Bank of Japan Working Paper Series

A Theory of Intrinsic Inflation Persistence

Takushi Kurozumi*
takushi.kurozumi@boj.or.jp

Willem Van Zandweghe**
willem.vanzandweghe@clev.frb.org

No.23-E-3
March 2023

Bank of Japan
2-1-1 Nihonbashi-Hongokucho, Chuo-ku, Tokyo 103-0021, Japan

* Monetary Affairs Department, Bank of Japan
** Research Department, Federal Reserve Bank of Cleveland

A Theory of Intrinsic Inflation Persistence*

Takushi Kurozumi† Willem Van Zandweghe‡

Abstract

We propose a novel theory of intrinsic inflation persistence by introducing trend inflation and Kimball (1995)-type aggregators of individual differentiated goods and labor in a model with staggered price- and wage-setting. Under nonzero trend inflation, the non-CES aggregator of goods and staggered price-setting give rise to a variable real marginal cost of goods aggregation, which becomes a driver of inflation. This marginal cost consists of an aggregate of the goods' relative prices, which depends on past inflation, thereby generating intrinsic inertia in inflation. Likewise, the non-CES aggregator of labor and staggered wage-setting lead to intrinsic inertia in wage inflation, which enhances the persistence of price inflation. With the theory we show that inflation exhibits a persistent, hump-shaped response to monetary policy shocks. We also demonstrate that lower trend inflation reduces inflation persistence and that a credible disinflation leads to a gradual decline in inflation and a fall in output.

JEL Classification: E31, E52

Keywords: Trend inflation, Non-CES aggregator, Credible disinflation

---

*The authors are grateful to Argia Sbordone (discussant), Marco Bassetto, Susanto Basu, Toni Braun, Brent Bundick, Olivier Coibion, Ferre De Graeve, Michael Dotsey, Andrew Foerster, Jeffrey Fuhrer, Mark Gertler, Kinda Hachem, Narayana Kocherlakota, Pok-sang Lam, Michal Marencak, Kiminori Matsuyama, Emi Nakamura, Roberto Pinheiro, Juan Rubio-Ramírez, Slavik Sheremirov, Lee Smith, Joe Vavra, Raf Wouters, and participants at 2017 NASMES, 2017 SNDE, Spring 2017 MMM, 2015 Federal Reserve System Meeting on Macroeconomics, SWET 2017, 2019 CEF, 6th CIGS End of Year Macroeconomics Conference, 2020 Econometric Society World Congress, and seminars at the Federal Reserve Banks of Boston, Cleveland, and Kansas City, the Federal Reserve Board, the National Bank of Belgium, and the University of Kansas for comments and discussions. Martin DeLuca provided excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Japan, the Federal Reserve Bank of Cleveland, or the Federal Reserve System.

†Monetary Affairs Department, Bank of Japan, 2-1-1 Nihonbashi-Hongokucho, Chuo-ku, Tokyo, 103-8660, Japan (E-mail: takushi.kurozumi@boj.or.jp).

‡Research Department, Federal Reserve Bank of Cleveland, P.O. Box 6387, Cleveland, OH 44101, USA (E-mail: willem.vanzandweghe@clev.frb.org).

"Taken as a whole, accordingly, the results suggest that it is worth searching for explanations of inflation inertia beyond the traditional ones that rely heavily on arbitrary lags."—Galí and Gertler (1999, p. 219).

1 Introduction

The well-known persistent response of inflation to monetary policy shocks has been documented by a large empirical literature. Christiano et al. (2011), for instance, use a structural vector autoregression (VAR) to show that inflation responds gradually to a shock to the monetary policy rate and that its peak response is delayed until some time after the shock. The source of inflation persistence is a longstanding question of interest to academic economists and monetary policy-makers. Many previous studies have accounted for inflation persistence by embedding price indexation to past inflation (Christiano et al., 2005; Smets and Wouters, 2007) or backward-looking rule-of-thumb price-setters (Galí and Gertler, 1999) in dynamic stochastic general equilibrium (DSGE) models.[1] These assumptions generate intrinsic inertia in inflation, but remain controversial because they are ad hoc assumptions that rely on non-optimizing price-setting behavior. Moreover, the price indexation implies that all prices change in every period, which contradicts the micro evidence that many individual prices remain unchanged for several months, as argued by Woodford (2007). In addition, Benati (2008) questions the assumptions, because they "hardwire" intrinsic inertia of inflation in models and imply that the degree of intrinsic inflation persistence is policy invariant, which contrasts with the result of his historical empirical analysis that the degree of inflation persistence varies across monetary policy regimes.[2]

Our paper proposes a novel theory of intrinsic inflation persistence by introducing trend inflation and Kimball (1995)-type aggregators of individual differentiated goods and labor—which relax the requirement of a constant elasticity of substitution (CES) between the goods or labor—in a DSGE model with Calvo (1983)-style staggered price- and wage-setting.[3]

[1] Woodford (2007) reviews different theories of intrinsic inertia in inflation. Fuhrer (2011) discusses the distinction between "intrinsic" versus "inherited" persistence in inflation.

[2] Hofmann et al. (2012) present empirical evidence of changes in wage dynamics over time, and similarly argue that hardwiring the degree of intrinsic persistence in wage inflation can be misleading.

[3] In the macroeconomic literature, Kimball (1995)-type non-CES aggregators have been widely used as

Nonzero trend inflation affects inflation dynamics in the model because in each period some prices remain unchanged in line with micro evidence.[4] Then, under staggered price-setting, the Kimball-type non-CES aggregator of individual goods gives rise to a variable real marginal cost of *aggregating* the goods, which becomes a driver of inflation. Moreover, the real marginal cost consists of an aggregate of the goods' relative prices, which depends on current and past inflation rates, thereby generating intrinsic inertia in inflation. Likewise, under staggered wage-setting, the non-CES aggregator of labor leads to intrinsic inertia in wage inflation, which enhances the persistence of price inflation. Therefore, our model provides a theoretical justification for intrinsic inflation persistence without relying on ad hoc backward-looking price- and wage-setting behavior and hence responds to the suggestion of Galí and Gertler (1999) in the opening quote. Consequently, a plausibly calibrated version of the model shows that inflation exhibits a persistent, hump-shaped response to monetary policy shocks, as documented by the empirical literature.

Why does the Kimball-type non-CES aggregator of individual differentiated goods lead the real marginal cost of aggregating the goods to vary under nonzero trend inflation and staggered price-setting? The real marginal cost equalizes each good's ratio of the relative price to marginal product for profit maximization, where the marginal product depends inversely on the demand for the good. Suppose that an expansionary monetary policy shock hits the economy. Then, the dispersion of relative prices increases, since firms that can adjust their goods' prices raise them, while other firms keep prices unchanged and thus have their relative prices eroded by inflation. The increased price dispersion leads to a shift in demand away from goods with higher relative prices toward those with lower ones, thus moving the marginal product of each good in the same direction as its relative price. In the case of the CES aggregator, the price elasticity of demand for goods is constant, allowing each good's marginal product to shift proportionally with its relative price and hence their ratio (the real marginal cost) remains constant. By contrast, the non-CES aggregator can give rise to a

a source of strategic complementarity in price setting; see, e.g., Eichenbaum and Fisher (2007), Smets and Wouters (2007), and Levin et al. (2008). Dotsey and King (2005) introduce a Kimball-type aggregator in a state-dependent price-setting model to show that it enhances the persistence of output and inflation.

[4] For micro evidence on price setting, see, e.g., Klenow and Kryvtsov (2008), Nakamura and Steinsson (2008), and Nakamura et al. (2018). Ascari and Sbordone (2014) survey the literature on the role of trend inflation in inflation dynamics.

positive superelasticity of demand, which assigns a larger elasticity of demand to goods with higher relative prices.[5] Accordingly, the marginal products of individual goods with higher relative prices show muted increases, whereas those of individual goods with lower relative prices exhibit sharper decreases. The real marginal cost then rises to equalize each good's ratio of the relative price to marginal product. Therefore, an increase in the dispersion of relative prices arising under nonzero trend inflation and staggered price-setting generates a rise in the real marginal cost of aggregating individual goods.

Our model provides a microfoundation of intrinsic inertia in price and wage inflation by relating the degrees of intrinsic inertia to structural parameters of the model. Consequently, the model is not subject to the criticism by Benati (2008) of models in which intrinsic inflation persistence is policy invariant. In particular, the degrees of intrinsic inertia in price and wage inflation are related to the rate of trend inflation, which represents the inflation target of the monetary authority in our model. We then show that lower trend inflation reduces inflation persistence. A number of empirical studies, such as Cogley and Sargent (2002), Stock and Watson (2007), Cogley et al. (2010), and Fuhrer (2011), indicate that inflation persistence has decreased in the U.S. since the early 1980s, around the time of the Volcker disinflation.[6] The leading explanation for the decrease in inflation persistence by existing studies, such as Benati and Surico (2008), Carlstrom et al. (2009), and Davig and Doh (2014), emphasizes a more active monetary policy response to inflation.[7] Our paper provides a new explanation: the fall in trend inflation caused the decrease in inflation persistence.

The paper also contributes to the literature on disinflation, including Ball (1994a), Fuhrer and Moore (1995), and Mankiw and Reis (2002). As Fuhrer (2011) points out, intrinsic inertia in inflation plays a key role in New Keynesian (NK) models, where a credible permanent reduction in trend inflation induces a gradual adjustment of inflation to its new trend rate and a decline in output. These responses align closely with historical evidence that disinflations tend to be gradual and accompanied by a recession (see, e.g., Gordon, 1982; Ball, 1994b;

[5]The term superelasticity of demand refers to the elasticity of the elasticity of demand.

