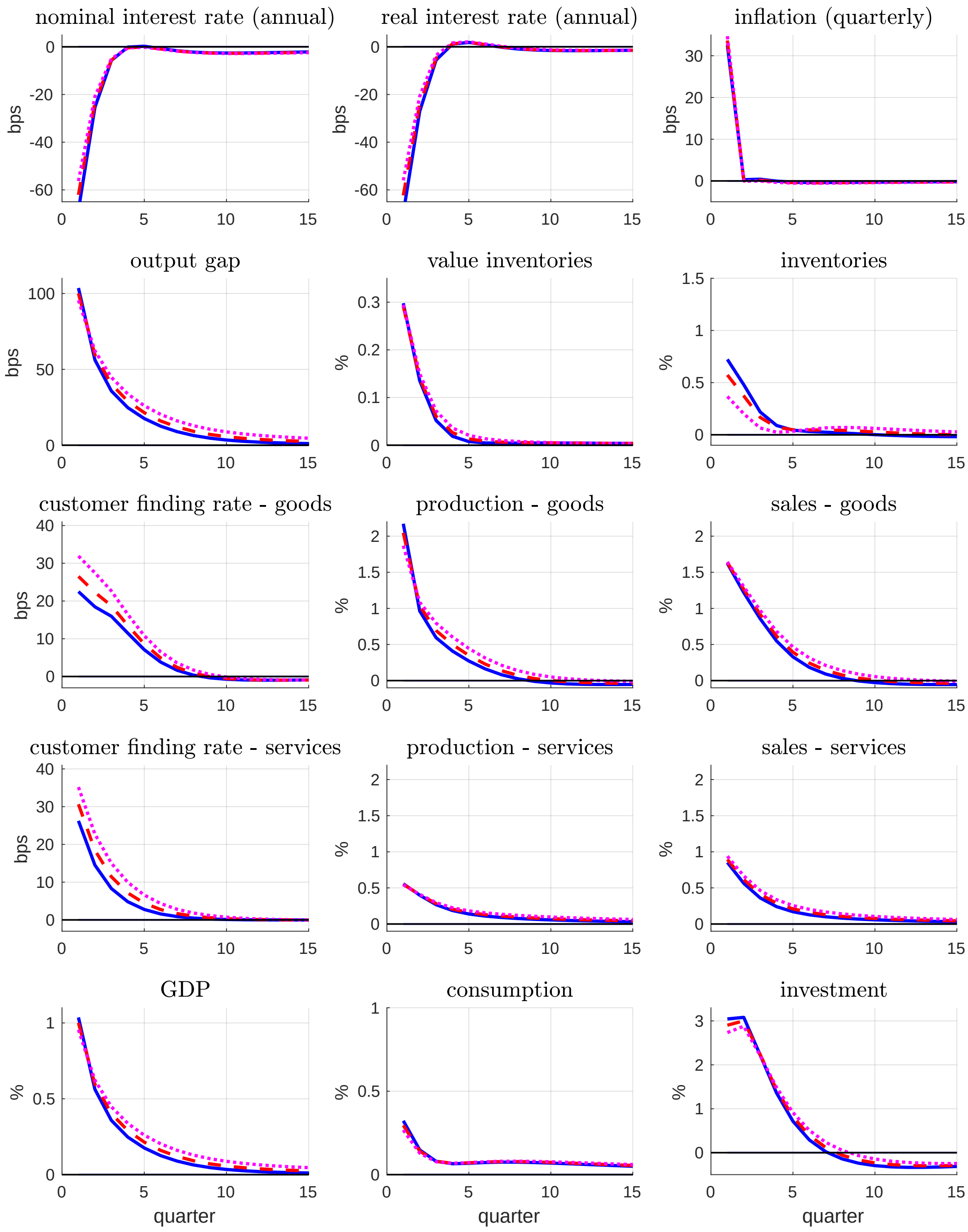
Notes. The blue/solid lines display IRFs of a TFP shock when the mean customer-finding rate is the same in the service sector is equal to 0.506 instead of 0.888 so equal to the one in the goods sector. Other parameter values are set equal to the benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

H.5 Alternative curvature in friction of selling services

Our moment-matching procedure is 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, is consistent with key inventory, production, and sales data. Results in the main text for the moment-matching parameter set are based on the assumption that the analogue parameter for the service sector, ν_s, takes on the same value. Figure 22 show the IRFs related to a monetary-policy shock for the benchmark (red/dashed) and two alternatives, namely $\nu_s = \nu_g - 0.1$ (magenta/dotted) and $\nu_s = \nu_g + 0.1$ (blue/solid). The results for the customer-finding rate are as expected. That is, a lower value of ν_s implies that the customer-finding rate, $\mu_s \theta_{s,t}^{1-\nu_s}$ responds more strongly to changes in tightness. This translates into a larger increase in the customer-finding rate which dampens the increase in inventories.

