Rationing Under Sticky Prices

Tom D. Holden, Deutsche Bundesbank[*] 26/09/2025

Abstract: If prices are sticky, following large shocks, firms would like to ration demand to avoid selling goods at a price below marginal cost. However, the standard assumption in solving sticky price models is that firms sell the entire quantity demanded at their price. This paper investigates the consequences of allowing firms to ration under sticky prices, in a continuous time model with idiosyncratic demand shocks and endogenous price rigidity. Allowing rationing massive reduces the welfare costs of positive trend inflation. The loss of variety caused by rationing becomes the main welfare cost of variations in inflation. Rationing helps the model match empirical results from both micro & macro data. It produces a convex, backwards-bending Phillips curve. While expansionary monetary policy increases the model's analogue of observed real GDP, it decreases the welfare relevant output aggregator.

Keywords: rationing, Phillips curve, inflation, New Keynesian.

JEL codes: E31, E52, D45

Latest version and slides available at https://www.tholden.org/.

[*] Address: Mainzer Landstraße 46, 60325, Frankfurt am Main, Germany.

E-mail: thomas.holden@gmail.com. Website: https://www.tholden.org/.

The views expressed in this paper are those of the author and do not represent the views of the Deutsche Bundesbank, the Eurosystem or its staff.

The author would like to thank Klaus Adam, Florin Bilbiie, Justin Bloesch, Corina Boar, Richard Dennis, Keshav Dogra, Fiorella De Fiore, Jordi Galí, Mike Golosov, Sebastian Graves, Refet Gürkaynak, Marcus Hagedorn, Matteo Iacoviello, Cosmin Ilut, Callum Jones, Peter Karadi, Oleksiy Kryvtsov, Campbell Leith, Vivien Lewis, Olivier Loisel, Jochen Mankart, Jean-Baptiste Michau, Giovanni Nicolò, Tereza Ranošová, Ricardo Reis, Frank Schorfheide, Luminita Stevens, Harald Uhlig, Daniel Villar Vallenas, Henning Weber, Nils Wehrhöfer and Elisabeth Wieland for helpful comments and assistance.

1 Introduction

Economies worldwide ground to a halt under supply constraints in the early 2020s. Covid restrictions prevented many people from working. The Suez Canal was blocked by the ship Ever Given, preventing goods from reaching Europe from Asia. The Russian invasion of Ukraine led to the end of Russia's gas exports to Europe. These supply constraints were accompanied by high inflation, stockouts in some consumer goods (Cavallo & Kryvtsov 2023b) and delivery delays for goods such as cars.[1] As I write this in 2025, it looks like we are heading for yet another supply crunch in the U.S., caused by Trump's tariffs. This will inevitably be accompanied by increased stockouts again.

Stockouts and delivery delays are both forms of rationing, as they are ultimately a choice of the supplier. While supply disruptions increase marginal costs, still marginal costs remain finite. If a firm desperately wanted a production input while the Suez Canal was blocked, they could have put it in an airplane instead. If car manufacturers really wanted microchips delivered in 2022 rather than 2023, they could have offered semiconductor manufacturers high enough prices to get them to switch from producing chips for GPUs and mobile phones. Instead, they sold consumers the substitute good "car-in-2023" instead of the good "car-in-2022" they were ideally looking for.

Firms had another choice though. They could have raised prices. If prices had risen with the increase in marginal costs, then all goods would have remained available. Consumers who were prepared to pay could still have obtained the goods they wanted. Thus, sticky prices seem essential for supply disruptions to lead to stockouts or other forms of rationing.

Rationing is also common in normal times. Over 10% of all consumer goods are out of stock in normal times in the U.S., according to the evidence of Cavallo & Kryvtsov (2023b). This paper builds a dynamic model of rationing under sticky prices to understand the implications of rationing for monetary policy

---

[1] See e.g. https://www.thedrive.com/news/new-cars-piling-up-at-german-port-will-mean-longer-wait-for-us-buyers, https://www.cnbc.com/2021/05/07/chip-shortage-is-starting-to-have-major-real-world-consequences.html, or https://www.thisismoney.co.uk/money/cars/article-11831443/How-long-wait-new-car-delivered-revealed.html.

and the broader macroeconomy.

Unfortunately, essentially all prior dynamic models of sticky prices have been solved under the simplifying assumption that firms satisfy all demand at their posted price, even if that results in them selling at a price below marginal cost. This is true for both Calvo, Rotemberg and menu-cost approaches to modelling price rigidities. While tractable, this seems deeply implausible.

If a firm cannot adjust their nominal price, then their real price will be declining over time. A lower real price implies higher demand for their good, and so higher sales. With short run decreasing returns to scale, higher sales in turn means higher real marginal costs. So, the firm's real price is declining, while their real marginal cost is increasing. If the price remains fixed, eventually the firm's marginal cost will equal or exceed its price. No firm would want to continue to sell their good in this state. Instead, they would ration demand, only selling up to the quantity at which price equals marginal cost.

Does rationing really matter in practice? I will present new evidence that supports the ubiquity of rationing, but a simple back of the envelope calculation is also instructive. Perhaps one reason the prior literature has been happy to rule out rationing is that they have had the following misleading calculation in mind: *"Mark-ups are 10%, inflation is 2%, prices are updated at least once per year, real prices will not hit marginal cost."* But this is not the right calculation when firms face short run decreasing returns to scale. The estimates of Abraham et al. (2024) using data from Belgian firms imply that around $\frac{2}{5}$ of all labour and intermediate inputs are fixed at annual frequency, implying a total share of fixed inputs in production, α, of around $\frac{3}{5}$.[2] Thus, firm marginal costs are roughly proportional to $y^{\frac{\alpha}{1-\alpha}} = y^{\frac{3}{2}}$, where y is their output. Meanwhile,

[2] From Table 3, column (3) or (4) of Abraham et al. (2024), we see that we cannot reject that the share of all capital inputs that are fixed at annual frequency is 100% at a 1% (or lower) significance level, and we cannot reject that the shares of all labour or intermediate inputs includes that are fixed at annual frequency are both 40% at a 5% (or lower) significance level. (I err on the side of high fixed shares as fixed shares would be higher in higher frequency data.) Ignoring intermediates, with a capital share of $\frac{1}{3}$, this gives a total fixed share in production of $1 \times \frac{1}{3} + \frac{2}{5} \times \frac{2}{3} = \frac{3}{5}$. Boehm, Flaaen & Pandalai-Nayar (2019) find that intermediates are perfect complements to other inputs, so given their fixed share $\frac{2}{5}$ is less than $\frac{3}{5}$, I am justified in taking $\frac{3}{5}$ as the overall fixed share.

firms face demand proportional to $\left(\frac{p}{P}\right)^{-\epsilon}$, where p is their nominal price, P is the price level, and $\epsilon \approx 10$ in standard calibrations. So, if the price level increases by 2% (over a year, say), but the firm's nominal price stays fixed, then firm sales increase by 2% × 10 = 20%, which means marginal costs increase by $\frac{3}{2} \times 20\% = 30\%$. A 30% rise in marginal costs is more than enough to erode standard calibrations of firm level mark-ups. Thus, we should expect firms with one year old prices to be rationing.

This simple calculation is likely to understate firms' incentives to ration. Firstly, firms face high frequency fluctuations in demand. At times of high demand, marginal costs will be high, making rationing more tempting. Secondly, inflation can be much higher than 2%. It was 7% over the period from June 2021 to June 2022 in the U.S..[3] With 7% inflation and a fixed nominal price, it would take less than a quarter for marginal costs to have risen by 25%. Thirdly, demand is also growing over time due to aggregate income growth. Even holding wages fixed, 2% demand growth implies a 3% increase in marginal costs over a year. Finally, marginal costs are increasing over time due to irregular replacement of broken machines, and imperfect maintenance. Firms face non-convex adjustment costs in new investment (Cooper & Haltiwanger 2006; Khan & Thomas 2008) and maintenance rates are below depreciation rates (Kabir, Tan & Vardishvili 2024).[4] Thus, in between installations of new machines, capital stocks will be declining and marginal costs will be increasing.[5]

---

[3] https://fred.stlouisfed.org/series/PCEPI, 100 × change in logarithms over a year.

[4] Kabir, Tan & Vardishvili (2024) find that annual maintenance expenditure is around 6.2% of the value of the capital stock, while their (caveated) estimate of annual depreciation is around 9.4% of the value of the capital stock.

[5] How much on average capital stocks are decreasing over the life of a price will depend on just how often firms make significant capital investments, and how correlated these times are with price change times. It seems natural to suppose that any firm going to the significant trouble of installing new machines would also take the much smaller step of updating its price at the same time. Using data extracted from Figure 1 of Cooper & Haltiwanger (2006) reveals that in any year, around 57% of all firms do not invest enough to cover depreciation (6.9% in their data) plus 2% growth, and 49% of all firms do not invest enough to cover just depreciation. This suggests that firms increase their capital stock less often than they update prices. (The price adjustment estimates of Blanco et al. (2024b) imply around 24% of firm prices last for at least a year.) This is consistent with net investments being accompanied by price changes.

Price adjustment. A natural question is why firms with price at or below marginal cost do not just update their price to restore their mark-up.[6] In a Golosov Lucas (2007) menu cost economy, with constant returns and no micro or macro uncertainty, it is clear that paying the menu cost is always optimal for a firm with price equal to marginal cost. Waiting t weeks to change prices is dominated by changing prices now but setting a higher price such that in t weeks your real price is what you would have set had you waited.

However, any micro or macro uncertainty can destroy this result. Once there is uncertainty, it can be optimal to tolerate rationing or a price below marginal cost in order to avoid repeated price changes. For example, suppose your price is currently too low, but you expect aggregate or idiosyncratic productivity to improve soon (perhaps due to mean reversion), at which time your current price will be comfortably above marginal cost. Modern menu cost models rely on random menu costs (Dotsey, King & Wolman 1999) and free price change opportunities (Nakamura & Steinsson 2010) to match the micro data, so in these models there is an even greater incentive to temporarily tolerate rationing or a price below marginal cost. Maybe now the menu cost is high, but next period it could be much lower. At the risk of oversimplifying, modern menu cost models work hard to look more like a Calvo model, and in a Calvo model, many firms get stuck with price below marginal cost. For example, price change hazard functions appear flat (Klenow & Kryvtsov 2008; Nakamura & Steinsson 2008; Klenow & Malin 2010), so old prices (with a higher probability of being lower than marginal cost) are no more likely to be adjusted than new prices.

Moreover, models that allow for rationing will be consistent with much lower price adjustment frictions than models that do not. In a model without rationing, firms risk substantial losses if they do not adjust their price. To match the data in which despite this, they do not adjust their price, the price adjustment frictions must be large. In a model with rationing though, the firm

Adam & Weber (2019) stress declining firm marginal costs over the firm life cycle. This is not inconsistent with rising marginal costs over the life of a price if productivity improvements (perhaps brought about by the installation of new machines) are accompanied by price changes.

[6] A version of this point was made in Barro (1977).

can always guarantee weekly positive profits no matter how old its price is, thus smaller adjustment frictions are needed to match the observed low frequency of price adjustment. If your prior is that adjustment frictions, like menu costs, are small, then you should place greater posterior weight on models with rationing, such as the one I present in this paper.

My model. My basic model is in continuous time, with Calvo-type price rigidity,[7] but I endogenize the price change arrival rate to capture the varying price adjustment rates we see in the data. Firms in the model are owned by conglomerates, who can choose the arrival rate of price adjustment opportunities for the firms they manage, following Blanco et al. (2024b). This provides aggregate state dependence, while matching the flat adjustment hazard functions found by Klenow & Kryvtsov (2008), Nakamura & Steinsson (2008) and Klenow & Malin (2010).

At all points in time, firms can freely choose their production. Optimally, they will meet demand if they can do so with price above marginal cost, otherwise they will produce up to the point at which price equals marginal cost, rationing demand. To smooth out the kink introduced by this decision, I assume that firms face demand shocks that are independent both across firms and over time. With a carefully chosen density, the model then admits aggregation with a finite dimensional state vector, permitting analytic results and easy simulation. Whereas the standard model without rationing is unstable at high inflation levels, the model with rationing is robustly stable, with reasonable behaviour even under extreme shocks.

For consumers, rationing is random. When they arrive at the shop, if they are lucky, the firm has recently restocked, and they can purchase their entire demand. If they are unlucky though, the shelves are empty, and they leave without any units of the good. Thus, when average stockout rates are high, consumers will be consuming a restricted set of varieties. They find this costly due to their love of variety. Love of variety is a standard feature of preferences in macro models, and is well supported empirically (Broda & Weinstein 2006;

[7] Early continuous time New Keynesian models were developed by Posch, Rubio-Ramírez & Fernández-Villaverde (2011), (2018).

2010).

Random rationing seems a reasonable first approximation to the rationing we see in reality. As an alternative, I could have modelled sellers as capping the quantity they would sell to any individual consumer. While during Covid we saw some shops placing quantity limits on a few items, this is clearly not the main way shops ration.[8] Furthermore, my numerical experiments suggest that sales-capped rationing generates an implausibly steep Phillips curve, giving further support to random rationing as the dominant variety.

Results preview. I stress three main results from my model. Firstly, I show that rationing leads to a convex, backward-bending Phillips curve relationship between output and inflation, in line with the evidence surveyed in the next section. The curvature emerges from the fact that high inflation erodes markups, leading to high rationing. This causes a drop in the range of varieties consumed, which reduces the welfare relevant measure of output. However, observed real GDP does not capture the gains from variety in consumption, so the relationship between the model's analogue of observed real GDP and inflation is upwards sloping.

Secondly, I show that the welfare relevant measure of output declines following expansionary monetary policy shocks (while the model's equivalent of observed real GDP increases). This follows from the fact that the welfare relevant Phillips curve is backward bending. Thus, at least in the vicinity of the model's steady state, "stimulative" monetary policy is undesirable, making monetary policy less useful for stabilisation.

Thirdly, I show that rationing drastically reduces the welfare costs of positive trend inflation, compared to a model without rationing. Without rationing, many firms are stuck with prices below marginal costs. And they sell large quantities at this price, precisely because it is low. This produces

[8] I would not want to argue that there is no role for sales-capped rationing. For example, if goods are semi-durable, consumers shop frequently, and they face storage constraints preventing them from hoarding, then the result can look a lot like quantity-capped rationing. Without storage constraints though, in equilibrium the result must look a lot like random rationing, with the unlucky consumers finding shop shelves empty at the same time their pantry is also empty.

substantial misallocation. Allowing rationing caps the welfare cost of having an old price (rationing firms have price equal to marginal cost), removing most of the costs of price dispersion.

The model also matches a range of further empirical evidence presented in the next section, despite only introducing one new parameter over a comparable model without rationing. This evidence includes new evidence from supermarket scanner data on sales over the life of a price, sectoral evidence on price adjustments and stockouts, as well as evidence on the macro response to monetary shocks.

Prior literature. Important early work examining rationing with sticky prices includes Barro & Grossman (1971), Drèze (1975) and Svensson (1984). Barro & Grossman look at outcomes in a one period model when both aggregate output and aggregate labour may be rationed. Drèze examines at equilibrium existence with the possibility of rationing in an Arrow-Debreu setup with price inequality constraints. Svensson looks at rationing in a dynamic monetary model with a single good. A little more recently, Corsetti & Pesenti (2005) worked in a proto-New Keynesian framework with prices set one period in advance, and were careful to restrict their model's shocks to ensure the absence of rationing.

I am aware of three papers that look at rationing in a New Keynesian setting. Huo & Ríos-Rull (2020) and Gerke et al. (2023) look at the rationing of labour supply that comes from sticky wages, but omit rationing on the price side. These papers both have infinite dimensional state vectors, which makes it challenging to understand the details of their mechanisms, and they rely on quantity-capped rationing rather than random rationing. Hahn (2022) looks at rationing under price rigidity in the steady state of a New Keynesian model with Calvo price frictions. While he is able to derive some interesting comparative statics results, his approach is not tractable for looking at dynamics, so he provides no dynamic results. Without idiosyncratic shocks, he also cannot hope to produce an empirically reasonable path of output over the life of a price, even in steady state, as we will see in Subsection 2.1. Finally, he only looks at quantity-capped rationing, not the more plausible random rationing specification.

Other papers have employed modelling devices to capture some of the effects of rationing, without the technical difficulties involved in properly modelling rationing. Michaillat & Saez (2015) use search and matching to smooth out the distinct rationed or non-rationed zones of the Barro & Grossman (1971) model. Liu (2025) extends this with endogenous match efficiency. While search is natural in labour markets, it clearly plays a lesser role in goods markets. Additionally, Kharroubi and Smets (2024) assume that firms stop producing altogether if their marginal cost is higher than their permanently fixed price at the demanded quantity.

Another relevant strand of the literature looks at stockouts in models of inventories. Contributions include Alessandria, Kaboski & Midrigan (2010), Kryvtsov & Midrigan (2013) and Bils (2016). They demonstrate the importance of inventory dynamics for a variety of macro questions. However, in all of these papers, firms always meet demand if they have stock available, even if the marginal value of that stock to the firm is greater than the price at which they can sell the good. Thus, in these models too, firms would like to ration in some circumstances. For the sake of tractability, my model will not feature inventories, but combining inventories and rationing is a promising avenue for future research.