[6]Owing to differences in methodology and measures of inflation, not all studies point to a change in inflation persistence in the post-World War II period. See, e.g., Pivetta and Reis (2007).

[7]Cogley et al. (2010) explain a decrease in the persistence of the inflation gap (i.e., the gap between actual and trend inflation) after the Volcker disinflation by introducing a time-varying inflation target in a Taylor (1993)-type monetary policy rule using a New Keynesian model.

Cecchetti and Rich, 2001). Without the intrinsic inertia in NK models, inflation jumps to its new trend rate, while output never deviates from its steady-state value. By contrast, in our model, a credible disinflation leads to a gradual decline in inflation and a fall in output even though price-setting decisions are purely forward-looking. This is because our model has intrinsic inertia of inflation through the aggregate of relative prices that arises from the Kimball-type non-CES aggregator of goods, as noted above.

A few previous studies have also explained inflation persistence using DSGE models without backward-looking price-setting behavior. Mankiw and Reis (2002) develop a sticky information model to account for the persistent response of inflation to monetary policy shocks. Dupor et al. (2010) introduce sticky information in a model with staggered price-setting and find that lagged inflation appears in an NK Phillips curve derived from the model. A similar finding is obtained by Sheedy (2010), who instead incorporates an upward-sloping hazard function in the model so that prices are more likely to be changed as they have remained fixed for longer. Compared with these studies, our paper emphasizes the role of trend inflation in inflation persistence and thus it is not subject to the criticism by Benati (2008) that monetary policy is relevant for inflation persistence. Damjanovic and Nolan (2010) employ a staggered price-setting model with trend inflation and a CES aggregator of individual goods, and indicate that a longer average duration of price change—their calibration of which is two years—makes relative price distortion—which has a first-order effect on the real marginal cost of *producing* the goods under nonzero trend inflation—more amplified and more persistent, thus generating a more persistent response of inflation to monetary policy shocks. At the same time, however, they show that the distortion also generates a counterfactual decline in output after an expansionary monetary policy shock.[8] Cogley and Sbordone (2008) embed not only price indexation to past inflation but also drifting trend inflation in a model with staggered price-setting under subjective expectations based on the anticipated utility model of Kreps (1998), which are distinct from rational expectations assumed in our paper.[9] They then empirically show that intrinsic inflation inertia arising from the price indexation is not needed for the model to explain U.S. inflation

[8]In their conclusion, Damjanovic and Nolan (2010) point out that "further work is required to understand this and reconcile it with how one typically thinks the economy responds to such a shock" (p. 1096).

[9]Such subjective expectations are needed to incorporate the drifting trend inflation in their model.

dynamics in the presence of the drifting trend inflation, which makes the gap between actual and trend inflation less persistent.[10]

The remainder of the paper proceeds as follows. Section 2 presents a model with trend inflation, staggered price- and wage-setting, and Kimball-type aggregators of goods and labor. Section 3 shows that a plausibly calibrated version of the model can explain the well-known persistent response of inflation to monetary policy shocks. Using the calibrated model, Section 4 shows that lower trend inflation reduces inflation persistence. Section 5 demonstrates that a credible disinflation leads to a gradual decline in inflation and a fall in output. Section 6 concludes.

2 Model

To account for inflation persistence, this paper uses a DSGE model with trend inflation, Calvo (1983)-style staggered price- and wage-setting, and Kimball (1995)-type aggregators of individual differentiated goods and labor. The model consists of a representative composite-good producer, individual-goods producing firms, a representative household with individual workers as its members, a representative labor packer, and a monetary authority. A key feature of the model is that each period a fraction of individual goods' prices remains unchanged in line with micro evidence, while other prices are set by firms that face demand curves with positive superelasticity arising from the non-CES aggregator of goods. Likewise, in each period, a fraction of individual workers' nominal wages remains unchanged, while other nominal wages are chosen for labor demand curves with positive superelasticity stemming from the non-CES aggregator of labor. The behavior of each economic agent is described in what follows.

2.1 Composite-good producer

There are a representative composite-good producer and a continuum of firms $f \in [0,1]$, each of which produces an individual differentiated good $Y_t(f)$. As in Kimball (1995), the

[10] Phaneuf et al. (2018) use a DSGE model with a roundabout production structure and working capital to demonstrate that even in the absence of intrinsic inertia, inflation can exhibit persistence inherited from the real marginal cost of producing individual goods.

composite good Y_t is produced by aggregating individual goods $\{Y_t(f)\}$ with

$$\int_0^1 F_p\left(\frac{Y_t(f)}{Y_t}\right) df = 1. \tag{1}$$

Following Dotsey and King (2005) and Levin et al. (2008), the function $F_p(\cdot)$ is assumed to be of the form

$$F_p\left(\frac{Y_t(f)}{Y_t}\right) = \frac{\gamma_p}{(1+\epsilon_p)(\gamma_p-1)}\left[(1+\epsilon_p)\frac{Y_t(f)}{Y_t} - \epsilon_p\right]^{\frac{\gamma_p-1}{\gamma_p}} + 1 - \frac{\gamma_p}{(1+\epsilon_p)(\gamma_p-1)},$$

where $\gamma_p \equiv \theta_p(1+\epsilon_p)$ and ϵ_p is a constant. In the special case of $\epsilon_p = 0$, the Kimball-type aggregator (1) is reduced to the CES one $Y_t = \left[\int_0^1 (Y_t(f))^{(\theta_p-1)/\theta_p}\, df\right]^{\theta_p/(\theta_p-1)}$, where $\theta_p > 1$ denotes the elasticity of substitution between individual differentiated goods.

The composite-good producer maximizes profit $P_t Y_t - \int_0^1 P_t(f) Y_t(f)\, df$ subject to the Kimball-type aggregator (1), given the composite good's price P_t and individual goods' prices $\{P_t(f)\}$. Combining the first-order conditions for profit maximization and the aggregator (1) yields

$$\frac{Y_t(f)}{Y_t} = \frac{1}{1+\epsilon_p}\left[\left(\frac{P_t(f)}{P_t\, d_{p,t}}\right)^{-\gamma_p} + \epsilon_p\right], \tag{2}$$

$$d_{p,t} = \left[\int_0^1 \left(\frac{P_t(f)}{P_t}\right)^{1-\gamma_p} df\right]^{\frac{1}{1-\gamma_p}}, \tag{3}$$

$$1 = \frac{1}{1+\epsilon_p}\, d_{p,t} + \frac{\epsilon_p}{1+\epsilon_p}\, e_{p,t}, \tag{4}$$

where $d_{p,t}$ is the Lagrange multiplier on the aggregator (1) and

$$e_{p,t} \equiv \int_0^1 \frac{P_t(f)}{P_t}\, df. \tag{5}$$

The Lagrange multiplier $d_{p,t}$ represents the *real marginal cost of aggregating individual differentiated goods* (or producing the composite good), and consists of the aggregate of relative prices of individual goods that corresponds to the Kimball-type aggregator (1), as shown in (3). In the special case of $\epsilon_p = 0$, where the Kimball-type aggregator (1) becomes the CES one as noted above, eqs. (2)–(4) can be reduced to $Y_t(f)/Y_t = (P_t(f)/P_t)^{-\theta_p}$,

$P_t = \left[\int_0^1 (P_t(f))^{1-\theta_p}\, df\right]^{1/(1-\theta_p)}$, and $d_{p,t} = 1$, respectively. The last equation indicates that the real marginal cost is constant in the case of the CES aggregator. Moreover, if all firms share the same production technology (as assumed later) and all individual goods' prices are flexible, the prices are all identical and thus eqs. (3) and (4) imply that $d_{p,t} = 1$ even in the case of the non-CES aggregator, i.e., $\epsilon_p \neq 0$.

Eq. (2) is the demand curve for each individual good $Y_t(f)$ and has the feature that the superelasticity (i.e., the elasticity of the elasticity) of demand is not necessarily zero. The elasticity and the superelasticity of demand are given by $\eta_{p,t} = \theta_p \left[1 + \epsilon_p - \epsilon_p (Y_t(f)/Y_t)^{-1}\right]$ and $-\epsilon_p \theta_p (Y_t(f)/Y_t)^{-1}$, respectively. Thus, the parameter ϵ_p governs the superelasticity, in particular its sign. Since $\theta_p > 1$, a negative value of ϵ_p leads to a positive superelasticity, which causes the elasticity $\eta_{p,t}$ to vary inversely with relative demand $Y_t(f)/Y_t$, that is, relative demand for each individual good to become more price-elastic for an increase in the relative price of the good and less price-elastic for a decrease in the relative price. As is well understood, this feature induces strategic complementarity in price setting, because firms that face the increasing elasticity keep their goods' relative prices near those of other firms (when they can adjust prices). By contrast, in the case of $\epsilon_p = 0$, the elasticity becomes constant (i.e., $\eta_{p,t} = \theta_p$), so that the superelasticity is zero.