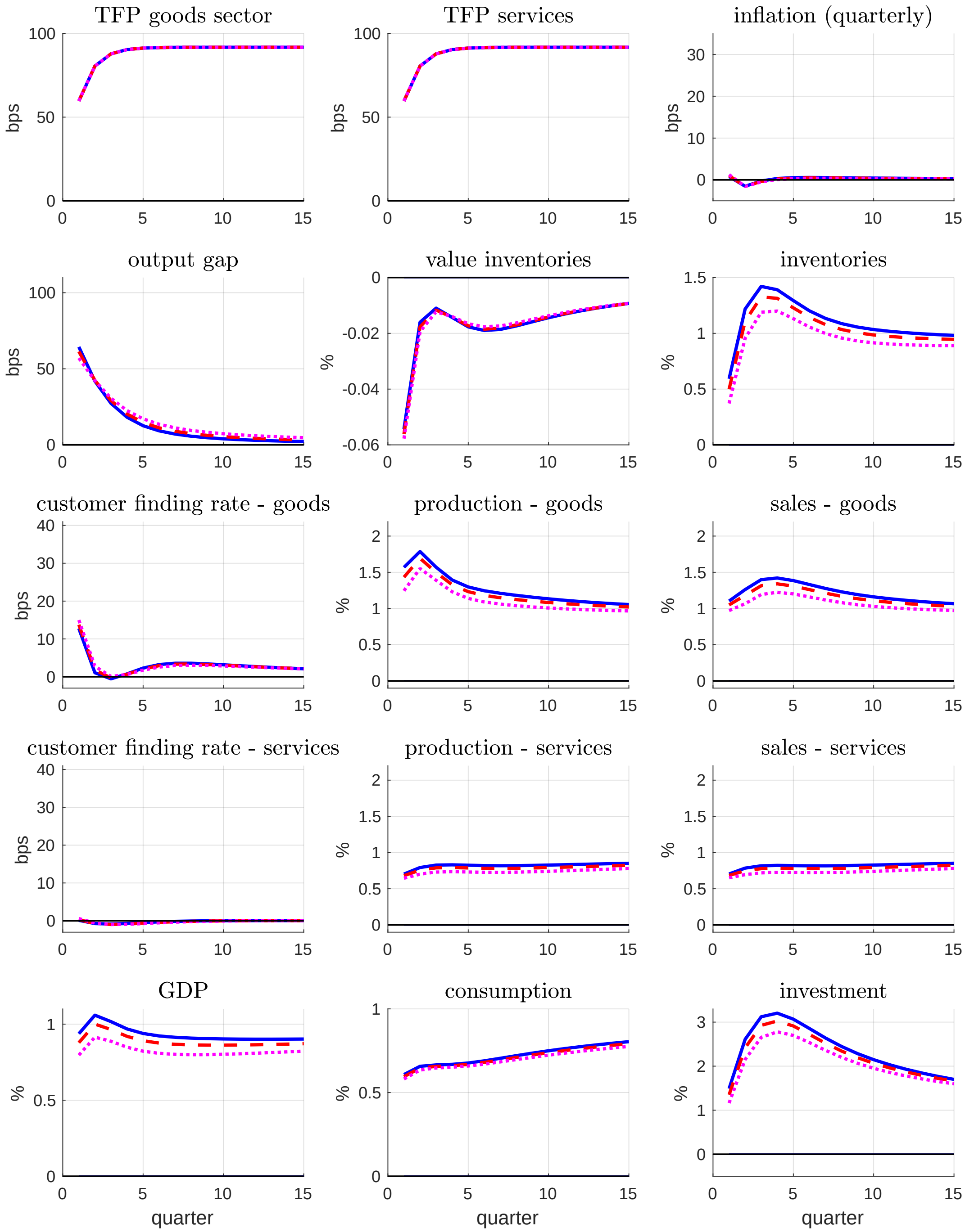
The corresponding results for a TFP shock are shown in figure 23. Results are very robust to changes in the value of ν_s. But there are some quantitative differences. The response of the customer-finding rate in the service sector is (even) smaller for the higher value of ν_s which goes together with stronger responses for the production and sales of services. Interestingly, these results spill over to the goods sector which also displays a smaller response of the customer-finding rate (12.8 instead of 15.0 basis points) and higher activity levels.