2 Empirical evidence for rationing

The previous arguments suggest rationing should be widespread. In line with this, Cavallo & Kryvtsov (2023b) found that around 11% of all goods in their data were out of stock in 2019, using daily web-scraped data from 17 large retailers in the U.S..[9] This may understate the true prevalence of rationing, since retailers can encourage consumers to substitute away from particular goods by, for example, lowering their ranking in search results, worsening their position on physical shelves, or by reducing advertising. Encouraging such substitution helps reduce stockouts, which may provide a reputational benefit for the store. "Shrinkflation" may also mask rationing. If I want 400 grams of cereal, but it is

[9] Mean of "AOOS" over the period from the replication data of Cavallo & Kryvtsov (2023b) (Cavallo & Kryvtsov 2023a).

now in sold in 375-gram boxes, I am unlikely to buy two boxes.

Unsurprisingly, Cavallo & Kryvtsov (2023b) found that stockouts increased massively during the Covid pandemic. More interestingly though, they found that in 2022 (January to July), still 22% of goods were out of stock in the U.S..[9] By 2022 many of the direct effects of Covid had subsided, but inflation was picking up worldwide. Thus, in line with the story of the model I will present, it appears that high inflation leads to increased rationing.

Rationing will also help to explain the observed convexity of the Phillips curve. For pre-Covid evidence on this, see, for example, Kumar & Orrenius (2016), Babb & Detmeister (2017) or Forbes, Gagnon & Collins (2022). The fact that the inflation of 2022 was not accompanied by huge output booms provides "natural experiment" evidence in further support of such convexity. Under rationing, the Phillips curve is convex as when demand is already high, further demand increases just lead to increased rationing, rather than increased output.

I will present three pieces of new evidence to support my model of rationing. Firstly, I will show that expansionary monetary shocks increase a measure of rationing and other shortages, providing causal evidence that high inflation leads to increased rationing. Secondly, I will show that high rates of price adjustment one month are followed by reductions in stockouts the next month, establishing a link between price frictions and rationing. Firms with new prices have high mark-ups, reducing the chance they ration. Thirdly, I will show that quantities sold are concave in the age of the good's price. Thus, goods with young prices experience relatively high output growth, while goods with older prices experience relatively low output growth. In the absence of rationing, we would expect sales to be growing exponentially in the age of a price, as inflation erodes the real price. If instead sales are concave in price age, then something must be pushing down the sales of goods with old prices. Rationing is the prime candidate. With high sales leading to high marginal costs, the firm prefers to limit sales to keep marginal costs below their price.

2.1 Rationing following monetary shocks

Figure 1 plots the impulse response to an expansionary monetary policy shock in the U.S., identified following the informationally robust approach of

Miranda-Agrippino & Ricco (2021). The specification closely follows that of their Figure 3, with the following changes:

1. I use Brave-Butters-Kelley RGDP (Brave, Cole & Kelley 2019; Brave, Butters & Kelley 2019) instead of industrial production. This monthly measure of real GDP produces results that are more readily comparable with the model.
2. I use PCEPI in place of CPI. PCEPI is closer to the model's welfare relevant price index.
3. I add the cumulation of the median rate of price changes, excluding sales, from the data of Montag & Villar (2025).[10,11] Deviations of this variable show the number of price changes per firm that happened following the shock that would not have happened had the shock not arrived.
4. I add the Caldara-Iacoviello-Yu U.S. Shortages Index (Caldara, Iacoviello & Yu 2025). This measure is constructed by searching U.S. newspapers for relevant keywords ("rationing", "shortage", "scarcity" or "bottleneck"),[12] which must occur in conjunction with an economics-related term. Only articles mentioning the "U.S.", "U.S.A.", "America", etc. or the name of a U.S. city are included in the measure I use here.

Let me stress that for the variables contained in both my specification and that of Miranda-Agrippino & Ricco (2021), my results are not qualitatively different to theirs. Likewise, RGDP in my specification follows a similar path to IP in theirs, and PCEPI in mine follows a similar path to CPI in theirs.

As you would expect, an expansionary monetary shock raises real GDP and prices, and lowers unemployment. More surprising is the response of the shortages index. The 1% shock causes this index to rise by about 30% on impact, before it decays back to trend over the next two years. This provides causal evidence that increases in demand lead to higher shortages. Since shortages are

[10] I thank the authors for providing me this data.

[11] Let f_t be the weighted median frequency of price changes across "Entry Level Item" (ELI) after adjusting for sales, at t (using each ELI's CPI weight). Then this series is $-\sum_{s=0}^{t}\log(1-f_t)$. Using the median frequency gives robustness to cross-sectional heterogeneity, and is in line with standard practice. The data construction is detailed in Montag & Villar (2025), following Nakamura et al. (2018).

[12] In private correspondence, Matteo Iacoviello reports that adding "stockout" or "delivery delay" to the keywords list made practically no difference to the index.

always ultimately a choice of the seller (if they employed enough resources, they could virtually always provide the good), this is direct evidence of higher rationing. Recall also that “rationing” is one of the keywords used to construct the index.

Additionally, following the shock, almost 5% of firms immediately make an additional price change. After two years about 10% of firms have made a price change they would not otherwise have made. This demonstrates the importance of endogenizing price adjustment frequencies, and will serve as a calibration target for the endogenous price adjustment in my model.

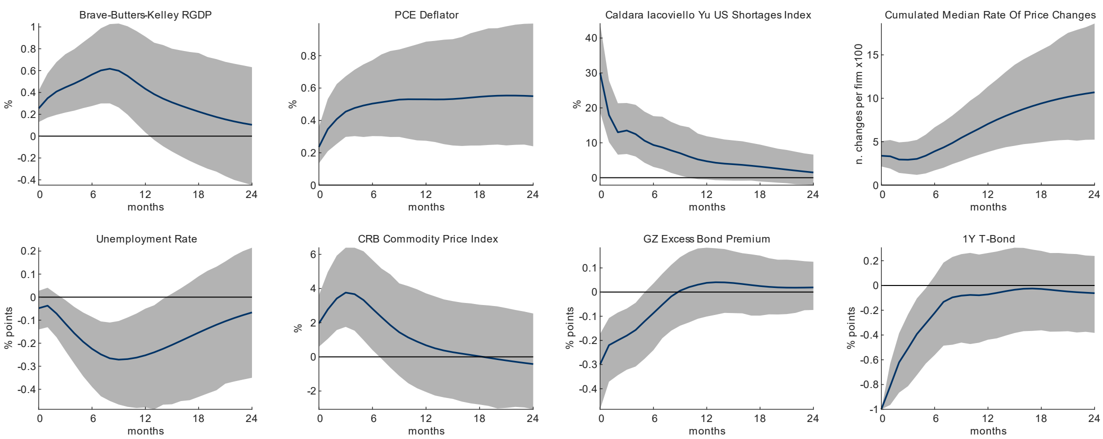


Figure 1: Impulse response to a 1% monetary policy shock. Informationally robust specification from Figure 3 of Miranda-Agrippino & Ricco (2021), but estimated using Brave-Butters-Kelley RGDP (Brave, Cole & Kelley 2019; Brave, Butters & Kelley 2019) in place of IP, PCEPI in place of CPI, and with both the cumulated median rate of price changes, excluding sales (Montag & Villar 2025),[11] and the Caldara-Iacoviello-Yu U.S. Shortages Index (Caldara, Iacoviello & Yu 2025) also in the VAR. 95% credible bands highlighted.

I also want to highlight one further property of these IRFs that is also present in Miranda-Agrippino & Ricco (2021), as well as in other high-frequency identified monetary VARs, like Bauer & Swanson (2023): the shock causes an immediate jump up in both the price level and output. Calvo or Rotemberg type models of price rigidity can never generate a jump in the price level following a shock, as the price level is a state variable in these models. By contrast, in a model with rationing the welfare relevant price level is no longer a state variable, since jumps in the level of rationing cause jumps in the weight placed on goods with relatively higher prices, due to rationing of lower price goods.

One caveat is in order though. At the most disaggregated level, the PCEPI index uses price indices constructed by the BLS (for the CPI), which are a geometric mean of gross product price growth for most goods.[13] Thus, the continuous time counterpart of PCEPI or CPI can only jump if a positive measure of firms adjust their price, which never happens in Calvo type models, and will not even happen under our model of endogenous price flexibility. However, there are many reasons for scepticism about the accuracy of monthly macro data, so I will target movements three months from the initial shock in my calibration. The fact that the welfare relevant price index does jump will increase the rate of change of measured PCEPI, as firms pass through cost increases.

Additionally, in practice, the BLS data collectors have to deal with many missing prices, for which they then use imputation based on price growth of other items. Stockouts (from rationing) are a major source of missing prices. Thus, the BLS-CPI imputation procedure ascribes average price changes from non-rationed goods to rationed goods. Since rationed goods are less likely to have changed price, this produces greater aggregate inflation than under a fixed weight index after an inflationary shock, bringing the CPI and PCEPI indices closer to the welfare relevant price index.

While menu cost models can potentially generate a jump in prices after a shock without rationing, cleanly identified monetary policy shocks are small and so are unlikely to lead to large amounts of price resetting. For example, Blanco et al. (2024a) calibrate a menu cost model to match both micro price data and the aggregate response of the price change frequency to inflation, and find that 1% increases in the money supply are mostly absorbed by output, not prices, in the short run. Similarly, in Figure 1 we saw that only around 5% of firms change prices immediately following a hypothetical 1% monetary shock. For those price changes to generate the observed 0.3% rise in the price level, the

[13] https://www.bls.gov/opub/hom/cpi/calculation.htm#price-relatives. Only a few goods use a Laspeyres formula.

adjusting firms would have to be increasing prices by over 50%.

2.2 Price adjustment and stockouts

I now want to present more disaggregated evidence linking price adjustment and stockouts, to support the mechanisms of my model. For this, I use the data provided in the replication package of Cavallo & Kryvtsov (2023). They have monthly web-scraped data on stockouts from online stores in 7 countries (Canada, China, France, Germany, Japan, Spain, USA), covering 37 product categories, at roughly the three-digit level. For my exercise, I just use their pre-Covid data from 2019, as my model will not capture the direct effects of Covid disruptions.

I run the following regression specification:

$$y_{i,j,t} = \beta x_{i,j,t-1} + \alpha_{i,j} + \gamma_{i,t} + \delta_{j,t} + \varepsilon_{i,j,t}.$$

Here, i indexes countries, j indexes product categories and t indexes time in months (February to December 2019). On the left, I have $y_{i,j,t}$, the average stockout rate for a country-category-month in percent. On the right, I have $x_{i,j,t-1}$, the one-month lag of the average annualized rate of price adjustment for a country-category-month multiplied by 100, as well as a fully saturated set of fixed effects.

I find $\hat{\beta} = -0.155$, with a standard error of 0.044 and a p-value for the null of $\beta = 0$ of below 0.001. (Standard errors are three-way clustered (Cameron, Gelbach & Miller 2011), with cluster indices (i,j), (i,t), (j,t).) Thus, a one percentage point increase in the proportion of firms adjusting prices one month is associated with a decrease in stockouts the next month of $12 \times 0.155\% = 1.86\%$ (given fixed effects).[14] This provides direct evidence for the role of price adjustment in stockouts. Firms with new prices choose high initial mark-ups to protect themselves against future inflation. These high mark-ups also reduce

[14] Of course, we cannot be sure this is entirely causal. But reverse causation would produce the opposite sign. If firms anticipated future stockouts, they would increase their price today. Perhaps some shock directly causes both an increase in price adjustment and a persistent reduction in stockouts (not via the price adjustment channel), but it is hard to imagine what this could be, particularly given the fixed effects.

the chance the firm rations.

2.3 Evidence from scanner data

I now want to disaggregate even further, and to present evidence from micro scanner data of the prevalence of rationing. By looking directly at quantities sold, I can measure not only stockouts, but also less direct forms of rationing, such as changes in product placement. I use data from a former chain of Chicago supermarkets called "Dominick's Finer Foods", made freely available by the Kilts Center for Marketing at Chicago Booth.[15] The data covers the period 1989 to 1994, during which time annual PCEPI inflation was between around 2% and around 5%.[16] While newer data is always preferable, supermarket practices have not changed so dramatically in the last thirty years, and the use of open data ensures replicability.

The data records the prices and quantities sold of products from 29 broad categories,[17] from 93 stores, over 399 weeks. The 29 broad categories are further refined into 86 narrower categories.[18] Where possible, I use the item code information provided by the supermarket to match goods which are newer versions of former products. For each good, at each store, I drop the following observations:

- Those with price equal to the first price observed for the good. (We do not observe the start of the first price spell, so we cannot construct price age for those observations.)
- Those with price equal to the final price observed for the good. (Maybe the

[15] https://www.chicagobooth.edu/research/kilts/research-data/dominicks.

[16] https://fred.stlouisfed.org/series/PCEPI.

[17] Analgesics, Bath Soap, Bathroom Tissues, Beer, Bottled Juices, Canned Soup, Canned Tuna, Cereals, Cheeses, Cigarettes, Cookies, Crackers, Dish Detergent, Fabric Softeners, Front-end-candies, Frozen Dinners, Frozen Entrees, Frozen Juices, Grooming Products, Laundry Detergents, Oatmeal, Paper Towels, Refrigerated Juices, Shampoos, Snack Crackers, Soaps, Soft Drinks, Toothbrushes, Toothpastes.

[18] The split into narrower categories was unavailable for "Refrigerated Juices", so I allocated goods in this category into the following eleven narrower categories based on their description field: Orange Juice, Orange Drinks, Apple Juice and Cider, Cranberry Juices and Cranberry Juice Blends, Other Fruit/Vegetable Juices, Fruit Punch and Mixed Fruit Drinks, Lemonade, Iced Tea, Dairy-based Drinks and Shakes, Puddings, Colored Easter Eggs. The CSV file giving the allocation of items to categories is contained in the replication materials for this paper.

good disappeared due to changing tastes, in which case the concavity in sales over the span of the final price could reflect demand, not supply.)

- Those with price less than the cumulative maximum price for the good at that store. (This ensures we are only looking at sales after a price rise, not a price cut. It would be unsurprising if sales initially increased after a price cut. We want to pick up the increase in sales after a price rise coming from inflation eroding real prices. This filter also takes out sales during which demand may be distorted by different advertising levels.)
- Those occurring at the same time as a change in price, or the week after a missing observation (which could have hidden a change in price). (Keeping observations the period of a price change could be a source of endogeneity, due to the same demand shock influencing both quantities sold and the decision to change prices.)
- Those with a price age greater than four years. (There are relatively few prices that ever last so long. Including them would reduce estimation reliability due to the use of average output over the life of a price in my regression specification.)

I estimate the following linear model for quantity sold as a function of the age of the price:

$$\frac{y_{i,j,k,l,t}-y_{i,j,k,l,t-1}}{\bar{y}_{i,j,k,l}} = \beta_{A(i,j,k,l,t)} + \alpha_{i,j,k} + \gamma_{i,j,t} + \delta_{i,k,t} + \sigma^{(1)}_{i,j,A(i,j,k,l,t)}\sigma^{(2)}_{i,j,k}\sigma^{(3)}_{i,j,t}\sigma^{(4)}_{i,k,t}\varepsilon_{i,j,t}.$$

Here, i indexes the 92 narrow categories, j indexes the 93 stores, k indexes the 10,166 products, l indexes product prices (there are 947,660 total product-price pairs; the same product receives a different l in two periods if its price differs), and t indexes the 398 weeks. $A(i,j,k,l,t)$ is the age in weeks of the i,j,k,l price at t,[19] and $y_{i,j,k,l,t}$ is the number of units sold of that item that week. $\bar{y}_{i,j,k,l}$ is the average of $y_{i,j,k,l,t}$ over the life of the price.

The left-hand side of this specification gives a measure of sales growth that

[19] For goods without missing observations, new prices start with age one (assuming that the price change occurred at the end of the previous week), so the first observed age will be two, as one week is dropped due to the price change. For goods with some missing observations, I renormalize ages so that the first included observation is age two.

is robust to the presence of zeros in $y_{i,j,k,l,t}$. Working in differences, not levels, ensures consistency even when products experience $I(1)$ demand shocks, due to entry or exit of substitute products, for example. On the right-hand side, $\beta_{A(i,j,k,l,t)}$ gives age fixed effects, our prime variable of interest. This means we allow sales growth to be an arbitrary function of price age. $\alpha_{i,j,k}$ gives category-store-product fixed effects, to absorb differing trends across products and stores. $\gamma_{i,j,t}$ gives category-store-time fixed effects to mop up changes in demand for specific category types in specific locations at specific times (think of the demand for candy around Halloween, concentrated in family neighbourhoods). $\delta_{i,k,t}$ gives product-time fixed effects to mop up changes in demand for particular products coming from seasonality or changing tastes. Importantly, these fixed effects mop-up product life-cycle effects. For example, the $\delta_{i,k,t}$ term will absorb the initial rise in popularity of a new product, and the decline in popularity, of old products. I model heteroskedasticity in the residual by category-store combined (separately) with age, product and time, as well as by product-time. This substantially improves the efficiency of my estimates.