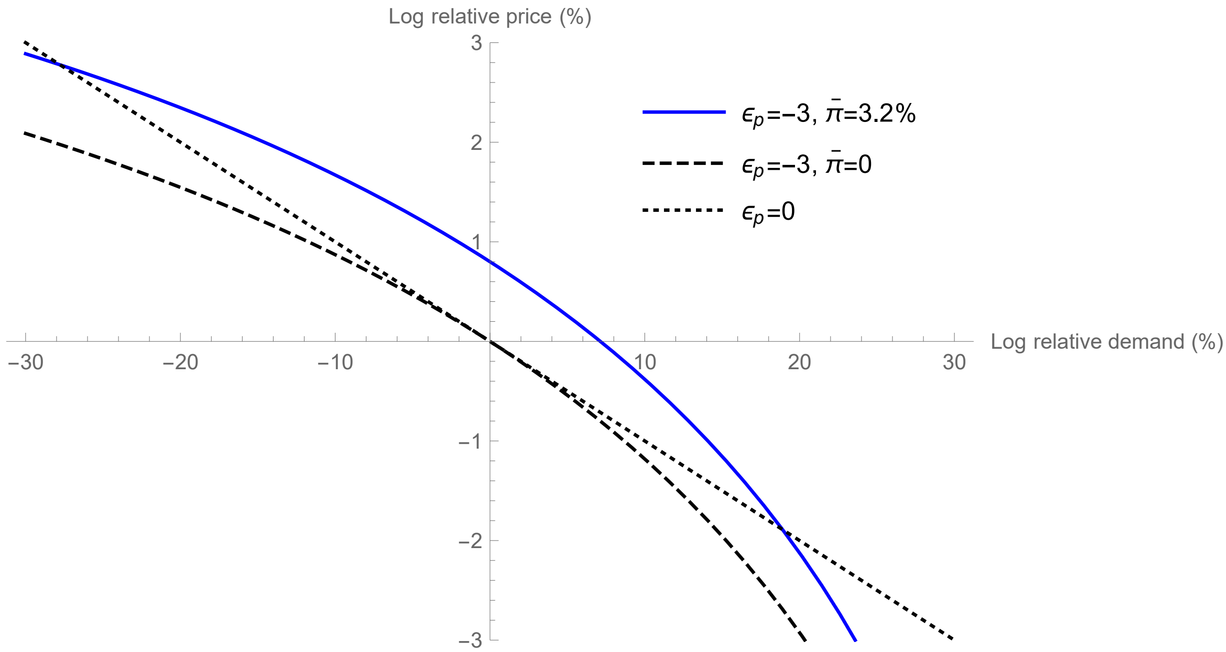
A positive superelasticity of demand (i.e., a negative value of ϵ_p) leads higher inflation to raise the real marginal cost of goods aggregation $d_{p,t}$. Figure 1 illustrates the log of the inverse demand curve

$$\frac{P_t(f)}{P_t} = \left[\frac{Y_t(f)}{Y_t}(1 + \epsilon_p) - \epsilon_p\right]^{-\frac{1}{\gamma_p}} d_{p,t},$$

keeping the real marginal cost $d_{p,t}$ at its steady-state value under firms' staggered price-setting that is explained in the next subsection. For each good $Y_t(f)$, the left-hand side of the inverse demand curve is its relative price and the right-hand side consists of its marginal product $[(Y_t(f)/Y_t)(1 + \epsilon_p) - \epsilon_p]^{-1/\gamma_p}$ and the real marginal cost $d_{p,t}$. The figure uses two values of the superelasticity parameter, $\epsilon_p = 0$ (the dotted line) and $\epsilon_p = -3$. For the latter, the inverse demand curve is shown under a trend inflation rate $\bar{\pi}\,(\equiv 4\log\pi)$ of zero (the dashed line) and 3.2 percent annually (the solid line).[11] If trend inflation is zero then the

[11]Throughout the paper, π denotes the gross trend inflation rate and $\bar{\pi}$ denotes the net annualized trend inflation rate.

Figure 1: Inverse demand curves with positive superelasticity and constant elasticity.



Notes: In the figure the dotted line displays the case of $\epsilon_p = 0$, that is, the CES aggregator of goods. The dashed and the solid lines illustrate the case of $\epsilon_p = -3$, that is, the Kimball (1995)-type non-CES aggregator under a trend inflation rate of zero and 3.2 percent annually, respectively. The values of other model parameters used here are reported in Table 1 below. The real marginal cost of goods aggregation $d_{p,t}$ is fixed at its steady-state value under firms' staggered price-setting that is explained in the next subsection.

real marginal cost is unity, so that the dashed line displays the marginal product of the good and shows that a rise in the relative price leads to a muted increase in the marginal product, whereas a decline in the relative price leads to a sharper decrease in the marginal product. Under the annualized trend inflation rate of 3.2 percent, a rise in the real marginal cost $d_{p,t}$ shifts up the demand curve, which magnifies the muted increase of the marginal product for a rise in the relative price and dampens the sharp decrease of the marginal product for a decline in the relative price. Thus, higher inflation, by increasing the dispersion in relative prices of individual differentiated goods, leads the real marginal cost to rise in order to equate each good's ratio of the relative price to marginal product.[12]

[12] Likewise, an increase in the degree of nominal price rigidity would increase price dispersion and the real marginal cost $d_{p,t}$. Conversely, if all individual goods' prices are flexible and all firms share the same production technology (as assumed later), the real marginal cost $d_{p,t}$ is unity even in the case of the non-CES aggregator (i.e., $\epsilon_p \neq 0$) as noted above, thus resulting in no shift in the inverse demand curve.

2.2 Firms

Each firm f produces one kind of differentiated good $Y_t(f)$ using the production technology

$$Y_t(f) = N_t(f), \tag{6}$$

where $N_t(f)$ is labor input of firm f. The firm minimizes cost $w_t N_t(f)$ subject to the technology (6), given the real wage (i.e., the relative price of labor input) $w_t \, (\equiv W_t/P_t)$. The first-order condition for cost minimization shows that each firm's *real marginal cost of producing its differentiated good* mc_t is identical and equal to the real wage:

$$mc_t = w_t. \tag{7}$$

In the face of the demand curve (2) and the real marginal cost mc_t, firms set their goods' prices on a staggered basis as in Calvo (1983). In each period, a fraction $\alpha_p \in (0,1)$ of firms keeps prices unchanged, while the remaining fraction $1-\alpha_p$ of firms sets the price $P_t(f)$ so as to maximize relevant profit

$$E_t \sum_{j=0}^{\infty} \alpha_p^j \, q_{t,t+j} \left(\frac{P_t(f)}{P_{t+j}} - mc_{t+j} \right) \frac{Y_{t+j}}{1+\epsilon_p} \left[\left(\frac{P_t(f)}{P_{t+j}\, d_{p,t+j}} \right)^{-\gamma_p} + \epsilon_p \right],$$

where E_t denotes the expectation operator conditional on information available in period t and $q_{t,t+j}$ is the (real) stochastic discount factor between period t and period $t+j$.

Using the equilibrium condition $q_{t,t+j} = \beta^j C_t/C_{t+j}$ for the household's log utility of consumption C_t with its subjective discount factor $\beta \in (0,1)$ and the composite-good market clearing condition

$$Y_t = C_t, \tag{8}$$

the first-order condition for profit maximization can be written as

$$E_t \sum_{j=0}^{\infty} (\alpha_p \beta)^j \left[\left(\frac{p_t^*}{d_{p,t+j}} \prod_{k=1}^{j} \frac{1}{\pi_{t+k}} \right)^{-\gamma_p} \left(p_t^* \prod_{k=1}^{j} \frac{1}{\pi_{t+k}} - \frac{\gamma_p}{\gamma_p - 1} mc_{t+j} \right) - \frac{\epsilon_p}{\gamma_p - 1} p_t^* \prod_{k=1}^{j} \frac{1}{\pi_{t+k}} \right] = 0, \tag{9}$$

where $\pi_t = P_t/P_{t-1}$ is the gross inflation rate of the composite good's price and p_t^* is the relative price set by firms that can adjust prices in period t. Moreover, under staggered price-setting, eqs. (3) and (5) can be reduced to, respectively,

$$(d_{p,t})^{1-\gamma_p} = \alpha_p \left(\frac{d_{p,t-1}}{\pi_t} \right)^{1-\gamma_p} + (1-\alpha_p)(p_t^*)^{1-\gamma_p}, \tag{10}$$

$$e_{p,t} = \alpha_p \left(\frac{e_{p,t-1}}{\pi_t} \right) + (1-\alpha_p)\, p_t^*. \tag{11}$$

Eq. (10) represents a law of motion of the real marginal cost of goods aggregation $d_{p,t}$, which consists of the relative-price aggregate corresponding to the Kimball-type goods aggregator (1) as shown in (3), thus yielding the law of motion under staggered price-setting.

The labor market clearing condition is given by $N_t = \int_0^1 N_t(f)\, df$, where N_t is labor input supplied by the labor packer. Combining this condition with the demand curve (2) and the production technology (6) leads to

$$Y_t = \frac{N_t}{\Delta_t}, \tag{12}$$

where

$$\Delta_t \equiv \frac{s_t + \epsilon_p}{1 + \epsilon_p} \tag{13}$$

represents the relative price distortion and

$$s_t \equiv \int_0^1 \left(\frac{P_t(f)}{P_t\, d_{p,t}} \right)^{-\gamma_p} df, \tag{14}$$

which can be reduced, under staggered price-setting, to

$$s_t\, (d_{p,t})^{-\gamma_p} = \alpha_p s_{t-1} \left(\frac{d_{p,t-1}}{\pi_t} \right)^{-\gamma_p} + (1-\alpha_p)(p_t^*)^{-\gamma_p}. \tag{15}$$

2.3 Household and labor packer

The representative household consumes the composite good C_t, purchases one-period bonds B_t, and has a continuum of members $h \in [0, 1]$, each of which supplies an individual differentiated labor service $N_t(h)$, so as to maximize the utility function

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\log(C_t) - \int_0^1 \frac{(N_t(h))^{1+\sigma_n}}{1+\sigma_n}\, dh \right]$$

subject to the budget constraint

$$P_t C_t + B_t = \int_0^1 W_t(h) N_t(h)\, dh + i_{t-1} B_{t-1} + T_t, \tag{16}$$

where $\sigma_n \geq 0$ is the inverse of the elasticity of labor supply, $W_t(h)$ is the nominal wage of the labor service $N_t(h)$, i_t is the gross interest rate on the bonds and is assumed to coincide with the monetary policy rate, and T_t consists of lump-sum taxes and transfers and firm profits received.