Figure 22: Monetary policy shock; different curvature parameters service sector friction



Notes. figure displays the IRFs of a monetary-policy shock for three different values of ν_s, the curvature parameter in the function that controls the friction of selling services. The blue/solid line corresponds to the case when $\nu_s = \nu_g + 0.1$ and the dotted line to the case when $\nu_s = \nu_g - 0.1$. Other parameter values are set equal to the benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

Figure 23: TFP shock; different curvature parameters service sector friction



Notes. figure displays the IRFs of a TFP shock for three different values of ν_s, the curvature parameter in the function that controls the friction of selling services. The blue/solid line corresponds to the case when $\nu_s = \nu_g + 0.1$ and the dotted line to the case when $\nu_s = \nu_g - 0.1$. Other parameter values are set equal to the benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

H.6 Alternative assumptions about inventory maintenance costs

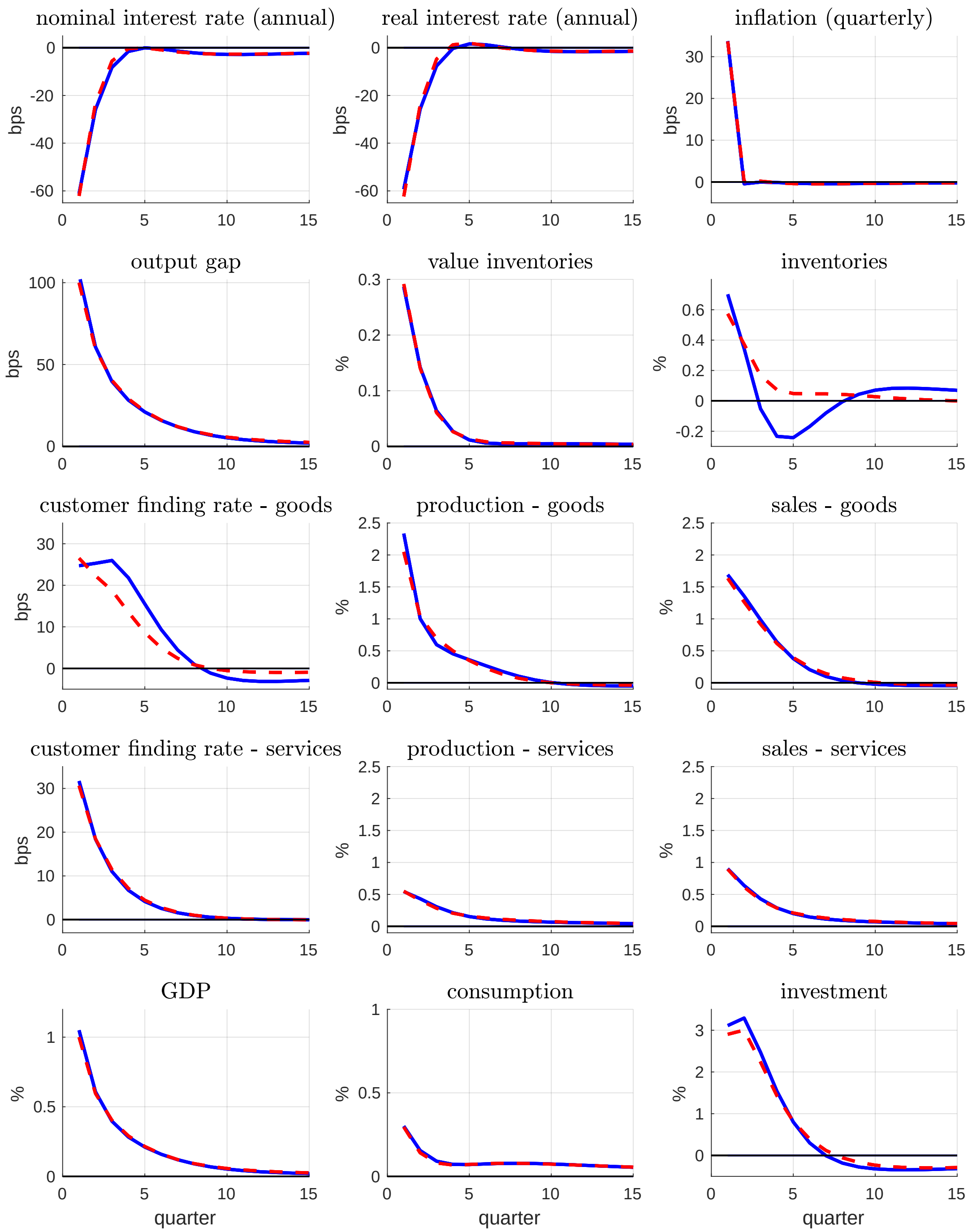
Our calibration of the maintenance cost parameter, η_x, is based on historical data. But such costs are likely to have become much smaller because of technological advances, especially in inventory planning. To study how such a reduction in costs affect model predictions, we consider a value for η_x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

Figures 24 and 25 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 $\lambda_{x,t}^f$ 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 the steady-state value of λ_x^f is likely to matter. Consistent with the analysis in section 3, we observe a stronger response in the goods-sector customer-finding rate when η_x is lower and the value of unsold goods is relatively more important. This dampens the increase in inventories somewhat.