After differencing, I am left with 21,474,126 observations. Estimating the model on these observations by feasible generalized least squares gives the estimates summarized in Figure 2. This figure plots $100 \sum_{a=3}^{\mathrm{AGE}} \beta_a$ as a function of AGE in the black solid line. I.e., it plots the average level of sales over the life of a price. Due to the fixed effects, this is only identified up to a linear trend, so the plot is normalized so that the impact is zero for age 2 weeks and age 130 weeks (2.5 years). The shaded area gives a 95% confidence band, and the dotted lines give a 99% band. These are constructed with four-way clustered standard errors (Cameron, Gelbach & Miller 2011), with clusters indexed by category-store combined (separately) with age, product and time, as well as by product-time. Hence, the clusters have indices $(i, j, A(i, j, k, l, t))$, (i, j, k), (i, j, t) and (i, k, t) respectively. The first grouping allows for heterogeneity in the effects of age across categories and stores. The second allows for arbitrary correlation across time for the residuals from any particular product in a particular store. The third allows for time-varying correlation between the residuals of all products in a category and store, capturing substitution across products within a category for example. The fourth allows for time-varying correlation between the residuals of the same product at different stores, capturing substitution between stores.

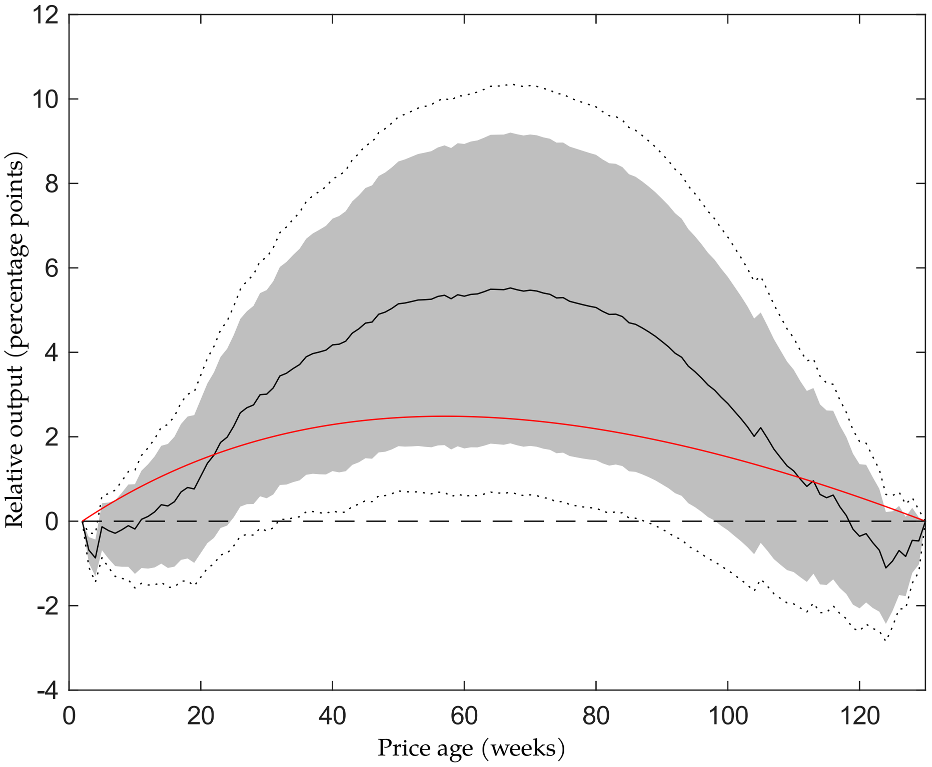


Figure 2: Average output over the life of a price ($100 \sum_{a=3}^{\text{AGE}} \beta_a$).

The effect is identified up to a linear trend, so I normalize to zero at ages 2 and 130 weeks. The black solid line gives the estimates. The black dashed line is at zero for reference. The shaded area gives a 95% confidence band, and the dotted lines give a 99% band. The red line gives the prediction of the model from Section 3.[20]

We see that relative to the normalization, sales grow for around a year, before starting to decline. This is consistent with firms rationing demand for products with old prices. While a good's nominal price is fixed, its real price is declining, leading to higher sales. But with decreasing returns, higher sales mean higher marginal costs. Eventually, marginal costs are higher than prices, so the firm rations demand. Under rationing, sales are a decreasing function of real price (due to decreasing returns again), so sales are then declining in price age. The result is that sales are a concave function of price age, as we see here. Without any rationing, log-sales would be linear in firm age (as long as demand

[20] In the notation of equation (3) from Section 3 this is $100 \log y_{\tau,t}$, detrended to be 0 at 2 and 130 weeks.

is roughly isoelastic), so after normalizing we would not find any statistically significant difference from zero.[21]

The red line in Figure 2 plots the prediction of the model I will present in Section 3. This is not a calibration target of the model, so it is reassuring that the model broadly matches the data. You might be surprised by the small size of the predicted effect of price age on sales, though. After all, if the price elasticity of demand is -10, then with 2% inflation, over a year sales should have increased by 20% without rationing. If firms started rationing from week 52 on, then that would still imply a normalized peak impact of 12%.[22] However, my model is one in which firms face idiosyncratic demand shocks that are independent across time. These shocks mean that for any price age, a firm's expected sales is a mix of their sales when their demand shock is high, so they ration, and their sales when their demand shock is low, so they meet demand. This reduces the sensitivity of average sales to price age, matching the data. A model without idiosyncratic demand shocks would predict implausibly high average sales growth for young prices.

Let me end this section by stressing that this paper is not about a fundamentally different model of price rigidity. Rather, it is about relaxing a simplifying assumption previously used in solving such models. As such, whatever evidence supports your favourite sticky price model will probably also support the same model extended to allow for rationing.

3 The model

I will now present my base model of rationing under sticky prices. Throughout the paper, I stick to the convention that upper case letters denote aggregate variables, while lowercase Latin letters denote firm specific variables. The model is in continuous time, with time measured in years throughout. Letters without time subscripts denote steady-state values. For simplicity, there is no aggregate uncertainty: I will only look at the impact of prior probability zero "MIT" shocks.

---

[21] Standard calibrations of Kimball (1995) demand can also generate concavity in sales over the life of a price, without any rationing.

[22] Something like this would be true in the Hahn (2022) model, for example.

3.1 Firms and aggregators

The model will feature a continuum of firms of measure one. Firms are only able to adjust their price when they are hit by a shock from a non-homogenous Poisson process. In particular, price change opportunities arrive at time t with rate $\lambda_t > 0$, where $\int_{-\infty}^{t} \lambda_v \, \mathrm{d}v = \infty$ for all t. As a result, the time t density of firms that last adjusted their price at time τ is given by $\lambda_\tau e^{-\int_\tau^t \lambda_v \, \mathrm{d}v}$. Note that, as required for this to be a density, $\int_{-\infty}^{t} \lambda_\tau e^{-\int_\tau^t \lambda_v \, \mathrm{d}v} d\tau = \int_{-\infty}^{t} \frac{\mathrm{d}}{\mathrm{d}\tau} e^{-\int_\tau^t \lambda_v \, \mathrm{d}v} d\tau = 1 - e^{-\int_{-\infty}^{t} \lambda_v \, \mathrm{d}v} = 1$. I index firms with the time at which they last updated their price τ, so this density will appear frequently.

Firms will face demand shocks that are independent both across firms, and across time t. This means that over even an arbitrarily small interval of time, a firm will face all possible values of the demand shock.[23] I write $y_{\zeta,\tau,t}$ for the output of a firm at time t, that last updated their price at time τ, that is hit by a demand shock of level $\zeta \in [0,1]$. Demand shocks ζ will be drawn from a $\mathrm{Beta}(\theta, 1)$ distribution, where $\theta > 0$, meaning they have probability density function $g(\zeta) = \theta \zeta^{\theta - 1}$. This implies the mean of the demand shock is $\frac{\theta}{\theta+1} \approx 1 - \frac{1}{\theta}$ and the variance of the demand shock is $\frac{\theta}{(\theta+1)^2(\theta+2)} \approx \frac{1}{\theta^2}$. Demand shocks are essential for tractability as they smooth out the kink introduced by the rationing decision. This particular distribution for the demand shocks is needed for the model to have a finite dimensional state. θ is the only non-standard parameter in the entire model, apart from the two parameters determining the costs of price adjustment (to be introduced shortly). I will calibrate θ to match the evidence from Cavallo & Kryvtsov (2023b) that around 11% of all goods are rationed in normal times.

I also introduce purchaser-good-time-specific shocks denoted by $\psi \in [0,1]$ that control whether a given purchaser can buy a given good. I write $y_{\psi,\zeta,\tau,t}$ for the consumption of a buyer with shock ψ, at time t, of the good produced by a firm that last updated their price at time τ, that is hit by a demand shock of level ζ. I will focus on the special case in which ψ is uniformly distributed and:

[23] This is no more mathematically problematic than having shocks that are independent across a continuous measure of firms in a discrete time model. Obviously, I will be careful not to attempt to measure any unmeasurable quantity!

$$y_{\psi,\zeta,\tau,t} = \begin{cases} y^*_{\zeta,\tau,t}, & \psi \le \bar{\psi}_{\zeta,\tau,t} \\ 0, & \psi > \bar{\psi}_{\zeta,\tau,t} \end{cases},$$

where $y^*_{\zeta,\tau,t}$ is the buyer's purchase when not rationed, and $\bar{\psi}_{\zeta,\tau,t}$ gives their probability of not being rationed. In this special case, total output of a given firm will satisfy:

$$y_{\zeta,\tau,t} = \int_0^1 y_{\psi,\zeta,\tau,t}\,\mathrm{d}\psi = y^*_{\zeta,\tau,t}\bar{\psi}_{\zeta,\tau,t}.$$

We can either think of firms as choosing their total sales, with $\bar{\psi}_{\zeta,\tau,t}$ adjusting to ensure $y_{\zeta,\tau,t} = y^*_{\zeta,\tau,t}\bar{\psi}_{\zeta,\tau,t}$, or we can think of firms as choosing the probability that a consumer will not be rationed, $\bar{\psi}_{\zeta,\tau,t}$, with total sales adjusting. The former will be more convenient in practice.

The aggregate good Y_t is produced by a competitive industry of "aggregators" with access to the technology:

$$Y_t = D^{-\frac{\epsilon}{\epsilon-1}}\left[\int_{-\infty}^{t} \lambda_\tau e^{-\int_\tau^t \lambda_v\,\mathrm{d}v} \int_0^1 \zeta g(\zeta) \int_0^1 y_{\psi,\zeta,\tau,t}^{\frac{\epsilon-1}{\epsilon}}\,\mathrm{d}\psi\,\mathrm{d}\zeta\,\mathrm{d}\tau\right]^{\frac{\epsilon}{\epsilon-1}}. \qquad (1)$$

Here, $\epsilon > 1$ is the elasticity of substitution across varieties, and $D = \frac{\theta}{\theta+1}$ is a scale factor chosen to ensure that if $y_{\psi,\zeta,\tau,t}$ is one for all ψ, ζ and τ, then $Y_t = 1$. This aggregator is essentially the standard Dixit-Stiglitz one. The only changes are the weighting by the density of firms that last updated at time τ, and the inner integrals over the possible draws of the demand shock and the rationing shock. Demand is higher for varieties receiving a higher draw of ζ. To understand the ζ integral, you should think of there being a positive measure of firms that last updated their price at time τ. Of these infinitely many firms, a density proportional to $g(\zeta)$ of them will receive demand shock ζ at time t. The interpretation of the ψ integral is similar.

Like normal, aggregators choose their input quantities to maximize their profits:

$$P_t Y_t - \int_{-\infty}^{t} \lambda_\tau e^{-\int_\tau^t \lambda_v\,\mathrm{d}v} p_\tau \int_0^1 g(\zeta) \int_0^1 y_{\psi,\zeta,\tau,t}\,\mathrm{d}\psi\,\mathrm{d}\zeta\,\mathrm{d}\tau,$$

where P_t is the aggregate price, and p_τ is the price of all varieties that last updated their price at time τ.[24] In doing so, they face the supply constraints $y_{\psi,\zeta,\tau,t} \le 0$ for all ψ, ζ, τ and t with $\psi > \bar{\psi}_{\zeta,\tau,t}$. The first order conditions of this

[24] I am assuming here that all firms updating their price at the same time will choose the same price. This will be true in equilibrium.

problem imply that firms face the demand constraint:

$$y_{\zeta,\tau,t} \leq y^*_{\zeta,\tau,t} := \left(\frac{D}{\zeta}\frac{p_\tau}{P_t}\right)^{-\epsilon} Y_t. \tag{2}$$

Demand places an upper bound on firm sales, not a lower bound.

From inspecting this problem, it may not be obvious whether aggregators make zero profits in equilibrium. It turns out that with this random rationing specification, aggregators do indeed make zero profits. This is because the aggregator has constant returns to scale conditional on $\bar{\psi}_{\zeta,\tau,t}$.[25]

Firms produce output using the decreasing returns to scale production function:

$$y_{\zeta,\tau,t} = v^{1-\alpha}_{\zeta,\tau,t}, \quad \text{where } v_{\zeta,\tau,t} = A_t l_{\zeta,\tau,t}.$$

Here, $v_{\zeta,\tau,t}$ is their effective labour input, $l_{\zeta,\tau,t}$ is their actual labour input, $A_t > 0$ is aggregate productivity and $\alpha \in (0,1)$ is the fixed share in production. The use of letter v for the effective labour input anticipates the extended model in which $v_{\zeta,\tau,t}$ will be a bundle of variable inputs. Labour will be supplied at the aggregate wage W_t. For convenience, I define the wage of effective labour by $\widehat{W}_t := \frac{W_t}{A_t}$.

Firms' flow of real production profits is given by:

$$o_{\zeta,\tau,t} = \frac{p_\tau}{P_t} y_{\zeta,\tau,t} - \widehat{W}_t v_{\zeta,\tau,t} = \frac{p_\tau}{P_t} v^{1-\alpha}_{\zeta,\tau,t} - \widehat{W}_t v_{\zeta,\tau,t}.$$

I assume firms can choose how much to produce at all points in time, after learning their demand shock. Thus, $v_{\zeta,\tau,t}$ (or $l_{\zeta,\tau,t}$) is a choice variable for the firm. Note that no matter the price p_τ, $o_{\zeta,\tau,t} = 0$ if $v_{\zeta,\tau,t} = 0$, but:

$$\frac{\mathrm{d}\tilde{o}_{\zeta,\tau,t}}{\mathrm{d}v_{\zeta,\tau,t}} = (1-\alpha)\frac{p_\tau}{P_t} v^{-\alpha}_{\zeta,\tau,t} - \widehat{W}_t \to \infty$$

as $v_{\zeta,\tau,t} \to \infty$. Thus, the firm can always ensure positive production profits by choosing a small enough $v_{\zeta,\tau,t}$. A small enough $v_{\zeta,\tau,t}$ will also satisfy the firm's demand constraint, (2), and hence the firm will always make strictly positive profits, and will always choose $v_{\zeta,\tau,t} > 0$ so $y_{\zeta,\tau,t} > 0$.

---

[25] With quantity-capped rationing, aggregators do make profits. In that case, the presence of sales limits mean that the aggregators face decreasing returns to scale, and so positive aggregator profits are consistent with perfect competition. Another way to see this is to note that the true price index would integrate over a sum of the actual price of goods, and the Lagrange multipliers on the sales limits, but aggregators do not "pay" the Lagrange multipliers, resulting in profit.

Firms choose $v_{\zeta,\tau,t}$ to maximize $o_{\zeta,\tau,t}$ subject to the demand constraint, (2). In Appendix A I show that this leads them to choose:

$$v_{\zeta,\tau,t} = \min\left\{\left[\left(\frac{D}{\zeta}\frac{p_\tau}{P_t}\right)^{-\epsilon} Y_t\right]^{\frac{1}{1-\alpha}}, \left(\frac{p_\tau}{P_t}\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{1}{\alpha}}\right\},$$

so:

$$y_{\zeta,\tau,t} = \min\left\{\left(\frac{D}{\zeta}\frac{p_\tau}{P_t}\right)^{-\epsilon} Y_t, \left(\frac{p_\tau}{P_t}\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{1-\alpha}{\alpha}}\right\}.$$

In both of these expressions, the first term in the curly brackets gives the outcome without rationing, in which firms meet demand. In this case, price is above marginal cost, and sales are decreasing in the good's real price. The second term in the curly brackets in these expressions gives the outcome with rationing. In this case, price equals marginal cost, and sales are increasing in the good's real price. Note that the firm can calculate their maximum output in advance of the realisation of the shock. Thus, rationing does not require the firm to possess implausible amounts of information.

High values of ζ mean higher demand, and so make rationing more likely. To be specific, define:

$$\bar{\zeta}_{\tau,t} := D\left(\frac{p_\tau}{P_t}\right)^{1+\frac{1-\alpha}{\epsilon\alpha}}\left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{1-\alpha}{\epsilon\alpha}} Y_t^{-\frac{1}{\epsilon}},$$

then the firm will ration at least some buyers if $\zeta > \bar{\zeta}_{\tau,t}$, and the firm will never ration if $\zeta < \bar{\zeta}_{\tau,t}$. In particular, in equilibrium, a buyer's probability of not being rationed when visiting a firm that last updated their price at τ with demand shock ζ is:

$$\bar{\psi}_{\zeta,\tau,t} = \min\left\{1, \left(\frac{\bar{\zeta}_{\tau,t}}{\zeta}\right)^{\epsilon}\right\}.$$

High values of $\bar{\zeta}_{\tau,t}$ mean that rationing only takes place with extreme draws of the demand shock, whereas low values of $\bar{\zeta}_{\tau,t}$ mean rationing is likely. Increases in aggregate demand Y_t reduce $\bar{\zeta}_{\tau,t}$, increasing the chance of rationing. Likewise, when effective wages $\widehat{W}_t$ are high, so marginal costs are high, rationing is likely. Finally, note that having a high real price makes rationing less likely.