Assuming additive separability in preferences and complete contingent-claims markets for consumption implies that all members make a joint consumption–saving decision. Thus, combining the first-order conditions for utility maximization with respect to consumption and bond holdings yields the consumption Euler equation

$$1 = E_t\left(\frac{\beta C_t}{C_{t+1}} \frac{i_t}{\pi_{t+1}}\right). \tag{17}$$

The representative labor packer supplies labor input N_t to firms by aggregating individual labor services $\{N_t(h)\}$ with

$$\int_0^1 F_w\left(\frac{N_t(h)}{N_t}\right) dh = 1, \tag{18}$$

where the function $F_w(\cdot)$ takes the same form as $F_p(\cdot)$, but with parameters ϵ_w, θ_w, and γ_w (instead of ϵ_p, θ_p, and γ_p). Note that $\theta_w > 1$ and $\gamma_w \equiv \theta_w(1+\epsilon_w)$. As is similar to the goods aggregator (1) with $\epsilon_p \leq 0$, we consider the case of $\epsilon_w \leq 0$ in the following sections. The labor packer maximizes profit $W_t N_t - \int_0^1 W_t(h) N_t(h)\, dh$ subject to the Kimball-type labor aggregator (18), given the nominal wage W_t and individual labor services' nominal wages $\{W_t(h)\}$. Combining the first-order conditions for profit maximization and the aggregator (18) yields

$$\frac{N_t(h)}{N_t} = \frac{1}{1+\epsilon_w}\left[\left(\frac{W_t(h)}{W_t\, d_{w,t}}\right)^{-\gamma_w} + \epsilon_w\right], \tag{19}$$

$$d_{w,t} = \left[\int_0^1 \left(\frac{W_t(h)}{W_t}\right)^{1-\gamma_w} dh\right]^{\frac{1}{1-\gamma_w}}, \tag{20}$$

$$1 = \frac{1}{1+\epsilon_w} d_{w,t} + \frac{\epsilon_w}{1+\epsilon_w} e_{w,t}, \tag{21}$$

where $d_{w,t}$ is the Lagrange multiplier on the Kimball-type labor aggregator (18) and

$$e_{w,t} \equiv \int_0^1 \frac{W_t(h)}{W_t}\, dh. \tag{22}$$

The Lagrange multiplier $d_{w,t}$ represents the real marginal cost of providing the labor input N_t, by aggregating individual labor services, and consists of the aggregate of their relative nominal wages that corresponds to the Kimball-type labor aggregator (18), as shown in (20).

Given the demand curve (19), nominal wages are chosen on a Calvo-style staggered basis. In each period, a fraction $\alpha_w \in (0,1)$ of nominal wages is kept unchanged, while the remaining fraction $1-\alpha_w$ of wages is chosen so as to maximize the relevant utility function

$$E_t \sum_{j=0}^{\infty} (\alpha_w \beta)^j \left[-\frac{\left(N_{t+j|t}(h)\right)^{1+\sigma_n}}{1+\sigma_n} + \Lambda_{t+j} \frac{W_t(h)}{P_{t+j}} N_{t+j|t}(h) \right]$$

subject to the demand curve

$$N_{t+j|t}(h) = \frac{N_{t+j}}{1+\epsilon_w} \left[\left(\frac{W_t(h)}{W_{t+j}\, d_{w,t+j}} \right)^{-\gamma_w} + \epsilon_w \right],$$

where Λ_t is the real value of the Lagrange multiplier on the household's budget constraint (16) and meets the first-order condition $\Lambda_t = 1/C_t$ for the log utility of consumption. Using the composite-good market clearing condition (8), the first-order condition for utility maximization with respect to the nominal wage can be written as

$$E_t \sum_{j=0}^{\infty} (\alpha_w \beta)^j \frac{N_{t+j}}{Y_{t+j}} \left[\begin{array}{l} \left(\dfrac{W_t^*/W_t}{d_{w,t+j}} \displaystyle\prod_{k=1}^{j} \frac{1}{\pi_{w,t+k}} \right)^{-\gamma_w} \\ \times \left(\dfrac{W_t^*}{W_t} \displaystyle\prod_{k=1}^{j} \frac{1}{\pi_{t+k}} - \frac{\gamma_w}{\gamma_w - 1} \left\{ \frac{N_{t+j}}{1+\varepsilon_w} \left[\left(\frac{W_t^*/W_t}{d_{w,t+j}} \prod_{k=1}^{j} \frac{1}{\pi_{w,t+k}} \right)^{-\gamma_w} + \varepsilon_w \right] \right\}^{\sigma_n} \frac{Y_{t+j}}{w_{t+j}} \prod_{k=1}^{j} \frac{w_{t+k}}{w_{t+k-1}} \right) \\ - \dfrac{\epsilon_w}{\gamma_w - 1} \dfrac{W_t^*}{W_t} \displaystyle\prod_{k=1}^{j} \frac{1}{\pi_{t+k}} \end{array} \right]$$

$$= 0, \tag{23}$$

where W_t^* is the nominal wage that is chosen in period t, and

$$\pi_{w,t} \equiv \frac{W_t}{W_{t-1}} = \frac{w_t}{w_{t-1}} \pi_t \tag{24}$$

denotes wage inflation. Moreover, under staggered wage-setting, eqs. (20) and (22) can be reduced to, respectively,

$$(d_{w,t})^{1-\gamma_w} = \alpha_w \left(\frac{d_{w,t-1}}{\pi_{w,t}} \right)^{1-\gamma_w} + (1 - \alpha_w) \left(\frac{W_t^*}{W_t} \right)^{1-\gamma_w}, \tag{25}$$

$$e_{w,t} = \alpha_w \left(\frac{e_{w,t-1}}{\pi_{w,t}} \right) + (1 - \alpha_w) \frac{W_t^*}{W_t}. \tag{26}$$

Eq. (25) represents a law of motion of the real marginal cost of labor aggregation $d_{w,t}$, which consists of the relative-nominal wage aggregate that corresponds to the Kimball-type labor aggregator (18), as shown in (20).

2.4 Monetary authority

The monetary authority conducts policy according to a Taylor (1993)-type rule. This rule adjusts the interest rate in response to deviations of inflation from its trend rate and deviations of output from its trend level, and allows for policy inertia:

$$\log i_t = \rho \log i_{t-1} + (1 - \rho)[\log i + \phi_\pi (\log \pi_t - \log \pi) + \phi_Y (\log Y_t - \log Y)] + \varepsilon_{i,t}, \tag{27}$$

where i is the gross steady-state interest rate; π is the gross trend inflation rate; Y denotes steady-state output; $\rho \in [0, 1)$, $\phi_\pi \geq 0$, and $\phi_Y \geq 0$ represent, respectively, the degrees of policy inertia, the policy response to inflation, and the one to output; and $\varepsilon_{i,t} \sim N(0, \sigma_i^2)$ is an i.i.d. monetary policy shock.

2.5 Log-linearized equilibrium conditions

The equilibrium conditions in the model consist of (4), (7)–(13), (15), (17), (21), and (23)–(27).

To demonstrate intrinsic inertia of inflation in the model, we derive a generalized NK Phillips curve (GNKPC) by log-linearizing the equilibrium conditions around the steady state with nonzero trend inflation π (see Appendix A for the details of the derivation). For the steady state to be derived explicitly, we assume a unit elasticity of labor supply (i.e., $\sigma_n = 1$), which is a common value in the macroeconomic literature. We also assume, to ensure that the steady state is well defined, that the following conditions are satisfied:

$$\alpha_p \max(\pi^{\gamma_p}, \pi^{\gamma_p - 1}, \pi^{-1}) < 1, \quad \alpha_w \max(\pi^{2\gamma_w}, \pi^{\gamma_w}, \pi^{\gamma_w - 1}, \pi^{-1}) < 1. \tag{28}$$

These conditions are always met in the special case of zero trend inflation, i.e., $\pi = 1$.