In response to a monetary policy shock, we also find a somewhat stronger response of the customer-finding rate. Although inventories still display a strong response, initially, they now display a negative value for several periods. Of course, a re-calibration of ν_g would lower the response of the customer-finding rate and push the inventory IRF up.

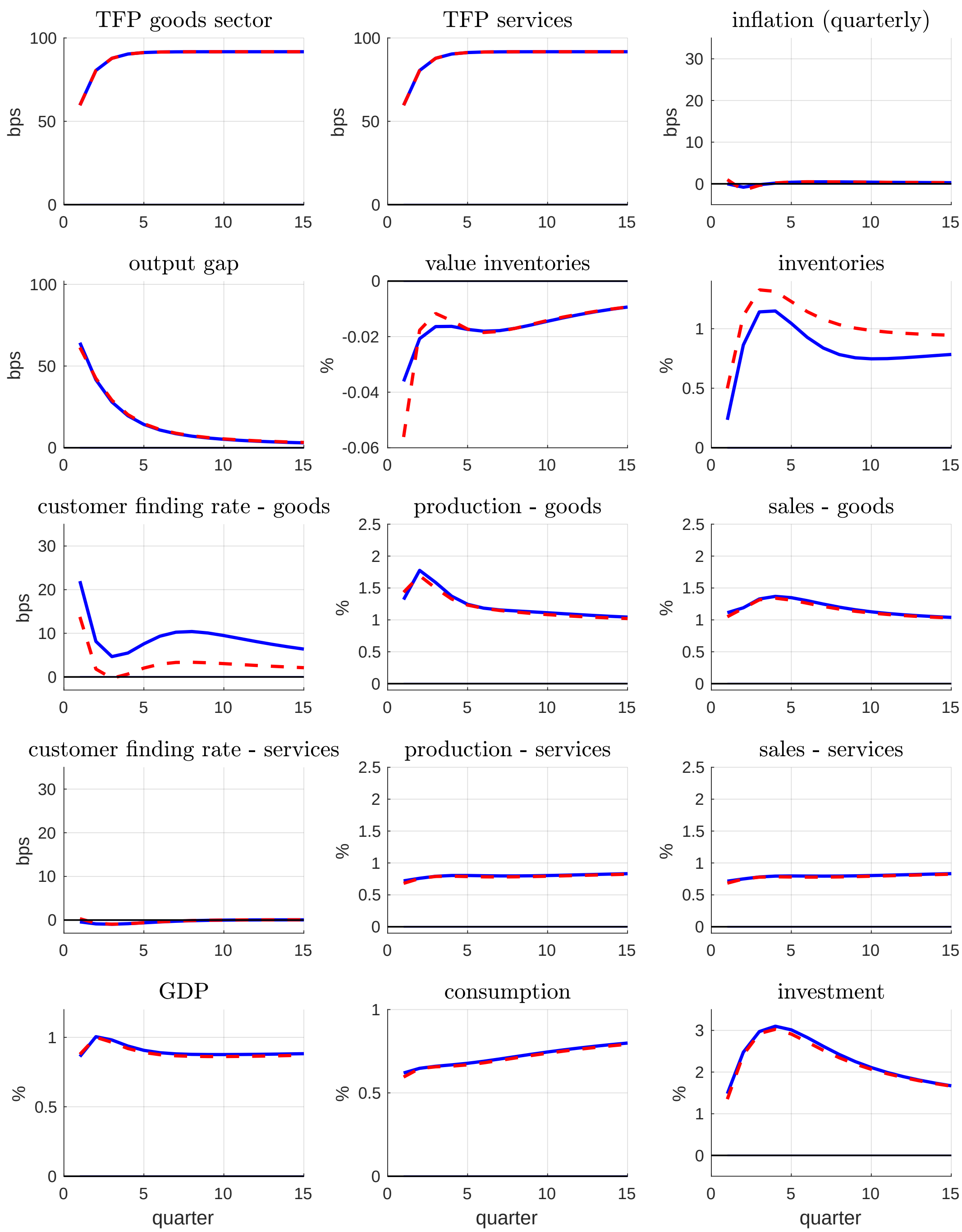
So the good news is that the predictions of our model remain valid even as inventory maintenance costs change.

Figure 24: Monetary-policy shock; lower maintenance costs inventories



Notes. The blue/solid lines display IRFs of a monetary-policy shock when the maintenance costs of inventories are only 20% of the benchmark value. Other parameter values are set equal to the benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.

Figure 25: TFP shock; lower maintenance costs inventories



Notes. The blue/solid lines display IRFs of a TFP shock when the maintenance costs of inventories are only 20% of the benchmark value. Other parameter values are set equal to the benchmark moment-matching values. The red/dashed lines display the IRFs based on benchmark moment-matching parameters used to generate figures 2 and 3.