In the limit as $\lambda_t \to \infty$ for all t, the model tends to one with quasi-flexible prices. In this limit, firms continuously adjust their prices, but still set prices at t before the realisation of their time t demand shock. I show in Appendix A that firms still ration with positive probability in this limit (i.e., $\bar{\zeta}_{\tau,t} \leq 1$), as long as $\theta \leq \frac{\alpha\epsilon-1}{1-\alpha}\epsilon$, which will hold in any reasonable calibration. Thus, we should also expect $\bar{\zeta}_{\tau,t} \leq 1$ when $\lambda_t < \infty$ and prices are sticky, meaning there is rationing for at least some values of the demand shock. In all numerical exercises I will check that $\bar{\zeta}_{\tau,t} \leq 1$ for all τ and t.

Returning to the general case with $\lambda_t < \infty$, and assuming that $\bar{\zeta}_{\tau,t} \leq 1$, a firm's expected output before the demand shock is realized is:[26]

$$\begin{aligned} y_{\tau,t} &:= \int_0^1 y_{\zeta,\tau,t}\, g(\zeta)\, \mathrm{d}\zeta \\ &= \left(\frac{1-\alpha}{\widehat{W}_t}\frac{p_\tau}{P_t}\right)^{\frac{1-\alpha}{\alpha}} - \frac{\epsilon}{\theta+\epsilon} D^\theta Y_t^{-\frac{\theta}{\epsilon}} \left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{\theta+\epsilon}{\epsilon}\frac{1-\alpha}{\alpha}} \left(\frac{p_\tau}{P_t}\right)^{\theta+\frac{\theta+\epsilon}{\epsilon}\frac{1-\alpha}{\alpha}}. \end{aligned} \tag{3}$$

This has a part that is increasing in the good's real price and a part that is decreasing. The combination of the two gives log-concavity in $\frac{p_\tau}{P_t}$, generating the concave log-sales over the life of a price that I previously plotted in the red line of Figure 2.[27]

Again, assuming that $\bar{\zeta}_{\tau,t} \leq 1$,[28] a firm's expected profits before the realization of the demand shock is given by:

$$\begin{aligned} o_{\tau,t} &:= \int_0^1 o_{\zeta,\tau,t}\, g(\zeta)\, \mathrm{d}\zeta \\ &= \alpha \left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{1-\alpha}{\alpha}} \left(\frac{p_\tau}{P_t}\right)^{\frac{1}{\alpha}} \\ &\quad - \frac{\epsilon}{\theta+\epsilon}\frac{\epsilon\alpha}{(1-\alpha)\theta+\epsilon} D^\theta \left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{\theta+\epsilon}{\epsilon}\frac{1-\alpha}{\alpha}} Y_t^{-\frac{\theta}{\epsilon}} \left(\frac{p_\tau}{P_t}\right)^{\theta+\frac{1}{\alpha}+\frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha}}. \end{aligned} \tag{4}$$

This is also log-concave in $\frac{p_\tau}{P_t}$.[29]

3.2 State dynamics and the short-run Phillips curve

The basic model will have two state variables, though calculating the

---

[26] See Appendix A for derivations of this and subsequent results.

[27] To see log-concavity in price, write this expression as $Ax^a - Bx^{a+b}$, where $x = \frac{p_\tau}{P_t}$, $A, a, B, b > 0$ and $A - Bx^b > 0$ (as $\bar{\zeta}_{\tau,t} \leq 1$). Then the second derivative of its logarithm is $-\frac{a}{x^2} - x^{-2}\left(A - Bx^b\right)^{-2} Bbx^b\left[(b-1)A + Bx^b\right]$. As long as $\theta > 1$, $b > 1$, so this is negative.

[28] I cover the $\bar{\zeta}_{\tau,t} > 1$ case in Appendix A.

[29] By an identical argument to that of Footnote 27.

model's analogue of observed real GDP will add another nine. However, all these state variables will take the same form:

$$X_{j,t} := \int_{-\infty}^{t} \lambda_\tau e^{-\int_\tau^t \lambda_v \, \mathrm{d}v} p_\tau^{\chi_{j,1}} \, \mathrm{d}\tau,$$

where $j \in \mathbb{Z}$ and $\chi_{j,1}$ is a constant to be defined.[30] This implies that:

$$\dot{X}_{j,t} = \lambda_t \big(p_t^{\chi_{j,1}} - X_{j,t} \big),$$

where, as usual, dots above variables denote time derivatives.

Total demand for the variable production input, effective labour, is given by:

$$V_t := \int_{-\infty}^{t} \lambda_\tau e^{-\int_\tau^t \lambda_v \, \mathrm{d}v} \int_0^1 v_{\zeta,\tau,t} \, g(\zeta) \, \mathrm{d}\zeta \, \mathrm{d}\tau.$$

Assuming $\bar{\zeta}_{\tau,t} \le 1$ for all τ and t, I show in Appendix A that:

$$V_t = -\frac{\epsilon}{(1-\alpha)\theta + \epsilon} D^\theta \left(\frac{1-\alpha}{\widehat{W}_t} \right)^{\frac{1}{\alpha} + \frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha}} Y_t^{-\frac{\theta}{\epsilon}} P_t^{-\chi_{1,1}} X_{1,t} + \left(\frac{1-\alpha}{\widehat{W}_t} \right)^{\frac{1}{\alpha}} P_t^{-\chi_{2,1}} X_{2,t}, \quad (5)$$

where $\chi_{1,1} := \theta + \frac{1}{\alpha} + \frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha}$ and $\chi_{2,1} := \frac{1}{\alpha}$. Labour market clearing implies $V_t = A_t L_t$, where L_t is the household's labour supply.

Next, evaluating the integrals in the definition of the aggregator Y_t, equation (1), implies:

$$1 = -\frac{\epsilon}{\theta + \epsilon} D^\theta \left(\frac{1-\alpha}{\widehat{W}_t} \right)^{\frac{\theta+\epsilon}{\epsilon}\frac{1-\alpha}{\alpha}} Y_t^{-\frac{\theta+\epsilon}{\epsilon}} P_t^{-\chi_{1,1}} X_{1,t} + \left(\frac{1-\alpha}{\widehat{W}_t} \right)^{\frac{1-\alpha}{\alpha}} Y_t^{-1} P_t^{-\chi_{2,1}} X_{2,t}, \quad (6)$$

where I again assume $\bar{\zeta}_{\tau,t} \le 1$ for all τ and t.[31]

Holding fixed the values of the two states, $X_{1,t}$ and $X_{2,t}$, equations (5) and (6) can be combined with labour market clearing ($V_t = A_t L_t$) and the household's labour first order condition (to be given) to produce four equations in five unknowns (V_t, L_t, $\widehat{W}_t$, Y_t and P_t). The set of points satisfying these equations in (Y_t, P_t)-space gives the model's short-run Phillips curve.

Figure 3 plots the resulting short-run Phillips curve, under the model's baseline calibration which I will describe shortly. This figure answers the following question. Suppose that for $t < 0$, $P_t = \exp(\pi t)$, meaning inflation was constant at π, and suppose all state variables were at steady state at time 0. Then, suppose that at time 0, an unexpected monetary shock caused the price level to jump to P_0 from 1, where it would have been had no shock arrived. How

[30] The subscript ",1" anticipates the fact that other powers will enter this integral in the extended model.

[31] Again, proven in Appendix A.

does Y_0 vary with P_0, assuming that $P_t = P_0 \exp(\pi t)$ for $t \geq 0$? I should stress that since the welfare relevant price level P_t is not equal to the model's analogue to PCEPI, this plot cannot easily be compared to the results of Figure 1. I will perform a careful comparison of the model to Figure 1 in Section 4.

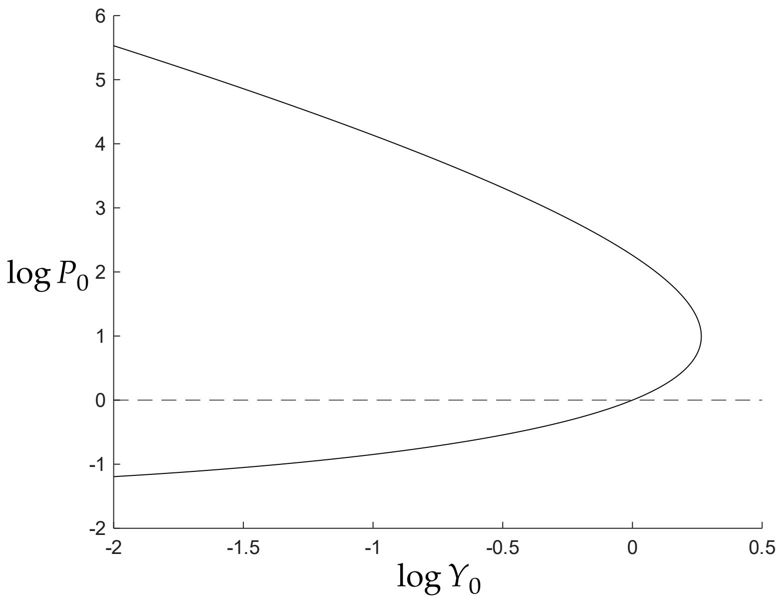


Figure 3: The model's short-run Phillips curve (solid line), and the short-run Phillips curve without rationing (dashed line). Percent deviation from steady state.

If $P_t = \exp(\pi t)$ for $t < 0$, with all state variables at steady state at time 0, how does Y_0 vary with a jump in P_0, assuming inflation continues at π after time 0?

We see that the model's short-run Phillips curve is convex and backwards-bending. Expansionary monetary policy can produce a jump in prices by generating a jump in rationing, which tilts the weights of the welfare relevant price index away from goods with old (low) prices which are likely to ration. Large enough monetary expansions generate so much rationing that output falls. This result is completely independent of price setting, as it is an impact result, before prices have adjusted.

Without rationing, the equivalent of equation (6) relates P_t & $X_{-2,t}$ with $\chi_{-2} := -(\epsilon - 1)$.[32] Thus, since $X_{-2,t}$ cannot jump, neither can P_t. Effectively, without rationing, the price level is a state variable. Thus, the model without rationing can never generate a jump in the welfare relevant price level. Of

[32] See Appendix B.

course, we do not observe the welfare relevant price level in reality, so this should not lead us to discard the model without rationing on its own. However, we will see that the faster adjustment of the welfare relevant measure with rationing will speed the adjustment of the observed price level, helping explain the data.

The convexity and backward-bending of the short-run Phillips curve with rationing can be seen analytically from equation (6) in the special case in which wages are also fixed in the short-run. With wages fixed, totally differentiating equation (6) implies that:

$$\frac{\mathrm{d}\log P_t}{\mathrm{d}\log Y_t} = \alpha \frac{1-\mathcal{A}_t}{\left[1+\alpha\frac{\theta}{\theta+\epsilon}(\epsilon-1)\right]\mathcal{A}_t - 1}, \tag{7}$$

where:

$$\mathcal{A}_t := D^{\theta}\left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha}} Y_t^{-\frac{\theta}{\epsilon}} P_t^{-(\chi_{1,1}-\chi_{2,1})}\frac{X_{1,t}}{X_{2,t}} > 0.$$

Thus, $\frac{\mathrm{d}\log P_t}{\mathrm{d}\log Y_t}$ is positive for moderate values of $\mathcal{A}_t$ (a standard upwards sloping Phillips curve), but negative for extreme values (a doubly backwards-bending Phillips curve).

3.3 Price setting

Just as all of the model's state variables take a similar form, so to do all of the forward-looking expressions that appear in the first order condition for firms' optimal price. In particular, they all take the form:

$$z_{j,\tau} := \int_{\tau}^{\infty} e^{-\int_{\tau}^{t}(\lambda_v + r_v)\,\mathrm{d}v} D^{\omega_{j,1}} \widehat{W}_t^{\omega_{j,2}} Y_t^{\omega_{j,3}} P_t^{\omega_{j,4}}\,\mathrm{d}t,$$

for $j \in \mathbb{Z}$, and constants $\omega_{j,1}$, $\omega_{j,2}$, $\omega_{j,3}$ and $\omega_{j,4}$ to be defined. Here, r_t is the real interest rate at t. Differentiating the definition of $z_{j,t}$ implies it satisfies the following differential equation:

$$\dot{z}_{j,\tau} = -D^{\omega_{j,1}} \widehat{W}_\tau^{\omega_{j,2}} Y_\tau^{\omega_{j,3}} P_\tau^{\omega_{j,4}} + (\lambda_\tau + r_\tau) z_{j,\tau}.$$

Firms updating their price at a time τ choose p_τ to maximize their value over the life of the price:

$$\begin{aligned} o_\tau &:= \int_{\tau}^{\infty} e^{-\int_{\tau}^{t}(\lambda_v + r_v)\,\mathrm{d}v} o_{\tau,t}\,\mathrm{d}t \\ &= -\frac{\epsilon}{\theta+\epsilon}\frac{\alpha\epsilon}{(1-\alpha)\theta+\epsilon}(1-\alpha)^{-\omega_{1,2}} p_\tau^{-\omega_{1,4}} z_{1,\tau} + \alpha(1-\alpha)^{-\omega_{2,2}} p_\tau^{-\omega_{2,4}} z_{2,\tau}, \end{aligned}$$

where $\omega_{1,1} := \theta$, $\omega_{1,2} := -\frac{\theta+\epsilon}{\epsilon}\frac{1-\alpha}{\alpha}$, $\omega_{1,3} := -\frac{\theta}{\epsilon}$, $\omega_{1,4} := -\chi_{1,1} = -\left(\theta + \frac{1}{\alpha} + \frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha}\right)$,

$\omega_{2,1} := 0$, $\omega_{2,2} := -\frac{1-\alpha}{\alpha}$, $\omega_{2,3} := 0$, $\omega_{2,4} := -\chi_{2,1} = -\frac{1}{\alpha}$.[33] Thus, firms optimally set p_τ such that:

$$\epsilon\left[\frac{\epsilon}{\theta(1-\alpha)+\epsilon}-\frac{\epsilon-1}{\theta+\epsilon}\right](1-\alpha)^{\frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha}}p_\tau^{\theta+\frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha}}z_{1,\tau}=z_{2,\tau}.$$

In the quasi-flexible price limit with $\lambda_\tau \to \infty$, this implies they would set the price p_τ^{QF} with:

$$\frac{p_\tau^{\mathrm{QF}}}{P_\tau}=\left[\left[\frac{\epsilon^2}{\theta(1-\alpha)+\epsilon}-\epsilon\frac{\epsilon-1}{\theta+\epsilon}\right]^{-\frac{\alpha\epsilon}{\theta}}D^{-\alpha\epsilon}Y_\tau^{\alpha}\left(\frac{\widehat{W}_\tau}{1-\alpha}\right)^{1-\alpha}\right]^{\frac{1}{1+(\epsilon-1)\alpha}}.$$

For comparison, without rationing in the quasi-flexible price limit, firms would set the price p_τ^{QFNR}:[34]

$$\frac{p_\tau^{\mathrm{QFNR}}}{P_\tau}=\left[\left[(1-\alpha)\left(\frac{\epsilon}{\epsilon-1}\right)\frac{\theta+\epsilon}{\theta(1-\alpha)+\epsilon}\right]^{1-\alpha}D^{-\alpha\epsilon}Y_t^{\alpha}\left(\frac{\widehat{W}_\tau}{1-\alpha}\right)^{1-\alpha}\right]^{\frac{1}{1+(\epsilon-1)\alpha}}$$

The two expressions agree when $\alpha=\frac{\theta+\epsilon}{\theta+\epsilon^2}$. At this point, the derivatives of the ratio $\frac{p_\tau^{\mathrm{QFNR}}}{p_\tau^{\mathrm{QF}}}$ with respect to α, ϵ or θ are all zero, and the second derivatives of the ratio with respect to those variables are all positive. Thus, at least locally around $\alpha=\frac{\theta+\epsilon}{\theta+\epsilon^2}$, with quasi-flexible prices, firms set higher prices if they cannot ration than if rationing is allowed. This is intuitive. If rationing is not allowed, firms worry about making large losses if demand is very high. To protect against this, they set a higher price.

3.4 Price adjustment rate choice

I endogenize λ_t, broadly following Blanco et al. (2024b). This is important as I wish to analyse the effects of changing steady-state inflation, and it is not plausible to assume that λ_t remains fixed as the long-run inflation rate increases. Higher trend inflation should mean more frequent price adjustment.

To endogenize λ_t, I assume that all firms are owned by conglomerates, with each conglomerate owning countably many firms (still a measure zero subset of the set of all firms). Each conglomerate will choose the rate of price adjustment λ_t for the firms it owns, to maximize average firm value over its firms minus a price adjustment cost of $\frac{\kappa_1}{1+\kappa_2}(\max\{0,\lambda_t-\underline{\lambda}\})^{1+\kappa_2}$ labour units,

[33] See Appendix A for this derivation and those of the rest of the results in this Subsection. I continue to assume $\bar{\zeta}_{\tau,t}\le 1$ for all τ and t throughout.