The GNKPC is then derived as

$$\hat{\pi}_t = \beta\pi E_t \hat{\pi}_{t+1} + \kappa_p \hat{mc}_t + \kappa_{pd} \hat{d}_{p,t} + \hat{d}_{p,t-1} + \beta\pi E_t \hat{d}_{p,t+1} + \varphi_{p,t} + \psi_{p,t}, \tag{29}$$

where hatted variables denote log-deviations from steady-state values, and $\varphi_{p,t}$ and $\psi_{p,t}$ are auxiliary variables that are additional drivers of inflation under nonzero trend inflation and satisfy

$$\varphi_{p,t} = \alpha_p \beta \pi^{\gamma_p - 1} E_t \varphi_{p,t+1} + \kappa_{p\varphi} \Big(\gamma_p (1 - \alpha_p \beta \pi^{\gamma_p - 1}) E_t \hat{d}_{p,t+1} + (\gamma_p - 1) E_t \hat{\pi}_{t+1} \Big), \tag{30}$$

$$\psi_{p,t} = \alpha_p \beta \pi^{-1} E_t \psi_{p,t+1} + \kappa_{p\epsilon\psi} E_t \hat{\pi}_{t+1}. \tag{31}$$

The composite coefficients κ_p, κ_{pd}, $\kappa_{p\varphi}$, and $\kappa_{p\epsilon\psi}$ in (29)–(31) consist of the model's structural parameters, including trend inflation π, and they are presented in Appendix B. Note that $\kappa_{p\varphi} = 0$ if $\pi = 1$ and that $\kappa_{p\epsilon\psi} = 0$ if $\pi = 1$ or $\epsilon_p = 0$. The law of motion of the real marginal cost of goods aggregation is given by

$$\hat{d}_{p,t} = \rho_{pd} \hat{d}_{p,t-1} + \kappa_{p\epsilon d}\, \hat{\pi}_t, \tag{32}$$

where

$$\rho_{pd} \equiv \frac{\alpha_p \pi^{-1}(1 + \epsilon_{p1}\pi^{\gamma_p})}{1 + \epsilon_{p1}}, \quad \kappa_{p\epsilon d} \equiv -\frac{\epsilon_{p1}\alpha_p \pi^{-1}(\pi^{\gamma_p} - 1)}{(1 + \epsilon_{p1})(1 - \alpha_p \pi^{-1})},$$

and $\epsilon_{p1} \equiv \epsilon_p[(1 - \alpha_p)/(1 - \alpha_p \pi^{\gamma_p - 1})]^{\gamma_p/(\gamma_p - 1)}$, so $\kappa_{p\epsilon d} = 0$ if $\pi = 1$ or $\epsilon_p = 0$.[13]

The presence of the real marginal cost $\hat{d}_{p,t}$ is a novel feature of the GNKPC (29). While the CES aggregator of goods (i.e., $\epsilon_p = 0$) keeps a unit real marginal cost (i.e., $d_{p,t} = 1$), the Kimball-type non-CES aggregator (1) with $\epsilon_p < 0$ leads the real marginal cost to vary

[13] Under the calibration of model parameters reported in Table 1 below, the values of κ_p, $\kappa_{p\varphi}$, ρ_{pd}, and $\kappa_{p\epsilon d}$ are positive, while those of κ_{pd} and $\kappa_{p\epsilon\psi}$ are negative.

under nonzero trend inflation and staggered price-setting. An economic intuition for this is as follows. Suppose that an expansionary monetary policy shock hits the economy. Then, the dispersion of relative prices increases, since price-adjusting firms raise their goods' prices, while non-adjusting firms have their relative prices eroded by inflation. The increased price dispersion leads to a shift in demand away from goods with higher relative prices toward those with lower ones, thus moving each good's marginal product in the same direction as its relative price. However, the non-CES aggregator assigns a larger elasticity of demand to goods with higher relative prices. This leads to not only muted increases in the marginal products of goods with higher relative prices but also sharper decreases in those of goods with lower relative prices. The real marginal cost then rises to equalize each good's ratio of the relative price to marginal product. Therefore, the real marginal cost $\hat{d}_{p,t}$ varies.

Moreover, the real marginal cost $\hat{d}_{p,t}$ is a source of intrinsic inertia in inflation. To see this, eq. (32) implies that the real marginal cost depends on current and past inflation rates: $\hat{d}_{p,t} = \kappa_{p\epsilon d} \sum_{j=0}^{\infty} \rho_{pd}^{j} \hat{\pi}_{t-j}$. Combining this and the GNKPC (29) leads to

$$\hat{\pi}_t = b_{p\epsilon 1} \sum_{j=1}^{\infty} \rho_{pd}^{j-1} \hat{\pi}_{t-j} + b_{p2} E_t \hat{\pi}_{t+1} + b_{p3} (\kappa_p \hat{mc}_t + \varphi_{p,t} + \psi_{p,t}) \,, \tag{33}$$

where $b_{p\epsilon 1} \equiv \kappa_{p\epsilon d} b_{p3}[1 + \rho_{pd}(\kappa_{pd} + \beta\pi\rho_{pd})]$, $b_{p2} \equiv \beta\pi b_{p3}(1 + \kappa_{p\epsilon d})$, and $b_{p3} \equiv 1/[1 - \kappa_{p\epsilon d}(\kappa_{pd} + \beta\pi\rho_{pd})]$. This shows that our model provides a theoretical justification for intrinsic inertia in inflation without relying on ad hoc backward-looking price-setting behavior. The degree of intrinsic inflation inertia can be summarized as the sum of the coefficients on lagged inflation rates, $\lambda_{p\epsilon} \equiv b_{p\epsilon 1} \sum_{j=1}^{\infty} \rho_{pd}^{j-1} = b_{p\epsilon 1}/(1 - \rho_{pd})$, and depends on the model's structural parameters, including trend inflation π and the superelasticity parameter ϵ_p. In particular, if $\pi = 1$ or $\epsilon_p = 0$ then $b_{p\epsilon 1} = 0$ because $\kappa_{p\epsilon d} = 0$, and thus $\lambda_{p\epsilon} = 0$ so intrinsic inertia of inflation is absent from (33).

In addition to intrinsic inertia in price inflation, the model has intrinsic inertia in wage inflation. The GNKPC for wage inflation (wage-GNKPC) can be derived as

$$\begin{aligned} \hat{\pi}_{w,t} = {} & \beta\pi^{\gamma_w+1} E_t \hat{\pi}_{w,t+1} + \kappa_w \left(2\hat{N}_t - \hat{w}_t\right) - (\tilde{\kappa}_{wd} - \kappa_{w\epsilon}) \left(\hat{N}_t - \hat{Y}_t\right) + \kappa_{wd} \hat{d}_{w,t} + \hat{d}_{w,t-1} \\ & + \beta\pi^{\gamma_w+1} E_t \hat{d}_{w,t+1} + \zeta_{w,t} + \varphi_{w,t} + \psi_{w,t}, \end{aligned} \tag{34}$$

where $\zeta_{w,t}$, $\varphi_{w,t}$, and $\psi_{w,t}$ are auxiliary variables that are additional drivers of wage inflation under nonzero trend inflation and satisfy

$$\begin{aligned}\zeta_{w,t} &= \alpha_w \beta \pi^{\gamma_w} E_t \zeta_{w,t+1} \\ &\quad + \kappa_{w\epsilon\zeta} \Big[(1 - \alpha_w \beta \pi^{\gamma_w}) \Big(2E_t \hat{N}_{t+1} - E_t \hat{w}_{t+1} + \gamma_w E_t \hat{d}_{w,t+1} \Big) + E_t \hat{w}_{t+1} - \hat{w}_t + \gamma_w E_t \hat{\pi}_{w,t+1} \Big],\end{aligned} \tag{35}$$

$$\begin{aligned}\varphi_{w,t} &= \alpha_w \beta \pi^{\gamma_w - 1} E_t \varphi_{w,t+1} \\ &\quad + \kappa_{w\varphi} \Big[(1 - \alpha_w \beta \pi^{\gamma_w - 1}) \Big(E_t \hat{N}_{t+1} - E_t \hat{Y}_{t+1} + \gamma_w E_t \hat{d}_{w,t+1} \Big) + \gamma_w E_t \hat{\pi}_{w,t+1} - E_t \hat{\pi}_{t+1} \Big],\end{aligned} \tag{36}$$

$$\psi_{w,t} = \alpha_w \beta \pi^{-1} E_t \psi_{w,t+1} + \kappa_{w\epsilon\psi} \Big[(1 - \alpha_w \beta \pi^{-1}) \Big(E_t \hat{N}_{t+1} - E_t \hat{Y}_{t+1} \Big) - E_t \hat{\pi}_{t+1} \Big]. \tag{37}$$

The composite coefficients κ_w, $\tilde{\kappa}_{wd}$, $\kappa_{w\epsilon}$, κ_{wd}, $\kappa_{w\epsilon\zeta}$, $\kappa_{w\varphi}$, and $\kappa_{w\epsilon\psi}$ in (34)–(37) are presented in Appendix B. Note that $\kappa_{w\varphi} = 0$ if $\pi = 1$ and that $\kappa_{w\epsilon\zeta} = \kappa_{w\epsilon\psi} = 0$ if $\pi = 1$ or $\epsilon_w = 0$. The law of motion of the real marginal cost of labor aggregation is given by

$$\hat{d}_{w,t} = \rho_{wd} \hat{d}_{w,t-1} + \kappa_{w\epsilon d}\, \hat{\pi}_{w,t}, \tag{38}$$

where

$$\rho_{wd} \equiv \frac{\alpha_w \pi^{-1} (1 + \epsilon_{w1} \pi^{\gamma_w})}{1 + \epsilon_{w1}}, \quad \kappa_{w\epsilon d} \equiv -\frac{\epsilon_{w1} \alpha_w \pi^{-1} (\pi^{\gamma_w} - 1)}{(1 + \epsilon_{w1})(1 - \alpha_w \pi^{-1})},$$

and $\epsilon_{w1} \equiv \epsilon_w [(1 - \alpha_w)/(1 - \alpha_w \pi^{\gamma_w - 1})]^{\gamma_w/(\gamma_w - 1)}$, so $\kappa_{w\epsilon d} = 0$ if $\pi = 1$ or $\epsilon_w = 0$.[14] Analogous to that of goods aggregation, eq. (38) implies that the real marginal cost of labor aggregation depends on current and past rates of wage inflation: $\hat{d}_{w,t} = \kappa_{w\epsilon d} \sum_{j=0}^{\infty} \rho_{wd}^j \hat{\pi}_{w,t-j}$. Combining this and the wage-GNKPC (34) leads to

$$\begin{aligned}\hat{\pi}_{w,t} &= b_{w\epsilon 1} \sum_{j=1}^{\infty} \rho_{wd}^{j-1} \hat{\pi}_{w,t-j} + b_{w2} E_t \hat{\pi}_{w,t+1} \\ &\quad + b_{w3} \Big[\kappa_w \Big(2\hat{N}_t - \hat{w}_t \Big) - (\tilde{\kappa}_{wd} - \kappa_{w\epsilon}) \Big(\hat{N}_t - \hat{Y}_t \Big) + \zeta_{w,t} + \varphi_{w,t} + \psi_{w,t} \Big],\end{aligned} \tag{39}$$

where $b_{w\epsilon 1} \equiv \kappa_{w\epsilon d} b_{w3} [1 + \rho_{wd} (\kappa_{wd} + \beta \pi^{\gamma_w + 1} \rho_{wd})]$, $b_{w2} \equiv \beta \pi^{\gamma_w + 1} b_{w3} (1 + \kappa_{w\epsilon d})$, and $b_{w3} \equiv$

[14] Under the calibration of model parameters reported in Table 1 below, the values of κ_w, $\tilde{\kappa}_{wd}$, $\kappa_{w\epsilon}$, $\kappa_{w\epsilon\zeta}$, $\kappa_{w\epsilon\psi}$, ρ_{wd}, and $\kappa_{w\epsilon d}$ are positive, while those of κ_{wd} and $\kappa_{w\varphi}$ are negative.