[34] See Appendix B for the model without rationing.

where $\kappa_1, \kappa_2 > 0$ and $\underline{\lambda} \geq 0$. This cost function has the reasonable property that if there is little price adjustment ($\lambda_t \leq \underline{\lambda}$) then there are no costs, unlike the cost function chosen by Blanco et al. (2024b) which is positive when adjustment rates are small. If $\underline{\lambda} > 0$ it also captures the free price adjustments stressed by "Calvo plus" models (Nakamura & Steinsson 2010). I will set $\underline{\lambda}$ to the minimum observed annual price adjustment rate in the Montag & Villar (2025) data used in Figure 1, $\underline{\lambda} = 0.73$. (This is at least a consistent estimator of the quantity of interest, if not necessarily an efficient one.) I calibrate κ_1 to match the time series mean (1978/01 – 2024/08) of the cross-sectional weighted median rate of price adjustment observed in the same data, $\lambda = 1.48$.[35] And I will calibrate κ_2 to match the response of λ_t to monetary shocks observed in Figure 1.

The derivation of the conglomerate's first order condition is a little involved, so I confine it to Appendix A, but the first order condition itself is quite simple. Define the total flow of profits at t by:

$$\begin{aligned} O_t &:= \int_{-\infty}^{t} \lambda_\tau e^{-\int_\tau^t \lambda_v \, \mathrm{d}v} o_{\tau,t} \, \mathrm{d}\tau \\ &= -\frac{\epsilon}{\theta+\epsilon} \frac{\alpha\epsilon}{(1-\alpha)\theta+\epsilon} D^\theta \left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{\theta+\epsilon}{\epsilon}\frac{1-\alpha}{\alpha}} Y_t^{-\frac{\theta}{\epsilon}} P_t^{-\chi_{1,1}} X_{1,t} \\ &\qquad + \alpha \left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{1-\alpha}{\alpha}} P_t^{-\chi_{2,1}} X_{2,t}. \end{aligned} \tag{8}$$

using equation (4),[36] and define the total value of all firms at time s over the lives of their current prices by:

$$Q_s^* := \int_{-\infty}^{s} \lambda_\tau e^{-\int_\tau^s \lambda_v \, \mathrm{d}v} \int_s^\infty e^{-\int_s^t (\lambda_v + r_v) \, \mathrm{d}v} o_{\tau,t} \, \mathrm{d}t \, \mathrm{d}\tau.$$

Then:

$$\dot{Q}_t^* = \lambda_t o_t - O_t + r_t Q_t^*,$$

and the conglomerate's first order condition implies:

$$\kappa_1 (\lambda_t - \underline{\lambda})^{\kappa_2} W_t = o_t - Q_t^*.$$

---

[35] See Footnote 11.

[36] Note that by equations (5) and (8):

$$\begin{aligned} \widehat{W}_t V_t + O_t = &-\left[\frac{(1-\alpha)\epsilon}{(1-\alpha)\theta+\epsilon} + \frac{\epsilon}{\theta+\epsilon}\frac{\alpha\epsilon}{(1-\alpha)\theta+\epsilon}\right] D^\theta \left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{\theta+\epsilon}{\epsilon}\frac{1-\alpha}{\alpha}} Y_t^{-\frac{\theta}{\epsilon}} P_t^{-\chi_{1,1}} X_{1,t} \\ &+ [(1-\alpha)+\alpha] \left(\frac{1-\alpha}{\widehat{W}_t}\right)^{\frac{1-\alpha}{\alpha}} P_t^{-\chi_{2,1}} X_{2,t} = \int_{-\infty}^{t} \lambda_\tau e^{-\int_\tau^t \lambda_v dv} \frac{p_\tau}{P_t} y_{\tau,t} \, d\tau. \end{aligned}$$

So, as expected, labour income plus total firm profits equals the real value of goods sold.

This is easy to understand. The right-hand side is the benefit of increasing the price adjustment rate. Firms that update their price have value o_t (over the life of their new price), while those that do not update their price on average have value Q_t^* (over the lives of their current prices). The left-hand side is the marginal cost of increasing the price adjustment rate.

3.5 Households and monetary policy

In period t the representative household maximizes:

$$\int_\tau^\infty e^{-\int_\tau^t \rho_v \, \mathrm{d}v} \left[\log Y_t - \Psi_t \frac{1}{1+\nu} \left(L_t + \frac{\kappa_1}{1+\kappa_2} (\lambda_t - \underline{\lambda})^{1+\kappa_2} \right)^{1+\nu} \right] \mathrm{d}t,$$

where $\nu > 0$ and for all t, $\Psi_t > 0$ and $\rho_t > 0$, with $\int_t^\infty \rho_v \, \mathrm{d}v = \infty$. Note that I have defined L_t so that it just includes production labour, not labour used in price adjustment.

The household faces the budget constraint:

$$Y_t + \frac{\dot{B}_t^{(i)}}{P_t} + \dot{B}_t^{(r)} = W_t \left(L_t + \frac{\kappa_1}{1+\kappa_2} (\lambda_t - \underline{\lambda})^{1+\kappa_2} \right) + i_t \frac{B_t^{(i)}}{P_t} + r_t B_t^{(r)} + \mathrm{T}_t,$$

where $B_t^{(i)}$ are their holdings of nominal bonds, which return i_t, $B_t^{(r)}$ are their holdings of real bonds, which return r_t, and where T_t contains all profits from owning firms and aggregators. The household's first order conditions then imply:

$$\Psi_t \left(L_t + \frac{\kappa_1}{1+\kappa_2} (\lambda_t - \underline{\lambda})^{1+\kappa_2} \right)^\nu = \frac{W_t}{Y_t}, \qquad r_t = \rho_t + \frac{\dot{Y}_t}{Y_t}, \qquad i_t = r_t + \pi_t,$$

where $\pi_t = \frac{\dot{P}_t}{P_t}$.

I assume that the central bank sets the nominal interest rate according to the "real rate rule" of Holden (2024), so in particular:

$$i_t = r_t + \pi_t^* + \phi(\pi_t - \pi_t^*),$$

where $\phi > 1$ and where π_t^* is an exogenous inflation target. Combining this equation with the Fisher equation derived above implies $\pi_t = \pi_t^*$ for all t. Hence, inflation will be effectively exogenous. This is helpful as we are interested in the relationship between output and inflation. The clearest way to study this relationship is to make one of the two exogenous. Making output exogenous risks multiplicity due to the backward bending Phillips curve, so it is more sensible to make inflation exogenous, as here. I can still study monetary policy shocks in this environment, as the central bank can undertake expansionary policy by increasing π_t^*, and contractionary by decreasing it.

3.6 Other aggregates

A number of other aggregates will prove useful. First, I need a measure of the average probability that a buyer from a particular firm will receive their order:

$$\bar{\psi}_{\tau,t} := \int_0^1 \bar{\psi}_{\zeta,\tau,t} g(\zeta)\, d\zeta = \frac{\theta \bar{\zeta}_{\tau,t}^{\epsilon} - \epsilon \bar{\zeta}_{\tau,t}^{\theta}}{\theta - \epsilon}.$$

I then need a measure of the average of this across all firms. For comparability with the fixed weights of the Cavallo & Kryvtsov (2023b) evidence on stockouts, it makes sense to take a simple average across firms, so I define:

$$\bar{\psi}_t := \int_{-\infty}^{t} \lambda_\tau e^{-\int_\tau^t \lambda_v \,\mathrm{d}v} \bar{\psi}_{\tau,t}\,\mathrm{d}\tau.$$

I calibrate θ so that $\bar{\psi} = 1 - 0.11$, as Cavallo & Kryvtsov (2023b) found an 11% stockout rate in the U.S. pre-Covid.

Next, I need the model's equivalent of the PCEPI index. At the most disaggregated level, the PCEPI index uses price indices constructed by the BLS (for the CPI), which are a geometric mean of gross product price growth for most goods.[37] When a price is not observed, due to a stockout, the BLS assumes that the good's price growth is equal to average price growth.

This suggests that over a small interval Δ, the PCEPI price index P_t^{PCEPI} should satisfy:

$$\begin{aligned}
&\frac{1}{\Delta}\left(\log P_t^{\mathrm{PCEPI}} - \log P_{t-\Delta}^{\mathrm{PCEPI}}\right)\\
&= \frac{1}{\Delta}\int_{-\infty}^{t-\Delta} \lambda_\tau e^{-\int_\tau^{t-\Delta} \lambda_v \,\mathrm{d}v} \left[\bar{\psi}_{\tau,t-\Delta}\lambda_t \Delta \left[\bar{\psi}_{t,t}\log\frac{p_t}{p_\tau} + (1-\bar{\psi}_{t,t})\log\frac{P_t^{\mathrm{PCEPI}}}{P_{t-\Delta}^{\mathrm{PCEPI}}}\right]\right.\\
&\qquad + \bar{\psi}_{\tau,t-\Delta}(1-\lambda_t\Delta)\left[\bar{\psi}_{\tau,t}0 + (1-\bar{\psi}_{\tau,t})\log\frac{P_t^{\mathrm{PCEPI}}}{P_{t-\Delta}^{\mathrm{PCEPI}}}\right]\\
&\qquad \left. + (1-\bar{\psi}_{\tau,t-\Delta})\log\frac{P_t^{\mathrm{PCEPI}}}{P_{t-\Delta}^{\mathrm{PCEPI}}}\right]\mathrm{d}\tau,
\end{aligned}$$

where we have split the integral to consider the multiple cases coming from the three following events: (1) was the price observed at $t - \Delta$? (2) did the price change in the interval $(t - \Delta, t]$? (3) was the price observed at t? Taking the limit as $\Delta \to 0$ and solving for $\frac{\mathrm{d}\log P_t^{\mathrm{PCEPI}}}{\mathrm{d}t}$ implies:

[37] See Footnote 13.

$$\pi_t^{\text{PCEPI}} := \frac{\mathrm{d}\log P_t^{\text{PCEPI}}}{\mathrm{d}t} = \lambda_t \bar{\psi}_{t,t} \frac{\int_{-\infty}^t \lambda_\tau e^{-\int_\tau^t \lambda_v \,\mathrm{d}v} \bar{\psi}_{\tau,t}^2 \frac{1}{\bar{\psi}_{\tau,t}} \log \frac{p_t}{p_\tau} \,\mathrm{d}\tau}{\int_{-\infty}^t \lambda_\tau e^{-\int_\tau^t \lambda_v \,\mathrm{d}v} \bar{\psi}_{\tau,t}^2 \,\mathrm{d}\tau}.$$

I have deliberately not simplified $\bar{\psi}_{\tau,t}^2 \frac{1}{\bar{\psi}_{\tau,t}}$ in the numerator to make clear that this is proportional to a weighted mean of $\frac{1}{\bar{\psi}_{\tau,t}} \log \frac{p_t}{p_\tau}$ across firms. Old firms will have low prices, and so will ration a lot, making $\frac{1}{\bar{\psi}_{\tau,t}}$ big. Thus, this measure will effectively give higher weight to the (large) price changes of older firms.

Finally, I need a measure of aggregate productivity. First, imagine that a constrained social planner wants to maximize aggregate output (without rationing) at t by choosing $v_{\zeta,\tau,t}$ for all $\zeta \in [0,1]$ and $\tau \le t$ subject to a fixed total effective labour supply, V_t. Then, I show in Appendix A that their choices imply total output Y_t of:

$$Y_t^{\text{SP}} := \left[\frac{\theta + 1}{\theta + \frac{\epsilon}{1 + \alpha(\epsilon - 1)}} \right]^{\frac{1+\alpha(\epsilon-1)}{\epsilon-1}} \left(\frac{\theta + 1}{\theta} V_t \right)^{1-\alpha}.$$

Given this, the natural measure of the economy's productivity is $\frac{Y_t}{Y_t^{\text{SP}}}$.

3.7 Detrended variables and stability

For the sake of simulation, it is helpful to define detrended versions of the model's variables, such that the detrended variables are stationary. The differential equations followed by these detrended variables will also inform us about the model's stability.

For the state variables, I define $\hat{X}_{j,t} := \frac{X_{j,t}}{P_t^{\chi_{j,1}}}$ for $j \in \mathbb{Z}$, and I define $\hat{p}_t := \frac{p_t}{P_t}$. Then:

$$\dot{\hat{X}}_{j,t} = \lambda_t \hat{p}_t^{\chi_{j,1}} - (\lambda_t + \chi_{j,1}\pi_t)\hat{X}_{j,t}.$$

Given a path of $\hat{p}_t$, this differential equation is stable if and only if $\lambda_t + \chi_{j,1}\pi_t > 0$, in which case when $\hat{X}_{j,t}$ is high, it will be pushed back towards trend. Recall that with rationing, my model has the state variables $X_{1,t}$ and $X_{2,t}$, with $\chi_{1,1} = \theta + \frac{1}{\alpha} + \frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha} > 0$ and $\chi_{2,1} = \frac{1}{\alpha} > 0$. Thus, as long as inflation does not go too negative, both state variables will be stable. By contrast, the state variable of the model without rationing is $X_{-1,t}$ with $\chi_{-1,1} := -\frac{\epsilon}{1-\alpha} < 0$ (see Appendix B). Since this is negative, if inflation gets too high then the state variable can explode to infinity, with output collapsing to zero. Marsal, Rabitsch & Kaszab

(2023) and Holden, Marsal & Rabitsch (2024) show that this instability is a major problem for empirically plausible calibrations. It is not even clear that a valid global solution exists to the basic New Keynesian model. Luckily, all of these problems go away when rationing is allowed.

For the forward-looking variables, I define $\hat{z}_{j,t} := \frac{z_{j,t}}{P_t^{\omega_{j,4}}}$ for $j \in \mathbb{N}$, so:

$$\dot{\hat{z}}_{j,t} = -D^{\omega_{j,1}} \widehat{W}_t^{\omega_{j,2}} Y_t^{\omega_{j,3}} + \left(\lambda_t + r_t - \omega_{j,4}\pi_t\right)\hat{z}_{j,t}.$$

Remembering that this equation is solved backwards in time, given the paths of other variables, $\omega_{j,4} < 0$ is sufficient for "stability" (with r_t, π_t positive). The forward-looking variables with rationing were $z_{1,t}$ and $z_{2,t}$, with $\omega_{1,4} = -\left(\frac{1}{\alpha} + \theta + \frac{\theta}{\epsilon}\frac{1-\alpha}{\alpha}\right) < 0$ and $\omega_{2,4} = -\frac{1}{\alpha} < 0$, so both variables are well behaved. Again, without rationing, neither of the two forward looking variables have this "stability" property.

3.8 Parameterization and calibration

I will show results for the model with rationing presented here, as well as for the equivalent model without rationing. (See Appendix B for the model without rationing.) I set most parameters to standard values for both models. I set $\rho := 2\%$ and $\pi^{\text{PCEPI}} := 2\%$, unless otherwise stated. For the model with rationing, hitting this target for PCEPI inflation requires true inflation of 2.04%. For the model without rationing, true inflation is also 2%.

Following Smets & Wouters (2007), I set $\epsilon := 10$ and $\nu := 2$. I set $\alpha := \frac{3}{5}$ following the argument of the introduction and the evidence of Abraham et al. (2024). Note that $\frac{3}{5}$ was consistent with the fixed share evidence of Abraham et al. (2024) at annual frequency. At higher frequencies, perhaps even higher calibrations of α would be justified. Choosing $\frac{3}{5}$ is thus relatively conservative. I normalize units by setting $A := 1$ and I normalize Ψ so that steady-state production labour supply agrees with that in an equivalent model with exogenous λ_t and $\Psi = 1$.

I set $\theta := 27$, as at this level the model implies that in steady state, the average probability of being rationed across goods, $1 - \bar{\psi} = 11\%$, matching the 11% stockouts found pre-Covid by Cavallo & Kryvtsov (2023b). Setting $\theta = 27$ implies the mean of ζ is 0.96 and its standard deviation is 0.03. This does not seem like an implausibly high level for an idiosyncratic demand shock.

As previously mentioned, I set $\underline{\lambda}$ to the minimum annual median price adjustment rate in the Montag & Villar (2025) data, $\underline{\lambda} = 0.73$, and I calibrate κ_1 to match the time series mean of the median rate of price adjustment from the same data, $\lambda = 1.48$. (This implies an expected price duration of 8 months.) I calibrate κ_2 to equate the value of:

$$\frac{\int_0^{\frac{1}{4}}(\lambda_t - \lambda)\,\mathrm{d}t}{\int_0^{\frac{1}{4}}(\pi_t^{\mathrm{PCEPI}} - \pi^{\mathrm{PCEPI}})\,\mathrm{d}t}$$

following a monetary policy shock to the value of this expression estimated from Figure 1, 8.2.[38] With rationing, this leads to $\kappa_1 = 0.016$ and $\kappa_2 = 3.75$. Without rationing, I set $\kappa_1 = 0.105$ and $\kappa_2 = 2.06$.[39] With this calibration, with rationing allowed, only 0.1% of all labour is used for price adjustment. By contrast, without rationing, 2.0% of all labour is used for price adjustment. This illustrates the degree to which rationing reduces the price adjustment frictions needed to match the data.

4 Results

I will first present comparative static results varying the steady-state inflation rate. Given these results suggest that welfare is higher in the model with rationing, I then spend some time discussing what drives this. I then examine the model's impulse responses to monetary policy shocks, starting with a comparison of the model's three-month Phillips curve to the results of Figure 1. Let me first remind you, though, of one important result we have already seen. In Figure 2, I showed that the model can match the concavity of firm sales over the life of a price that we see in scanner data, despite this not being a calibration target.

4.1 Comparative statics

This Subsection will present quite a number of graphs. In almost all the

[38] I choose the persistence of the monetary policy shock so that the resulting path for one-year bonds is inside the confidence bands from Figure 1. The annual decay rate of π_t^* is 3.9, giving an annual decay rate for one-year bonds of 5.1.

[39] A slightly lower κ_2 would have been preferable in the no-rationing case, but numerical difficulties prevented this.

following plots, black solid lines are from the model with rationing, and black dashed lines are from the model without rationing. (See Appendix B for the model without rationing.)

A first question to answer is how rationing varies as the rate of inflation varies. The first panel of Figure 6 answers this. It shows that average rationing levels are increasing in the steady-state inflation rate. This is driven by the fact that when inflation is high, mark-ups are eroded quickly, leading to greater rationing.

A natural follow-on question is: which firms ration? The middle panel of Figure 6 shows that for firms with new prices, stockout rates are actually decreasing in steady-state inflation. This is because when inflation is high, firms resetting their price choose a high initial mark-up, to protect themselves against future mark-up erosion. However, the third panel of Figure 6 shows that this effect for firms with new prices is dominated by the mark-up erosion channel. The dark blue line in that panel shows that if the steady-state inflation rate is 0.5%, then the stockout probability is almost constant over the life of a price. However, if the steady-state inflation rate is 8% (the dark red line), then the probability of being rationed increases quickly as the prices ages, as inflation erodes markups.

Figure 7 shows how output, production labour supply and welfare change with the long-run level of inflation. Both with and without rationing, the welfare maximising and output maximising inflation levels are very close to 0% (at least conditional on inflation being positive). Figure 7 makes clear that the costs of high steady-state inflation are far more serious when rationing is not allowed. While with rationing, 8% inflation leads to 2.6% worse welfare (consumption equivalent) than 0% inflation, without rationing the same loss is 36%. Inflation is bounded above in the model without rationing and with exogenous λ_t, and output falls to zero as inflation approaches this level. Allowing conglomerates to choose price adjustment rates removes this hard upper bound, but instead a substantial amount of labour is diverted to price adjustment when inflation is high.

Figure 8 looks in more detail at how rationing might be improving welfare relative to economies without rationing. The first panel looks at productivity (measured by $\log \frac{Y}{Y^{\mathrm{SP}}}$). With rationing, the productivity loss due to rationing and labour misallocation across firms varies from 4% to 6%. This is dwarfed by the productivity loss due to misallocation without rationing, which hits 25% at 8% inflation.

A priori, a potential candidate for productivity losses might have been mark-ups. But the bottom panels of Figure 8 show aggregate mark-ups (measured by $\frac{(1-\alpha)Y}{WL}$) and aggregate excess firm profit shares ($\frac{O}{Y} - \alpha$) are rapidly declining in inflation without rationing. This is driven by the combination of inflation eroding firm mark-ups, and firms with low (or negative) mark-ups selling large quantities. By contrast, when rationing is allowed, no firms set negative mark-ups, and firms with low mark-ups sell low quantities. Thus, aggregate mark-ups and profits barely change with inflation.

Finally, Figure 8 shows how the price adjustment rate varies with inflation. At 8% inflation, with rationing, price adjustment rates reach 1.91, while without rationing, they hit 2.66. While this may not seem like a huge difference, the consequences for price adjustment labour use are massive, due to the differing calibrations of κ_1 and κ_2 (that come from the differing losses from having an old price across the two models). With rationing, at 8% inflation, a moderate 1.0% of labour is used for price adjustment. Without, this figure is over 31%, explaining a substantial portion of the output loss.

4.2 Why might rationing be desirable?

The previous subsection showed welfare is higher in economies with rationing than in those without rationing.[40] This may be surprising. Is rationing not a bad thing?

Rationing's relative welfare benefits are primarily a consequence of the fact that in standard models, the firms with the most distorted prices are selling a lot, since the most distorted prices are very old and hence very low. High production by these firms with old prices pushes up marginal costs for all firms, in turn reducing output for firms with relatively undistorted prices. Thus,

[40] See also the results and discussion in Hahn (2022), who also examined static outcomes under rationing with sticky prices, but without idiosyncratic demand shocks.

without rationing, demand is shifted from firms with undistorted prices to firms with distorted prices. By contrast, if rationing is allowed, then these firms with old, highly distorted, prices will limit sales through rationing. With relatively low production of goods with old prices, there will be less pressure on marginal costs for firms with new prices, so those firms will produce more. Demand is shifted from firms with distorted prices to firms with undistorted ones, at least relative to the no rationing benchmark.

Furthermore, note that if a firm could adjust their price after observing their demand shock, they would choose a price that is increasing in ζ. Thus, fully flexible prices lead to reduced sales when ζ is high compared to the sticky or quasi-flexible benchmarks without rationing. Rationing also limits sales when ζ is high, so it is intuitive that increased rationing can bring the economy closer to the fully flexible benchmark.

At the micro level (looking at demand and supply of a single good), with arbitrary demand and cost curves and a fixed price, it is ambiguous whether average welfare (with quasilinear utility) is higher with rationing or with production of the full quantity demanded. But, in reality, we expect the demand curve ($\frac{p}{P} \propto y^{-\frac{1}{\epsilon}} \approx y^{-\frac{1}{10}}$) to be flatter than the marginal cost curve ($\mathrm{MC} \propto y^{\frac{\alpha}{1-\alpha}} \approx y^{\frac{3}{2}}$). In this case, we can see graphically that welfare should be higher when rationing is allowed than when firms are forced to satisfy demand, as shown in Figure 4. While the graphical argument of Figure 4 strictly only applies with linear marginal costs and linear demand, this result is more general. In Appendix C.1 I show that microeconomic welfare is higher with rationing with general isoelastic demand and marginal costs. Of course, with random rationing, these average welfare figures mask substantial heterogeneity across buyers. Some are getting their full order, while others get nothing. So, such average quasilinear micro-welfare results are far from the full story.

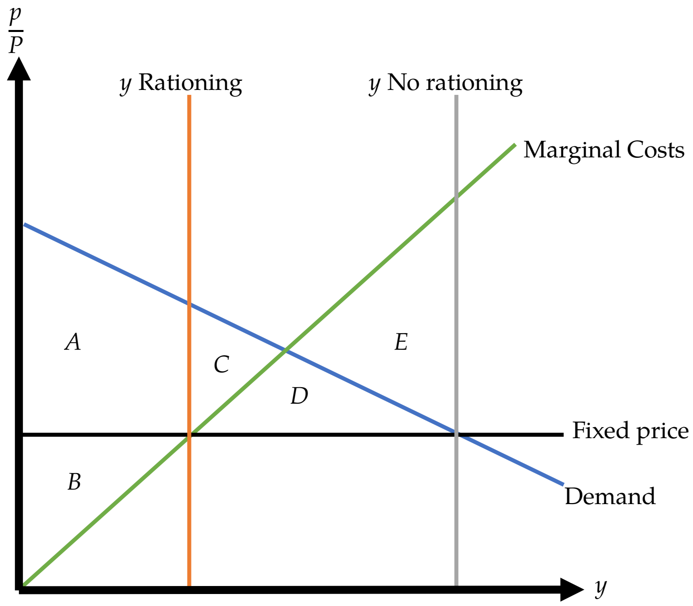


Figure 4: The microeconomics of rationing.

With rationing allowed, production is given by the orange line, and welfare is $A + B$.
Without rationing, production is given by the grey line, and welfare is $A + B + C - E$.
With demand flatter than marginal costs, $E > C$, and so welfare is higher with rationing.

The average benefits of rationing are even clearer if supply constraints really do mean that marginal costs go to infinity at some finite output level $\bar{y}$, as depicted in Figure 5. Then, if the quantity demanded at the current price is greater than $\bar{y}$, there is no way the micro market can clear without rationing, holding macro quantities fixed. Instead, as the firm increases production to try to satisfy demand, more and more of the economy's resources are devoted to this one micro market. This decreases aggregate production, pushing down demand for all products, including the current one, until demand for it is below $\bar{y}$. Thus, without rationing, macro quantities may have to move to clear a micro market, producing arbitrarily large distortions. With rationing, the equilibrium is at the point at which price equals marginal cost, as usual.

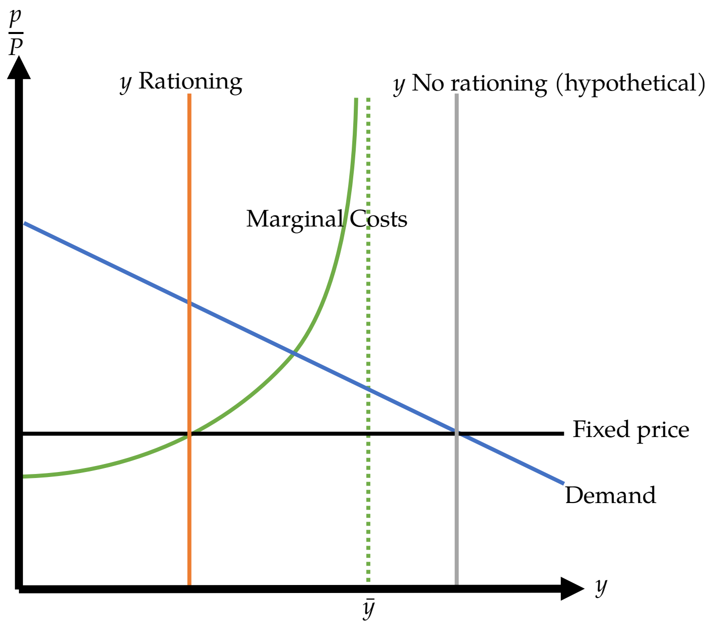
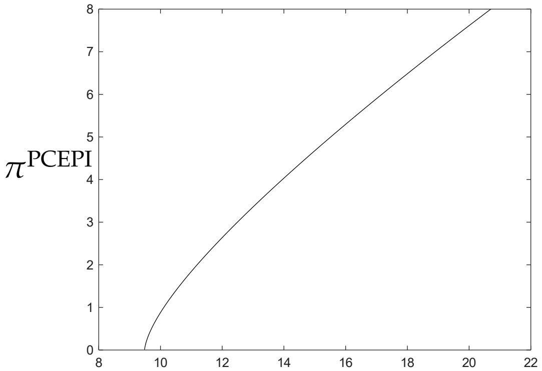


Figure 5: The microeconomics of rationing with supply constraints.

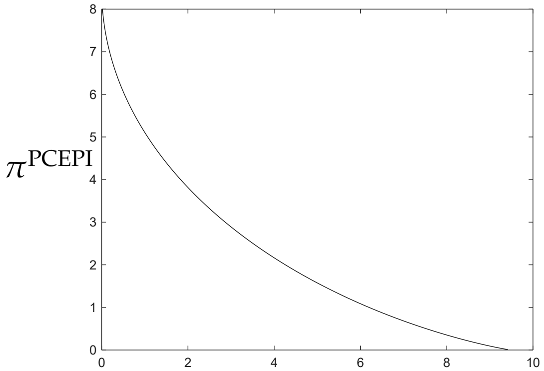
With rationing allowed, production is given by the orange line.

Without rationing, production should be given by the grey line, but it is impossible to ever produce this much, as maximum output is $\bar{y}$, the dashed green line.

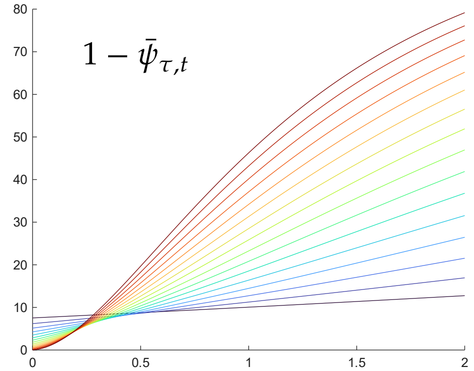
4.3 Results figures



Average stockout level, $1 - \bar{\psi}$ (percent).



Stockout rate at firms with new prices, $1 - \bar{\psi}_{t,t}$ (percent).

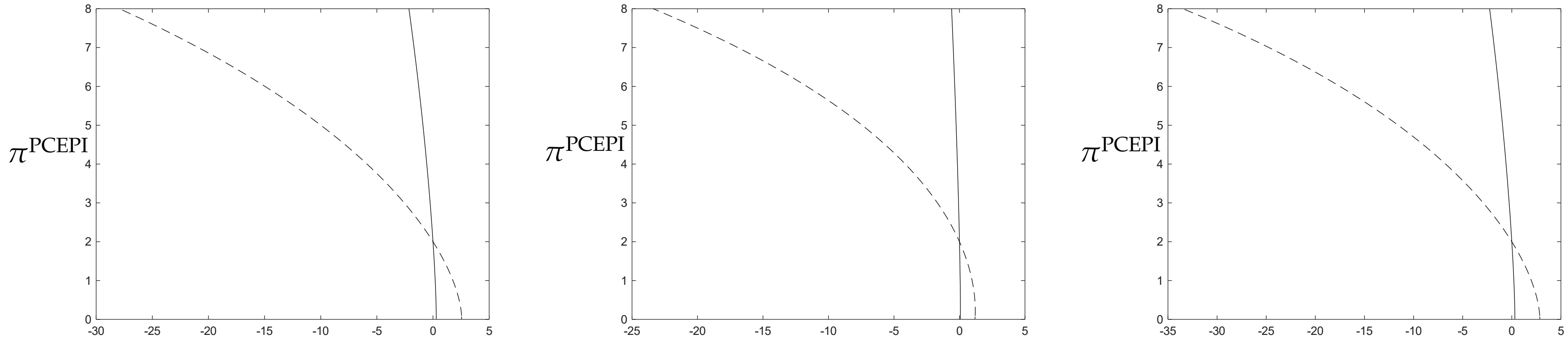


Stockout levels (percent) as a function of price age (years), with varying steady state inflation levels.

Dark blue corresponds to 0.5% inflation.

Dark red corresponds to 8% inflation.

Figure 6: Stockouts and rationing as a function of PCEPI inflation (percent).



Relative output: $\log Y$ (percent).

Relative production labour supply: $\log L$ (percent).

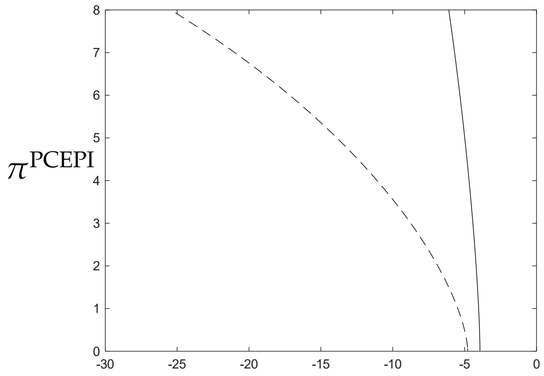
Relative welfare:

$$100\left(\log Y - \Psi\frac{1}{1+\nu}\left(L + \frac{\kappa_1}{1+\kappa_2}(\lambda - \underline{\lambda})^{1+\kappa_2}\right)^{1+\nu}\right)$$

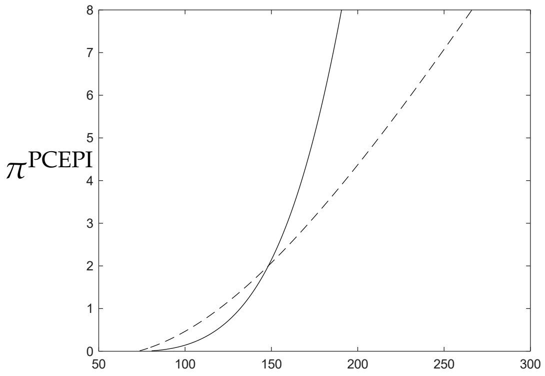
Figure 7: Output and welfare as a function of inflation (percent).

Black solid lines are the model with rationing. Black dashed lines are the model without rationing.

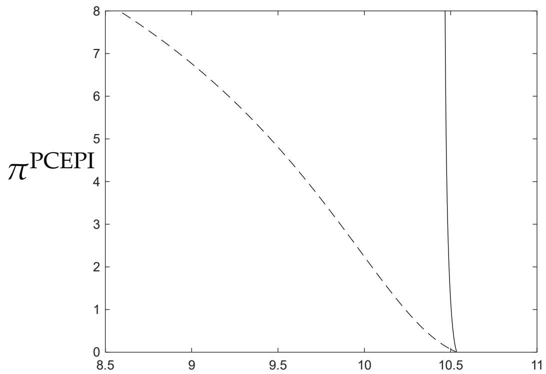
All plots are normalized to hit 0% on the horizontal axis when $\pi^{\text{PCEPI}} = 2\%$.



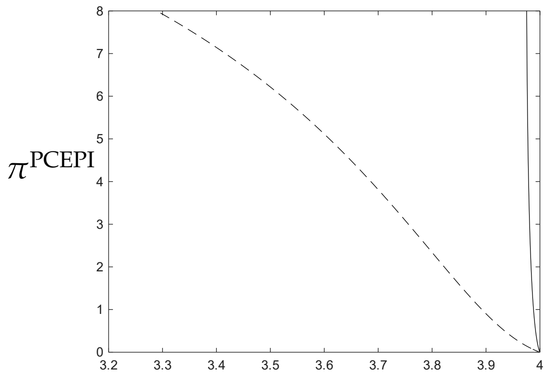
TFP loss from efficient benchmark: $\log \frac{Y}{Y^{\text{SP}}}$ (percent).