$1/[1-\kappa_{w\epsilon d}(\kappa_{wd}+\beta\pi^{\gamma_w+1}\rho_{wd})]$. Therefore, the model also provides a theoretical justification for intrinsic inertia in wage inflation that does not rely on backward-looking wage-setting behavior. Moreover, the degree of intrinsic inertia in wage inflation can be summarized as the sum of the coefficients on past rates of wage inflation, $\lambda_{w\epsilon} \equiv b_{w\epsilon 1}/(1-\rho_{wd})$, and depends on the model's structural parameters, including trend inflation π and the superelasticity parameter ϵ_w. Analogous to the GNKPC, if $\pi = 1$ or $\epsilon_w = 0$ then $b_{w\epsilon 1} = 0$ because $\kappa_{w\epsilon d} = 0$, and thus there is no intrinsic inertia of wage inflation in (39).

The complete set of log-linearized equilibrium conditions consists of (29)–(32), (34)–(38), and

$$\hat{Y}_t = E_t\hat{Y}_{t+1} - (\hat{\imath}_t - E_t\hat{\pi}_{t+1}), \tag{40}$$

$$\hat{\imath}_t = \rho\,\hat{\imath}_{t-1} + (1-\rho)\Big(\phi_\pi\hat{\pi}_t + \phi_Y\hat{Y}_t\Big) + \varepsilon_{i,t}, \tag{41}$$

$$\hat{mc}_t = \hat{w}_t, \tag{42}$$

$$\hat{\pi}_{w,t} = \hat{w}_t - \hat{w}_{t-1} + \hat{\pi}_t, \tag{43}$$

$$\hat{Y}_t = \hat{N}_t - \hat{\Delta}_t, \tag{44}$$

$$\hat{\Delta}_t = \alpha_p\pi^{\gamma_p}\hat{\Delta}_{t-1} + \frac{s}{s+\epsilon_p}\frac{\gamma_p\alpha_p\pi^{\gamma_p-1}(\pi-1)}{1-\alpha_p\pi^{\gamma_p-1}}\Big(\hat{\pi}_t + \hat{d}_{p,t} - \hat{d}_{p,t-1}\Big). \tag{45}$$

Eq. (40) is the spending Euler equation, (41) is the Taylor-type monetary policy rule, (42) is the equation for the real marginal cost of goods production, (43) is the definition of wage inflation, (44) is the aggregate production equation, and (45) is the law of motion of the relative price distortion $\hat{\Delta}_t$, where s is the steady-state value of s_t that is given by $s = (1-\alpha_p)/(1-\alpha_p\pi^{\gamma_p})[(1-\alpha_p\pi^{\gamma_p-1})/(1-\alpha_p)]^{\gamma_p/(\gamma_p-1)}$.

2.6 New Keynesian model with indexation

To show the implications of our model for inflation persistence, the model is compared with its NK counterpart with price and wage indexation. The counterpart can be obtained by assuming that prices and nominal wages that are kept unchanged in the above setting are instead updated by indexing to a weighted average of trend and recent past inflation: $P_t(f) = \pi^{1-\iota_p}\pi_{t-1}^{\iota_p}P_{t-1}(f)$ and $W_t(h) = \pi^{1-\iota_w}\pi_{t-1}^{\iota_w}W_{t-1}(h)$, where $0 \leq \iota_p, \iota_w \leq 1$. These assumptions generate the NK Phillips curve (NKPC) and wage-NKPC

$$\hat{\pi}_t = \frac{\iota_p}{1+\beta\iota_p}\hat{\pi}_{t-1} + \frac{\beta}{1+\beta\iota_p}E_t\hat{\pi}_{t+1} + \frac{(1-\alpha_p)(1-\alpha_p\beta)}{\alpha_p(1+\beta\iota_p)[1-\epsilon_p\theta_p/(\theta_p-1)]}\hat{mc}_t, \tag{46}$$

$$\begin{aligned}\hat{\pi}_{w,t} = {} & \frac{\iota_w}{1+\beta\iota_w}\hat{\pi}_{w,t-1} + \frac{\beta}{1+\beta\iota_w}E_t\hat{\pi}_{w,t+1} \\ & + \frac{(1-\alpha_w)(1-\alpha_w\beta)}{\alpha_w(1+\beta\iota_w)[1+\theta_w\sigma_n-\epsilon_w\theta_w/(\theta_w-1)]}\Big((1+\sigma_n)\hat{Y}_t - \hat{w}_t\Big),\end{aligned} \tag{47}$$

and imply that $\hat{\Delta}_t = \hat{d}_{p,t} = \varphi_{p,t} = \psi_{p,t} = \hat{d}_{w,t} = \zeta_{w,t} = \varphi_{w,t} = \psi_{w,t} = 0$. Thus, the NK counterpart consists of (40)–(43) and (46)–(47). In the special case of full indexation to trend inflation (i.e., $\iota_p = \iota_w = 0$), this model coincides with our model at zero trend inflation, i.e., $\pi = 1$. Thus, we can demonstrate the effect of trend inflation on inflation persistence by comparing our model with nonzero trend inflation and the NK counterpart with full indexation to trend inflation. Moreover, we can compare our model that provides a microfoundation of intrinsic persistence in inflation, with the NK counterpart that assumes intrinsic persistence stemming from indexation to recent past inflation, i.e., $\iota_p, \iota_w > 0$.

3 Impulse Response Analysis

This section analyzes impulse responses to monetary policy shocks in the log-linearized model presented in the preceding section and shows that a plausibly calibrated version of the model can account for the well-known persistent, hump-shaped response of inflation to monetary policy shocks.

3.1 Calibration of model parameters

The calibration of parameters in the quarterly model is summarized in Table 1. The elasticity of labor supply has already been fixed at $1/\sigma_n = 1$. As is common in the literature, we set the subjective discount factor at $\beta = 0.99$; the probability of no price change at $\alpha_p = 0.75$, which implies that the average frequency of price change is four quarters; and the parameter governing the elasticity of substitution between individual goods at $\theta_p = 10$, which implies a desired price markup of 11 percent. The corresponding parameters for wage setting and labor are chosen at $\alpha_w = 0.75$, in line with the micro evidence by Barattieri et al. (2014), and $\theta_w = 10$.

Table 1: Calibration of parameters in the quarterly model.

σ_n	Inverse elasticity of labor supply	1
β	Subjective discount factor	0.99
α_p	Probability of no price change	0.75
α_w	Probability of no wage change	0.75
θ_p	Parameter governing the elasticity of substitution between goods	10
θ_w	Parameter governing the elasticity of substitution between labor	10
ϵ_p	Parameter governing the superelasticity of demand for goods	-3
ϵ_w	Parameter governing the superelasticity of demand for labor	-3
π	Gross trend inflation rate	$1.032^{1/4}$
ρ	Degree of monetary policy inertia	0.8
ϕ_π	Degree of monetary policy response to inflation	1.5
ϕ_Y	Degree of monetary policy response to output	0.5/4
σ_i	Standard deviation of monetary policy shock	0.0025

To calibrate the parameter governing the superelasticity of demand for goods, we draw on the estimation results of Guerrieri et al. (2010). They obtain a benchmark estimate of -3.0 and an alternative estimate of -6.1.[15] Accordingly, we set $\epsilon_p = -3$ in the baseline calibration, which implies a steady-state superelasticity of demand for goods of $-\epsilon_p\theta_p = 30$, and also present results for an alternative value of $\epsilon_p = -6$. We are not aware of empirical evidence for the parameter governing the superelasticity of demand for labor, so we choose the parameter based on the evidence for goods, $\epsilon_w = -3$, which implies a steady-state superelasticity of demand for labor of $-\epsilon_w\theta_w = 30$, and also show results for an alternative value of $\epsilon_w = -6$.