Price adjustment rate (percent): 100λ.



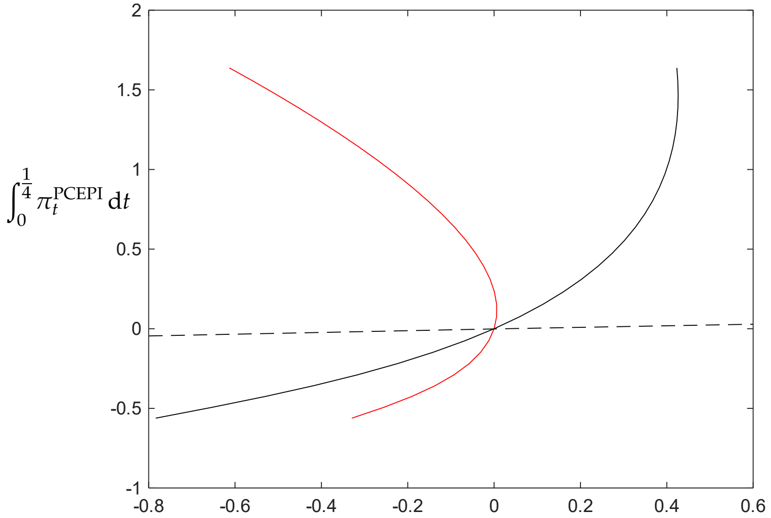
Aggregate mark-ups: $\log \frac{(1-\alpha)Y}{WL}$ (percent).



Excess firm profits shares: $\frac{O}{Y} - \alpha$ (percent).

Figure 8: Other consequences of varying inflation (percent).

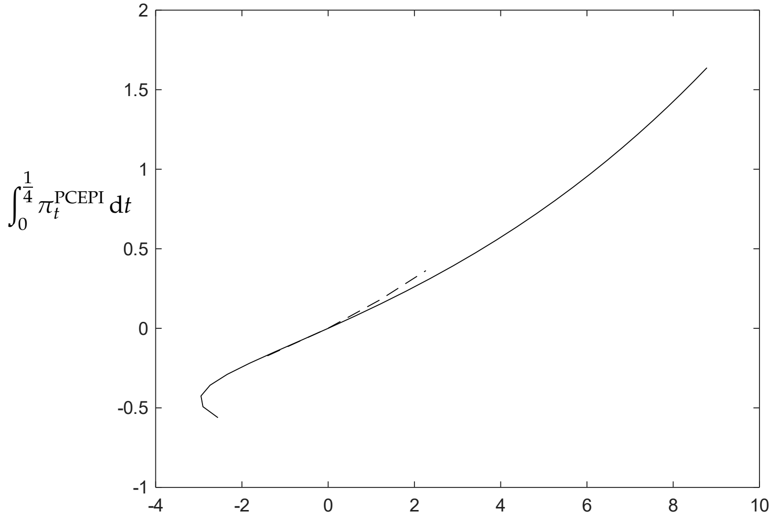
Black solid lines are the model with rationing. Black dashed lines are the model without rationing.



Black solid line: measured cumulated real GDP, $\log \int_0^{\frac{1}{4}} \frac{P_t Y_t}{P_t^{\text{PCEPI}}} \, \mathrm{d}t$, with rationing.

Red solid line: cumulated true output, $\log \int_0^{\frac{1}{4}} Y_t \, \mathrm{d}t$, with rationing.

Black dashed line: measured cumulated real GDP, $\log \int_0^{\frac{1}{4}} \frac{P_t Y_t}{P_t^{\text{PCEPI}}} \, \mathrm{d}t$, without rationing.

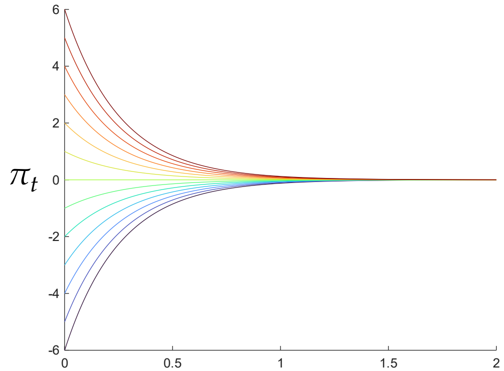


Black solid line: cumulated price adjustment rate, $100 \int_0^{\frac{1}{4}} \lambda_t \, \mathrm{d}t$, with rationing.

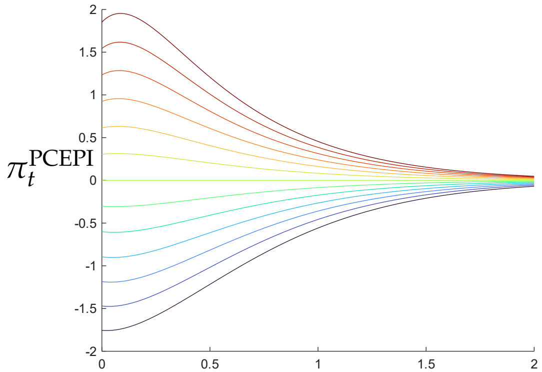
Black dashed line: cumulated price adjustment rate, $100 \int_0^{\frac{1}{4}} \lambda_t \, \mathrm{d}t$, without rationing.

Figure 9: The three-month Phillips curve with (solid lines) and without (dashed lines) rationing.

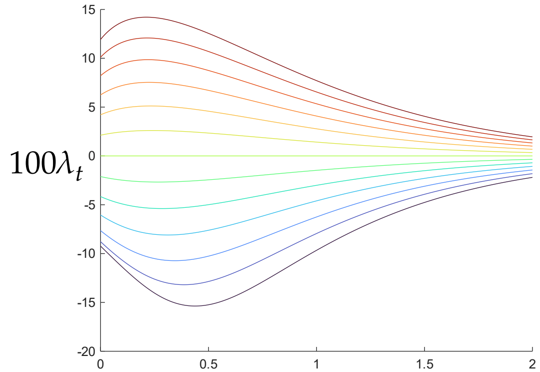
All variables in percent. All variables are relative to the no-shock counterfactual.



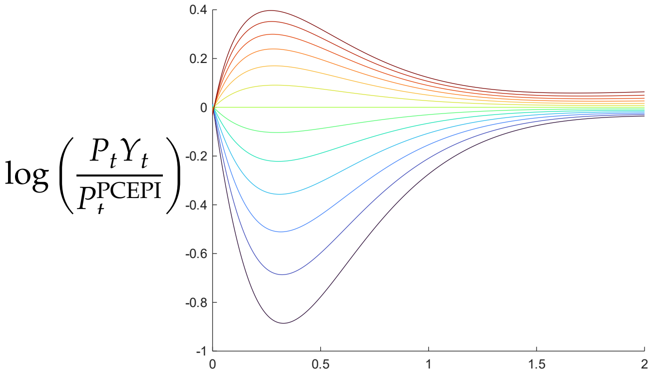
Driving inflation shocks (percent) as a function of time (years).



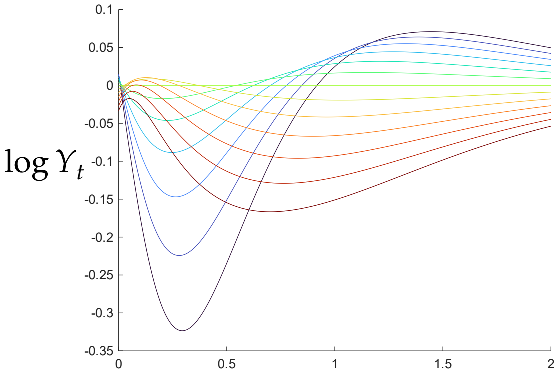
PCEPI inflation (percent) as a function of time (years).



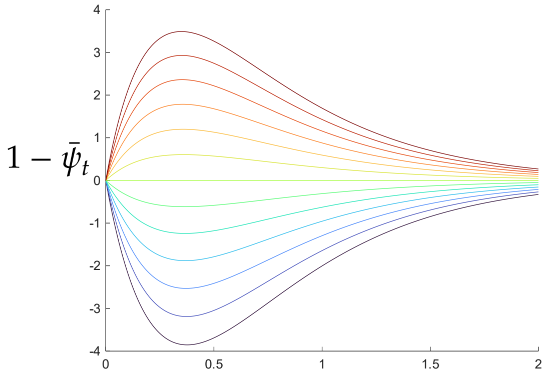
Price adjustment rate (× 100) as a function of time (years).



Measured real GDP response (percent) as a function of time (years).



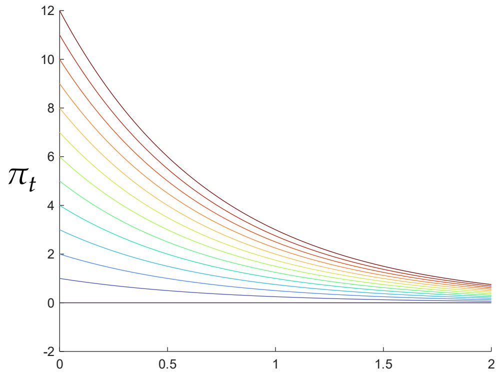
True output response (percent) as a function of time (years).



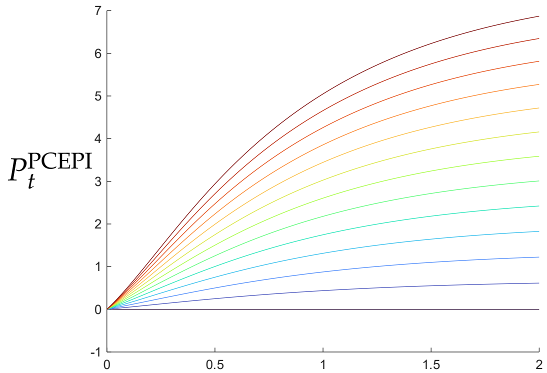
Stockout share response (percentage points) as a function of time (years).

Figure 10: Impulse responses to monetary shocks, with rationing.

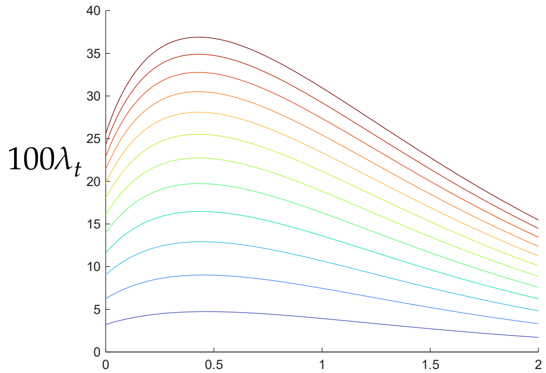
Colours are consistent across subplots. All responses are relative to the no-shock counterfactual.



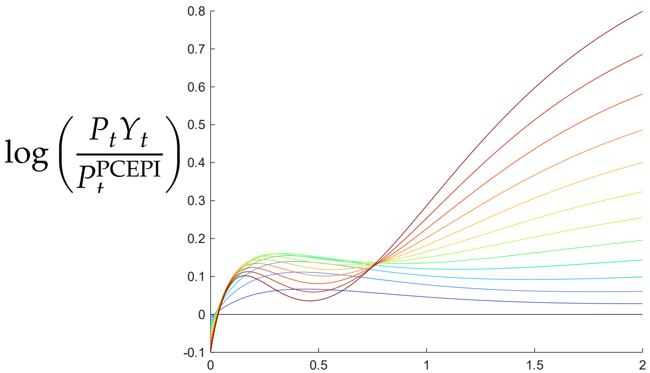
Driving inflation shocks (percent) as a function of time (years).



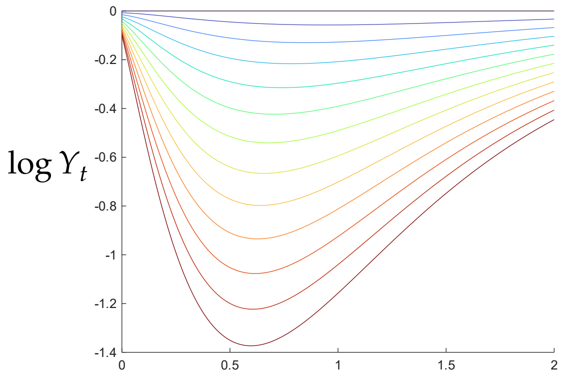
PCEPI **level** (percent) as a function of time (years).



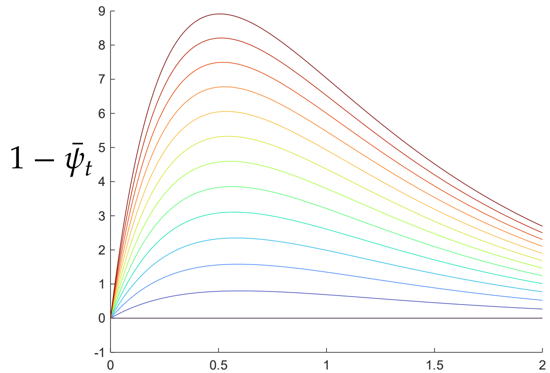
Price adjustment rate (× 100) as a function of time (years).



Measured real GDP response (percent) as a function of time (years).



True output response (percent) as a function of time (years).



Stockout share response (percentage points) as a function of time (years).

Figure 11: Impulse responses to more persistent monetary shocks, with rationing.

Colours are consistent across subplots. All responses are relative to the no-shock counterfactual.

4.4 Dynamics

I now examine the behaviour of the model in response to unexpected monetary policy shocks which vary π_t^* and hence π_t. I assume these shocks have prior probability zero, "MIT shock" style, and I assume the economy begins in steady state. In this simple model, other potential shocks are of limited interest due to divine coincidence.[41]

I consider driving monetary shocks of the form:

$$\pi_t^* = 2\% + \text{shock} \times \exp(-\varrho t),$$

for $t \geq 0$, for varying values of "shock". I set $\varrho := 3.9$ to ensure that following small shocks, the resulting impulse response for one-year bonds is inside the confidence bands from Figure 1. (This gives an annual decay rate for one-year bonds of 5.1.)

4.5 The three-month Phillips curve

I start by producing the three-month Phillips curve for the models with and without rationing, shown in the first panel of Figure 9. This plots cumulated output in the three months following a monetary shock of varying magnitude (relative to the no-shock counterfactual) again cumulated inflation in these three months (again relative to the no-shock counterfactual).

The black lines in this plot measure output as nominal GDP divided by the model's PCEPI index. Since there is no investment (etc.) in the model, this is the natural equivalent of measured real GDP. With rationing, the resulting Phillips curve slope around the origin is 1.2 (this is the solid black line). This is exactly the same as the three-month Phillips curve slope implied by the results in Figure 1 (produced using equivalent calculations). Thus, the model with rationing matches the observed three-month Phillips curve slope, without this being a calibration target. The model without rationing generates a three-month Phillips curve slope of only 0.05 (this is the dashed black line), completely failing to match the data.

The better empirical performance of the rationing model is partly explained

[41] Shocks to productivity, A_t, the disutility of labour supply, Ψ_t, or the discount rate, ρ_t, have essentially identical effects to their effects under quasi-flexible prices if monetary policy holds inflation constant.

by the flexibility in the welfare relevant price index discussed in Subsection 3.2. Increases in rationing lead this index to place greater weight on newer firms that ration less and have higher prices. The flexibility in the true price index drives flexibility in the model's version of the PCEPI index, as firm pass through cost changes. Additionally, with rationing, measured PCEPI price growth gets tilted towards the price growth of firms with old prices, following shocks that increase rationing. This is a consequence of the BLS's imputation procedure, which assumes unobserved prices have the average price growth rate, as demonstrated in Subsection 3.6.

The first panel of Figure 9 also demonstrates that the model with rationing produces substantial convexity in its (three-month) Phillips curve. Expansionary shocks lead to increases in rationing, dampening the output impact. Contractionary shocks reduce rationing, cushioning the output impact.

However, the solid black line in this panel is no way near as convex as the solid red line. Whereas the solid black line plots our model's equivalent of cumulated measured real GDP, the solid red line plots our model's cumulated true output. While moderate expansionary monetary policy shocks increase measured real GDP over three months, we see that they reduce the model's true output over the same period. The price index used in constructing real GDP cannot capture changes in consumers' gains from variety, and so when these gains fall (due to increased rationing) measured real GDP is overstated. For a monetary policy maker, this is alarming. At least in the vicinity of the steady state, changes in monetary policy cannot stimulate the economy, correctly measured.

The second panel of Figure 9 plots the three-month "price adjustment Phillips curve". The slope of this curve around the origin was a calibration target, and thus is of limited interest. However, it is interesting to see that following large contractionary shocks, this price adjustment Phillips curve bends backwards, and conglomerates increase the rate of price adjustment again. This is intuitive. Small contractionary shocks can be absorbed by merely skipping the regular positive price adjustments that come from trend inflation. But large contractionary shocks make price reductions desirable, necessitating an increase in the price adjustment rate.

4.6 Impulse responses to monetary shocks

To see how the economy evolves beyond the three-month horizon following monetary shocks, I now present impulse responses in the model with rationing. Given the results of the previous subsection, these will not contain any major surprises.