Regarding monetary policy, the trend inflation rate (or the monetary authority's inflation target) is chosen at $\pi = 1.032^{1/4}$ (i.e., 3.2 percent annually), which is the average inflation rate of the U.S. GDP deflator over the period 1959:Q1–2019:Q4.[16] The degree of policy

[15] Specifically, Guerrieri et al. (2010) obtain a benchmark estimate for the parameter $\Psi = -\epsilon_p\theta_p/[\theta_p(1-\epsilon_p)-1]$ of 0.78 under the assumption of $\theta_p = 6$ and an alternative estimate of $\Psi = 0.87$ under the assumption of $\theta_p = 11$. While our baseline calibration of $\theta_p = 10$ lies in between their two values, our baseline calibration of ϵ_p is based on their benchmark estimate, which is the more conservative one. Micro evidence from European household spending data on relatively narrow categories of retail goods also points to the superelasticity of demand, although its magnitude is smaller (see Dossche et al., 2010; Beck and Lein, 2020).

[16] To meet assumption (28) under the calibration of model parameters reported in Table 1, the trend inflation rate needs to be greater than -2.8 percent annually.

inertia, the policy responses to inflation and output, and the standard deviation of the policy shock are set at $\rho = 0.8$, $\phi_\pi = 1.5$, $\phi_Y = 0.5/4$, and $\sigma_i = 0.0025$, respectively.

3.2 Impulse responses to monetary policy shocks

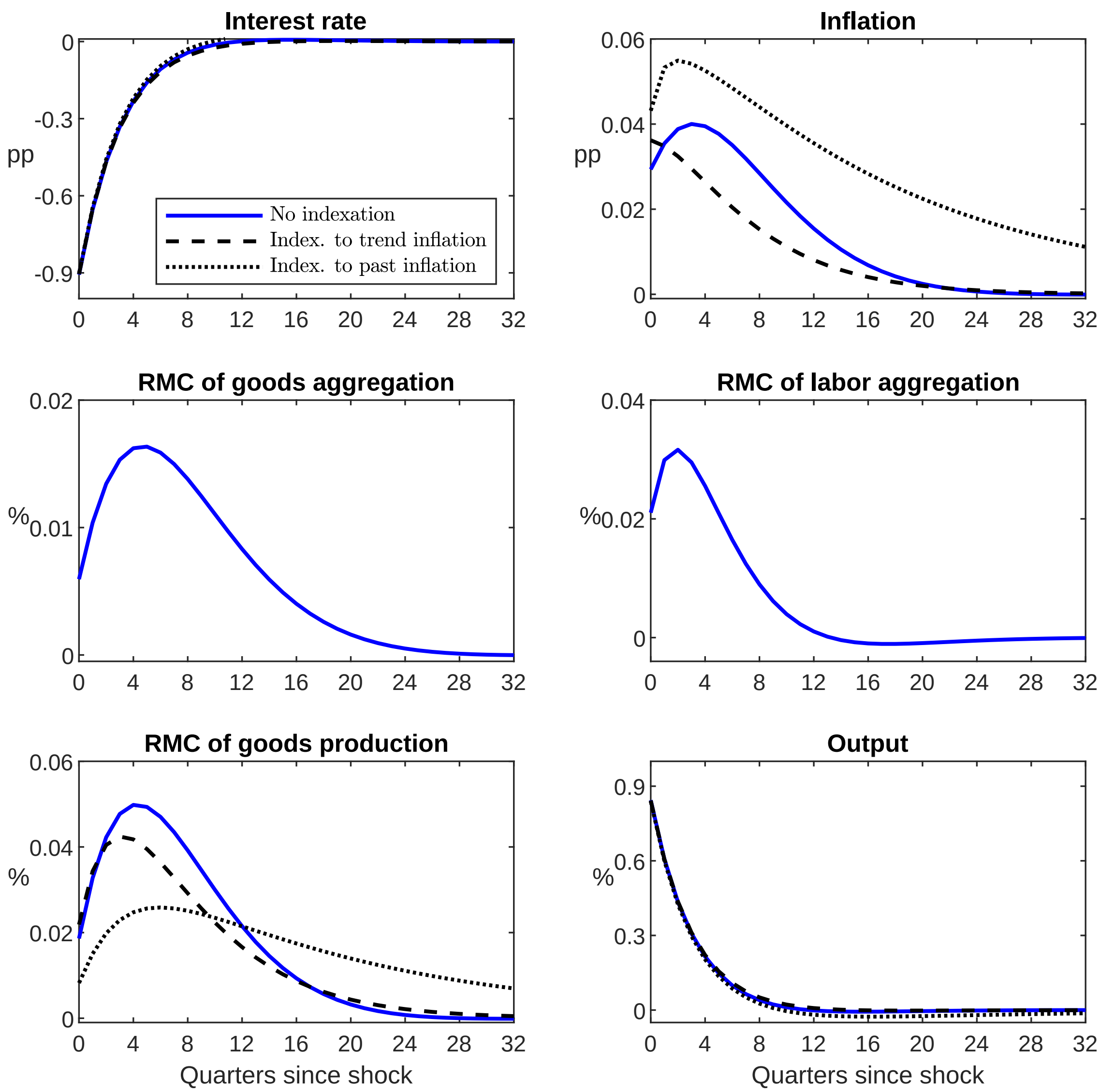
Empirical evidence indicates that the response of inflation to monetary policy shocks builds for some time before gradually diminishing. This subsection shows that our model can account for the hump-shaped response, using the calibration of parameters reported in Table 1.

Figure 2 illustrates the effects of an expansionary monetary policy shock on inflation in our model (the solid lines) and in its NK counterpart with full indexation to trend inflation (the dashed lines) or with partial indexation to past inflation (the dotted lines). For the NK counterpart with partial indexation to past inflation, the posterior mean estimates of Smets and Wouters (2007) are used for the parameter values of $\iota_p = 0.24$ and $\iota_w = 0.58$. The policy shock leads to an immediate drop in the interest rate, which then returns gradually to its pre-shock level in the upper left panel of the figure. The upper right panel shows that inflation exhibits a persistent response to the policy shock in our model, with a hump shape and a gradual decline, in line with the empirical evidence. Inflation rises for three quarters following the shock to reach a peak level and then declines gradually, similar to the response of inflation obtained in the NK counterpart with partial indexation to past inflation, which rises for two quarters after the shock. The crucial role of nonzero trend inflation for inflation persistence in our model is evident by comparing the responses of inflation in our model and in the NK counterpart with full indexation to trend inflation. Because this counterpart has the same form of log-linearized equilibrium conditions as our model with zero trend inflation as noted above, the difference between the solid and the dashed lines shows the effect of trend inflation on the inflation response.[17] Absent this effect, the response of inflation counterfactually peaks upon impact of the shock. Thus, the Kimball-type non-CES aggregators alone generate no hump-shaped response of inflation in our model without nonzero trend inflation.

The difference between the cases of positive trend inflation (the solid lines) and zero trend inflation (the dashed lines) is caused mainly by the presence of the real marginal

[17] Of course, the level of trend inflation is different between the two models, but is not relevant for the comparison of the impulse responses displayed in terms of log-deviations from steady-state values.

Figure 2: Impulse responses to an expansionary monetary policy shock.



Notes: The figure presents the impulse responses to a one standard deviation expansionary monetary policy shock under the calibration of model parameters reported in Table 1. The interest and inflation rates are displayed in annualized terms. The solid lines represent our model (with Kimball-type non-CES aggregators of goods and labor). The dashed and the dotted lines respectively show the NK counterparts with full indexation to trend inflation ($\iota_p = \iota_w = 0$) and with partial indexation to past inflation ($\iota_p = 0.24$, $\iota_w = 0.58$), where the real marginal costs (RMCs) of goods and labor aggregation exhibit no (first-order) responses.

costs (RMCs) of goods and labor aggregation, $\hat{d}_{p,t}$ and $\hat{d}_{w,t}$, as can be seen in the difference between the log-linearized equilibrium conditions (29)–(32) and (34)–(38) in the former case and (46)–(47) with $\iota_p = \iota_w = 0$ in the latter. As displayed in the middle panels of Figure 2, the RMCs of goods and labor aggregation exhibit persistent, hump-shaped responses to the shock, reflecting that they depend on current and past rates of price and wage inflation. The RMC of goods aggregation has a significant influence on inflation dynamics mainly through the GNKPC (29), where the past, present, and expected future values of the RMC drive inflation and the RMC in turn depends on current and past inflation rates, thus making inflation depend on its lags in the GNKPC (33).[18] Regarding the RMC of labor aggregation, it can affect inflation dynamics indirectly through its effects on the RMC of goods production, which coincides with the real wage, as shown in (7). However, the lower left panel indicates that the RMCs of goods and labor aggregation have modest effects on the RMC of goods production, as the latter RMC displays a similar response to that in the NK counterpart with full indexation to trend inflation.

In contrast to the key role of the RMC of goods aggregation, the relative price distortion $\hat{\Delta}_t$ makes little contribution to the response of inflation to the policy shock in our model.[19] The relative price distortion could affect the RMC of goods production and hence inflation dynamics through its effect on output, as the aggregate production equation (44) relates output to the relative price distortion. However, the lower right panel of Figure 2 shows that the response of output in our model is similar to those in the NK counterparts with full indexation to trend inflation and with partial indexation to past inflation, where output $\hat{Y}_t$

[18]The joint effect of the past, present, and expected future values of the RMC of goods aggregation on inflation can give a sense of the direct effect of the RMC on inflation (ignoring indirect effects through the RMC of goods production, inflation expectations, and the auxiliary variables in the GNKPC). The term $x_t \equiv \kappa_{pd}\hat{d}_{p,t} + \hat{d}_{p,t-1} + \beta\pi E_t\hat{d}_{p,t+1}$ in the GNKPC (29) declines upon impact of the expansionary monetary policy shock before gradually returning to the pre-shock level. The initial decline of x_t mutes the response of inflation following the shock, thus generating a hump shape. Under the calibration of model parameters reported in Table 1, we have $\kappa_{pd} = -2.032 \approx -2$ and $\beta\pi = 0.998 \approx 1$, so x_t is approximately equal to the second difference of $E_t\hat{d}_{p,t+1}$, i.e., $x_t \approx (E_t\hat{d}_{p,t+1} - \hat{d}_{p,t}) - (\hat{d}_{p,t} - \hat{d}_{p,t-1})$. This suggests that a larger (more concave) response of $\hat{d}_{p,t}$ would make the response of x_t more negative and the response of inflation more hump-shaped.