Figure 10 contains these impulse responses, for driving π_t^* shocks from +6% to −6%. In all the panels, the +6% shock is in dark red, while the −6% shock is in dark blue, with intermediate shocks in intermediary colours of the rainbow. These move PCEPI inflation by around $\pm$2%. Only the large "contractionary" shocks succeed in increasing true output one year out. These contractionary shocks actually increase felicity here, though it is unlikely that this result would survive in a model with a more plausible labour market.

Since PCEPI inflation moves much less than true inflation, even a +6% shock does not get PCEPI inflation up to the 7% level we saw post-Covid. To see how the model behaves at such inflation levels, in Figure 11 I repeat the previous exercise for positive π_t^* shocks between 0% and 12%, but now with the shock persistence, $\varrho = 2 \log 2$, implying the shock has a half-life of half a year. The second panel shows that the +12% π_t^* shock succeeds in raising the measured PCEPI price index by 5% relative to the counterfactual, after one year. Given steady-state inflation is 2%, this matches the 7% PCEPI inflation we saw in the U.S. from June 2021 to June 2022. (Incidentally, the fact that the gap between measured and true inflation is so big following large shocks may help explain some of the biases in consumers' inflation expectations.) This shock raises the stockout rate from 11% to about 20% at the peak, close to the 23% stockout level Cavallo & Kryvtsov (2023b) found for 2022. (Another untargeted moment matched!) The shock also raises λ_t from 1.48 to 1.85. This is some way off the levels of λ_t implied by the Montag & Villar (2025) data for the post-Covid period, suggesting further changes to my price adjustment cost function may be necessary.

5 Conclusion

In this paper I have shown how relaxing one small simplifying assumption from the workhorse model of sticky prices drastically alters the conclusions of that model. Allowing firms to ration removes most of the welfare costs of trend inflation yet leads "expansionary" monetary policy shocks to decrease the welfare relevant output measure. The model with rationing also matches the data remarkably well. With just one parameter controlling rationing, the model roughly matches the level of stockouts pre-Covid, the level of stockouts in the high inflation of 2022, the concavity of output over the life of a price and the slope of the three-month Phillips curve derived from high frequency monetary shocks. The model also produces a convex Phillips curve, as we see in the data. Allowing rationing appears essential to understanding the relationship between inflation and output and has dramatic implications for optimal monetary policy.

References

Abraham, Filip, Yannick Bormans, Jozef Konings & Werner Roeger. 2024. 'Price-Cost Margins, Fixed Costs and Excess Profits'. *The Economic Journal*: ueae037.

Adam, Klaus & Henning Weber. 2019. 'Optimal Trend Inflation'. *American Economic Review* 109 (2): 702–737.

Alessandria, George, Joseph P. Kaboski & Virgiliu Midrigan. 2010. 'Inventories, Lumpy Trade, and Large Devaluations'. *American Economic Review* 100 (5): 2304–2339.

Babb, Nathan R. & Alan K. Detmeister. 2017. 'Nonlinearities in the Phillips Curve for the United States : Evidence Using Metropolitan Data'. *Finance and Economics Discussion Series*. Finance and Economics Discussion Series.

Barro, Robert J. 1977. 'Long-Term Contracting, Sticky Prices, and Monetary Policy'. *Journal of Monetary Economics* 3 (3): 305–316.

Barro, Robert J. & Herschel I. Grossman. 1971. 'A General Disequilibrium Model of Income and Employment'. *The American Economic Review* 61 (1): 82–93.

Bauer, Michael D. & Eric T. Swanson. 2023. 'A Reassessment of Monetary Policy Surprises and High-Frequency Identification'. *NBER Macroeconomics Annual* 37: 87–155.

Bils, Mark. 2016. 'Deducing Markups from Stockout Behavior'. *Research in Economics* 70 (2): 320–331.

Blanco, Andrés, Corina Boar, Callum J. Jones & Virgiliu Midrigan. 2024a. 'Non-Linear Inflation Dynamics in Menu Cost Economies'. Working Paper. Working Paper Series. National Bureau of Economic Research.

———. 2024b. 'The Inflation Accelerator'. Working Paper. Working Paper Series. National Bureau of Economic Research.

Boehm, Christoph E., Aaron Flaaen & Nitya Pandalai-Nayar. 2019. 'Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake'. *The Review of Economics and Statistics* 101 (1): 60–75.

Brave, Scott A., R. Andrew Butters & David Kelley. 2019. 'A New "Big Data" Index of U.S. Economic Activity'. *Economic Perspectives* (1): 1–30.

Brave, Scott A., Ross Cole & David Kelley. 2019. 'A "Big Data" View of the U.S. Economy: Introducing the Brave-Butters-Kelley Indexes'. *Chicago Fed Letter*.

Broda, Christian & David E. Weinstein. 2006. 'Globalization and the Gains From Variety'. *The Quarterly Journal of Economics* 121 (2): 541–585.

Broda, Christian & David E Weinstein. 2010. 'Product Creation and Destruction: Evidence and Price Implications'. *American Economic Review* 100 (3): 691–723.

Cameron, A. Colin, Jonah B. Gelbach & Douglas L. Miller. 2011. 'Robust Inference With Multiway Clustering'. *Journal of Business & Economic Statistics* 29 (2): 238–249.

Cavallo, Alberto & Oleksiy Kryvtsov. 2023a. 'Replication Code and Data - "What Can Stockouts Tell Us About Inflation? Evidence from Online Micro Data", Journal of International Economics' 1.

———. 2023b. 'What Can Stockouts Tell Us about Inflation? Evidence from Online Micro Data'. *Journal of International Economics* 146. NBER International Seminar on Macroeconomics 2022: 103769.

Cooper, Russell W. & John C. Haltiwanger. 2006. 'On the Nature of Capital Adjustment Costs'. *The Review of Economic Studies* 73 (3): 611–633.

Corsetti, Giancarlo & Paolo Pesenti. 2005. 'International Dimensions of Optimal

Monetary Policy'. *Journal of Monetary Economics* 52 (2): 281–305.

Dotsey, Michael, Robert G. King & Alexander L. Wolman. 1999. 'State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output*'. *The Quarterly Journal of Economics* 114 (2): 655–690.

Drèze, Jacques H. 1975. 'Existence of an Exchange Equilibrium under Price Rigidities'. *International Economic Review* 16 (2): 301–320.

Forbes, Kristin J., Joseph E. Gagnon & Christopher G. Collins. 2022. 'Low Inflation Bends the Phillips Curve around the World'. *Economia* 45 (89): 52–72.

Gerke, Rafael, Sebastian Giesen, Matija Lozej & Joost Röttger. 2023. 'On Household Labour Supply in Sticky-Wage HANK Models'. SSRN Scholarly Paper. Rochester, NY.

Golosov, Mikhail & Robert E. Lucas Jr. 2007. 'Menu Costs and Phillips Curves'. *Journal of Political Economy* 115 (2): 171–199.

Hahn, Volker. 2022. 'Price Dispersion and the Costs of Inflation'. *Journal of Money, Credit and Banking* 54 (2–3): 459–491.

Holden, Tom D. 2024. *Robust Real Rate Rules*. Working Paper. Kiel, Hamburg: ZBW – Leibniz Information Centre for Economics.

Holden, Tom D., Ales Marsal & Katrin Rabitsch. 2024. 'From Linear to Nonlinear: Rethinking Inflation Dynamics in the Calvo Pricing Mechanism'.

Huo, Zhen & José-Víctor Ríos-Rull. 2020. 'Sticky Wage Models and Labor Supply Constraints'. *American Economic Journal: Macroeconomics* 12 (3): 284–318.

Kabir, Eugene Tan & Ia Vardishvili. 2024. 'Quantifying the Allocative Efficiency of Capital: The Role of Capital Utilization'.

Khan, Aubhik & Julia K. Thomas. 2008. 'Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics'. *Econometrica* 76 (2): 395–436.

Kharroubi, Enisse & Frank Smets. 2024. 'Monetary Policy with Profit-Driven Inflation'. *BIS Working Papers*. BIS Working Papers.

Kimball, Miles S. 1995. 'The Quantitative Analytics of the Basic Neomonetarist Model'. *Journal of Money, Credit and Banking* 27 (4): 1241–1277.

Klenow, Peter J. & Oleksiy Kryvtsov. 2008. 'State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?*'. *The Quarterly Journal of Economics* 123 (3): 863–904.

Klenow, Peter J. & Benjamin A. Malin. 2010. 'Microeconomic Evidence on Price-Setting☆'. In *Handbook of Monetary Economics*, edited by Benjamin M. Friedman & Michael Woodford, 3:231–284. Elsevier.

Kryvtsov, Oleksiy & Virgiliu Midrigan. 2013. 'Inventories, Markups, and Real Rigidities in Menu Cost Models'. *The Review of Economic Studies* 80 (1): 249–276.

Kumar, Anil & Pia M. Orrenius. 2016. 'A Closer Look at the Phillips Curve Using State-Level Data'. *Journal of Macroeconomics* 47. What Monetary Policy Can and Cannot Do: 84–102.

Liu, Yang. 2025. 'Inflation and Competition in an Old Keynesian Model'. Boston College.

Marsal, Ales, Katrin Rabitsch & Lorant Kaszab. 2023. 'From Linear to Nonlinear: Rethinking Inflation Dynamics in the Calvo Pricing Mechanism'. *Department of Economics Working Papers*. Department of Economics Working Papers.

Michaillat, Pascal & Emmanuel Saez. 2015. 'Aggregate Demand, Idle Time, and Unemployment *'. *The Quarterly Journal of Economics* 130 (2): 507–569.

Miranda-Agrippino, Silvia & Giovanni Ricco. 2021. 'The Transmission of Monetary Policy Shocks'. *American Economic Journal: Macroeconomics* 13 (3): 74–107.

Montag, Hugh & Daniel Villar. 2025. 'Post-Pandemic Price Flexibility in the U.S.: Evidence and Implications for Price Setting Models'.

Nakamura, Emi & Jón Steinsson. 2008. 'Five Facts about Prices: A Reevaluation of Menu Cost Models*'. *The Quarterly Journal of Economics* 123 (4): 1415–1464.

———. 2010. 'Monetary Non-Neutrality in a Multisector Menu Cost Model*'. *The Quarterly Journal of Economics* 125 (3): 961–1013.

Nakamura, Emi, Jón Steinsson, Patrick Sun & Daniel Villar. 2018. 'The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation*'. *The*

Quarterly Journal of Economics 133 (4): 1933–1980.

Posch, Olaf. 2018. 'Resurrecting the New-Keynesian Model: (Un)Conventional Policy and the Taylor Rule'. *CESifo Working Paper Series*. CESifo Working Paper Series.

Posch, Olaf, Juan F. Rubio-Ramírez & Jesús Fernández-Villaverde. 2011. 'Solving the New Keynesian Model in Continuous Time'. *2011 Meeting Papers*. 2011 Meeting Papers.

Smets, Frank & Rafael Wouters. 2007. 'Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach'. *American Economic Review* 97 (3): 586–606.

Svensson, Lars EO. 1984. *Sticky Goods Prices, Flexible Asset Prices, and Optimum Monetary Policy*. IIES.

Appendix A The full extended model

TODO: WRITE UP.

Appendix B The full extended model without rationing

TODO: WRITE UP.

Appendix C Further proofs

C.1 Microeconomic welfare under quantity-capped rationing versus meeting demand

Suppose the demand curve is $\hat{p} = Ay^{-\frac{1}{\epsilon}}$ ($\hat{p}$ is the real price, $A > 0$, y is quantity) and the marginal cost curve is $\hat{q} = By^{\frac{\alpha}{1-\alpha}}$ ($\hat{q}$ is real marginal cost, $B > 0$). Then the efficient quantity is $\left(\frac{A}{B}\right)^{1/\left(\frac{\alpha}{1-\alpha}+\frac{1}{\epsilon}\right)}$ and the efficient price is $\hat{p}^* = A^{\frac{\alpha\epsilon}{1+\alpha(\epsilon-1)}}B^{\frac{1-\alpha}{1+\alpha(\epsilon-1)}}$. Assume the demand curve is flatter than the marginal cost curve, so $\frac{1}{\epsilon} < \frac{\alpha}{1-\alpha}$, i.e. $\alpha > \frac{1}{\epsilon+1}$.

Welfare is the difference between the integral under the demand curve and the integral under the marginal cost curve, which is:

$$\frac{\epsilon}{\epsilon-1}Ay^{\frac{\epsilon-1}{\epsilon}} - (1-\alpha)By^{\frac{1}{1-\alpha}}.$$

Suppose that $\hat{p} < \hat{p}^*$. Then output under rationing is $\left(\frac{\hat{p}}{B}\right)^{\frac{1-\alpha}{\alpha}}$ and output without rationing is $\left(\frac{\hat{p}}{A}\right)^{-\epsilon}$. Thus, the welfare gain of rationing over producing the full quantity demanded is:

$$\begin{aligned}
&\frac{\epsilon}{\epsilon-1}A\left(\frac{\hat{p}}{B}\right)^{\frac{1-\alpha}{\alpha}\frac{\epsilon-1}{\epsilon}} - (1-\alpha)B\left(\frac{\hat{p}}{B}\right)^{\frac{1}{\alpha}} - \frac{\epsilon}{\epsilon-1}A\left(\frac{\hat{p}}{A}\right)^{-(\epsilon-1)} + (1-\alpha)B\left(\frac{\hat{p}}{A}\right)^{-\frac{\epsilon}{1-\alpha}} \\
&= \frac{\epsilon}{\epsilon-1}AB^{-\frac{1-\alpha}{\alpha}\frac{\epsilon-1}{\epsilon}}\hat{p}^{\frac{1-\alpha}{\alpha}\frac{\epsilon-1}{\epsilon}} - (1-\alpha)B^{-\frac{1-\alpha}{\alpha}}\hat{p}^{\frac{1}{\alpha}} - \frac{\epsilon}{\epsilon-1}A^{\epsilon}\hat{p}^{-(\epsilon-1)} \\
&\qquad + (1-\alpha)A^{\frac{\epsilon}{1-\alpha}}B\hat{p}^{-\frac{\epsilon}{1-\alpha}} \\
&= (1-\alpha)A^{\frac{\epsilon}{1+\alpha(\epsilon-1)}}B^{-\frac{(1-\alpha)(\epsilon-1)}{1+\alpha(\epsilon-1)}}\left[\frac{1}{1-\alpha}\frac{\epsilon}{\epsilon-1}\left[\left(\frac{\hat{p}}{\hat{p}^*}\right)^{\frac{1-\alpha}{\alpha}\frac{\epsilon-1}{\epsilon}} - \left(\frac{\hat{p}}{\hat{p}^*}\right)^{-(\epsilon-1)}\right]\right. \\
&\qquad \left. - \left[\left(\frac{\hat{p}}{\hat{p}^*}\right)^{\frac{1}{\alpha}} - \left(\frac{\hat{p}}{\hat{p}^*}\right)^{-\frac{\epsilon}{1-\alpha}}\right]\right].
\end{aligned}$$

Let $c := (1-\alpha)\frac{\epsilon-1}{\epsilon} \in (0,1)$, $a := \frac{1}{\alpha} > 1$, $b := \frac{\epsilon}{1-\alpha} > 1$, $z := \frac{\hat{p}}{\hat{p}^*} \in (0,1)$ and let $f\colon (0,1] \to \mathbb{R}$ be defined by:

$$f(x) = \frac{z^{ax} - z^{-bx}}{x} - \left(z^a - z^{-b}\right),$$

for all $x \in (0,1)$. Then the welfare gain of rationing is:

$$(1-\alpha)A^{\frac{\epsilon}{1+\alpha(\epsilon-1)}}B^{-\frac{(1-\alpha)(\epsilon-1)}{1+\alpha(\epsilon-1)}}f(c).$$

Note that since $\alpha > \frac{1}{\epsilon+1}$, $b > a$, so $z^b < z^a < 1$ and hence for $x \in (0,1]$, $z^{ax} = (z^a)^x < 1 < \left(z^b\right)^{-x} = z^{-bx}$.

Next, observe that $f(1) = 0$, so to prove the welfare gain of rationing is strictly positive for $x \in (0,1)$, it is sufficient to prove that $f'(x) < 0$ for all $x \in (0,1]$. Now:[1]

$$\begin{aligned}
x^2 f'(x) + z^{ax} - z^{-bx} &= \left(axz^{ax} + bxz^{-bx}\right)(\log z) = \frac{az^{ax} + bz^{-bx}}{a+b}\log\left(z^{ax}z^{bx}\right) \\
&= \left[\frac{z^{ax} + z^{-bx}}{2} - \frac{\left(z^{ax} - z^{-bx}\right)(b-a)}{2(a+b)}\right]\log\left(z^{ax}z^{bx}\right) \\
&< \frac{z^{ax} + z^{-bx}}{2}\log\left(z^{ax}z^{bx}\right) < \frac{z^{ax} + z^{-bx}}{2}\frac{2\left(z^{ax}z^{bx} - 1\right)}{z^{ax}z^{bx} + 1} = z^{ax} - z^{-bx},
\end{aligned}$$

using the fact that $\log(u) < \frac{2(u-1)}{u+1}$ for $u \in (0,1)$. Hence, $f'(x) < 0$ for all $x \in (0,1]$ and so $f(x) > 0$ for all $x \in (0,1)$. Therefore, the welfare gain of rationing over production of the total quantity demanded is strictly positive.

[1] This proof follows the one given here: https://math.stackexchange.com/questions/4989707/.