[19]This result contrasts with that of Damjanovic and Nolan (2010). They point out that a longer average duration of price change (e.g., two years) makes relative price distortion more amplified and more persistent under nonzero trend inflation, thus generating a more persistent response of inflation to monetary policy shocks. At the same time, however, they show that the distortion also generates a counterfactual decline in output after an expansionary monetary policy shock.

($= \hat{N}_t$) is not affected, up to the first order, by the relative price distortion.[20] This indicates that the relative price distortion plays little role for inflation dynamics in our model.

3.3 Roles of Kimball-type aggregators and nominal rigidities

We have indicated that the degree of intrinsic inflation inertia $\lambda_{p\epsilon}$ depends on structural parameters in our model. This subsection then shows that in our model, the Kimball-type non-CES aggregator of goods with $\epsilon_p < 0$ and the rigidities of prices and nominal wages (α_p, α_w) play key roles for inflation persistence, while the Kimball-type non-CES aggregator of labor with $\epsilon_w < 0$ mainly affects inflation dynamics by mitigating indeterminacy of equilibrium induced by higher trend inflation, which is likely in the case of the CES aggregator of labor, i.e., $\epsilon_w = 0$.[21]

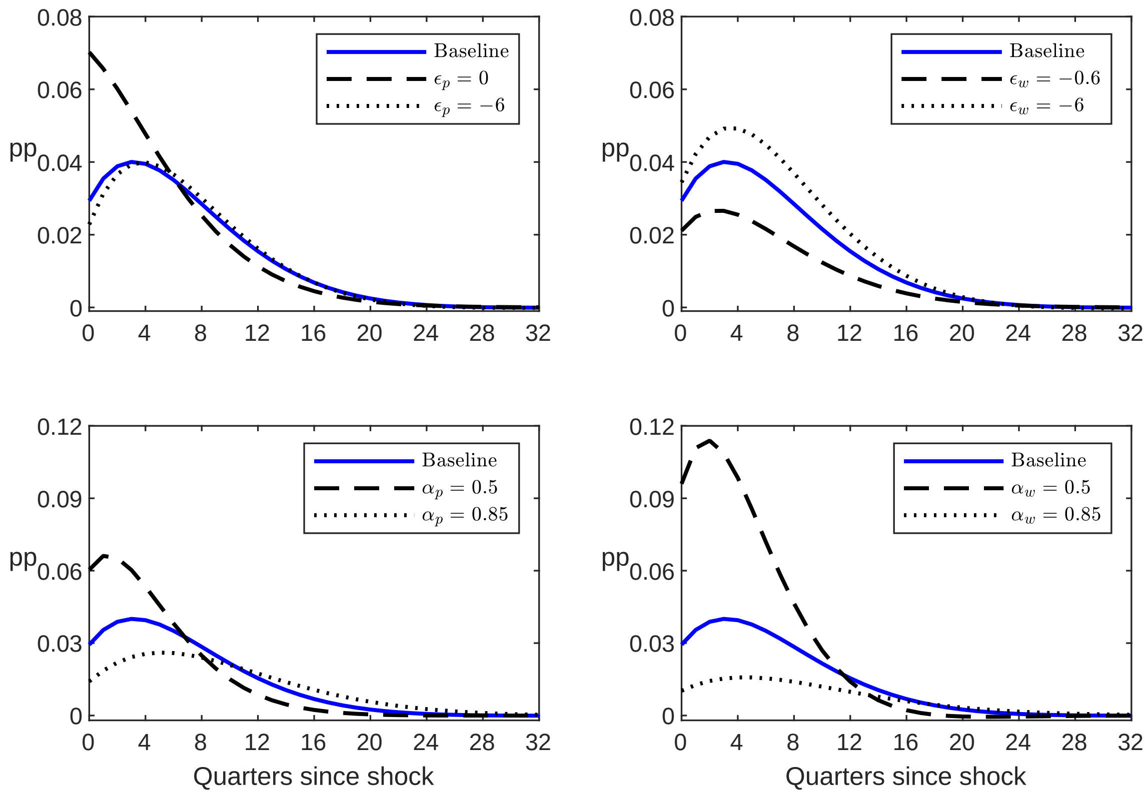
Figure 3 presents impulse responses of inflation to an expansionary monetary policy shock for alternative values of four parameters: the parameters that govern the superelasticities of demand for goods and labor (ϵ_p, ϵ_w) and the Calvo probabilities for staggered price- and wage-setting (α_p, α_w). The upper left panel of the figure shows the crucial role of the Kimball-type non-CES aggregator of goods by comparing the response of inflation under the baseline calibration reported in Table 1 (i.e., $\epsilon_p = -3$) with that in the case of the CES aggregator of goods (i.e., $\epsilon_p = 0$). The latter case implies that $\hat{d}_{p,t} = 0$ (and $\psi_{p,t} = 0$). With the CES aggregator, inflation peaks upon impact of the shock. Thus, nonzero trend inflation alone generates no hump-shaped response of inflation in the calibrated model without the Kimball-type non-CES aggregator of goods. The case of $\epsilon_p = -6$, which doubles the superelasticity of demand for goods compared with the baseline, further accentuates the hump shape of the inflation response, indicating that the higher superelasticity dampens the response of inflation early following the shock.

The upper right panel of Figure 3 illustrates a supporting role of the Kimball-type non-

[20]Empirical evidence points to a hump-shaped response of output to monetary policy shocks. Yet this is absent in our model. Adding habit formation in consumption preferences to the model generates a hump-shaped response of output and provides an additional source of inflation persistence. Such an extension of the model is examined in Section 3.4.

[21]In this context, higher trend inflation increases the likelihood of indeterminacy of equilibrium with the CES aggregator of goods, as shown by Ascari and Ropele (2009) and Coibion and Gorodnichenko (2011). Kurozumi and Van Zandweghe (2016) show that the Kimball-type non-CES aggregator of goods prevents the indeterminacy caused by higher trend inflation.

Figure 3: Roles of Kimball-type aggregators and nominal rigidities in the impulse response of inflation to monetary policy shocks.



Notes: The figure presents the responses of inflation to a one standard deviation expansionary monetary policy shock under the calibration of model parameters reported in Table 1, except as indicated in each panel. The solid lines represent the baseline case, while the dashed and the dotted lines represent the cases with alternative parameter values.

CES aggregator of labor. The panel considers the two alternative parameter values of $\epsilon_w = -0.6$ and $\epsilon_w = -6$, omitting the case of the CES aggregator of labor (i.e., $\epsilon_w = 0$), which induces indeterminacy of equilibrium. A higher superelasticity of demand for labor generates a larger response of inflation. That is because, under positive trend inflation, a higher superelasticity increases the slope of the wage-GNKPC, κ_w, which more than offsets the dampening effect of the real marginal cost of labor aggregation in the wage-GNKPC. With this dual implication of ϵ_w for the wage-GNKPC, the overall effect on inflation persistence is difficult to discern from the panel, so a quantitative measure of persistence is useful. A summary statistic of the persistence in impulse responses to a shock is the half-life, defined as the number of quarters until the size of the response falls to half of its size upon impact of the shock. The half-life of the inflation response is 12 quarters under the baseline calibration, 10 quarters when $\epsilon_w = -0.6$, and 12 quarters when $\epsilon_w = -6$. Thus, the lower superelasticity of demand for labor decreases inflation persistence somewhat.

The lower panels of Figure 3 show the importance of nominal rigidities for inflation persistence by comparing the response of inflation under the baseline calibration with those obtained under two alternative values of the probability of no price change (the lower left panel) and the probability of no wage change (the lower right panel). Higher price rigidity increases the persistence in the inflation response through higher intrinsic inflation inertia. Higher nominal wage rigidity also increases it through higher intrinsic inertia of wage inflation, which gives rise to a more persistent response of the real wage, that is, the real marginal cost of goods production in the model.

3.4 Model extension for empirical validation

Empirical evidence from VARs indicates that the peak response of inflation to a monetary policy shock is even more delayed than that in our calibrated model. Christiano et al. (2011) conduct a state-of-the-art VAR analysis and find that a monetary policy shock induces a peak response of inflation after eight quarters, longer than the three quarters in our model displayed in Figure 2. To reduce the discrepancy between the impulse responses of inflation in the model and the VAR evidence, we augment the model with three features: working capital, habit formation in consumption preferences, and the monetary policy response to output growth (see Appendix C for the details of the extended model). Christiano et al. (2011) and Phaneuf et al. (2018) show that working capital dampens the initial response of inflation to monetary policy shocks. Fuhrer (2000) demonstrates that habit formation gives rise to a gradual decline in inflation after an expansionary monetary policy shock. Coibion and Gorodnichenko (2011) estimate Taylor-type monetary policy rules on real-time data and find evidence of a sizeable policy response to output growth since the early 1980s. Such a policy response can generate a more gradual decline in inflation as weak output growth following the initial rise in output calls for a more accommodative monetary policy. We also introduce a (nonstationary) technology shock in the model and evaluate the impulse response of inflation to the shock.

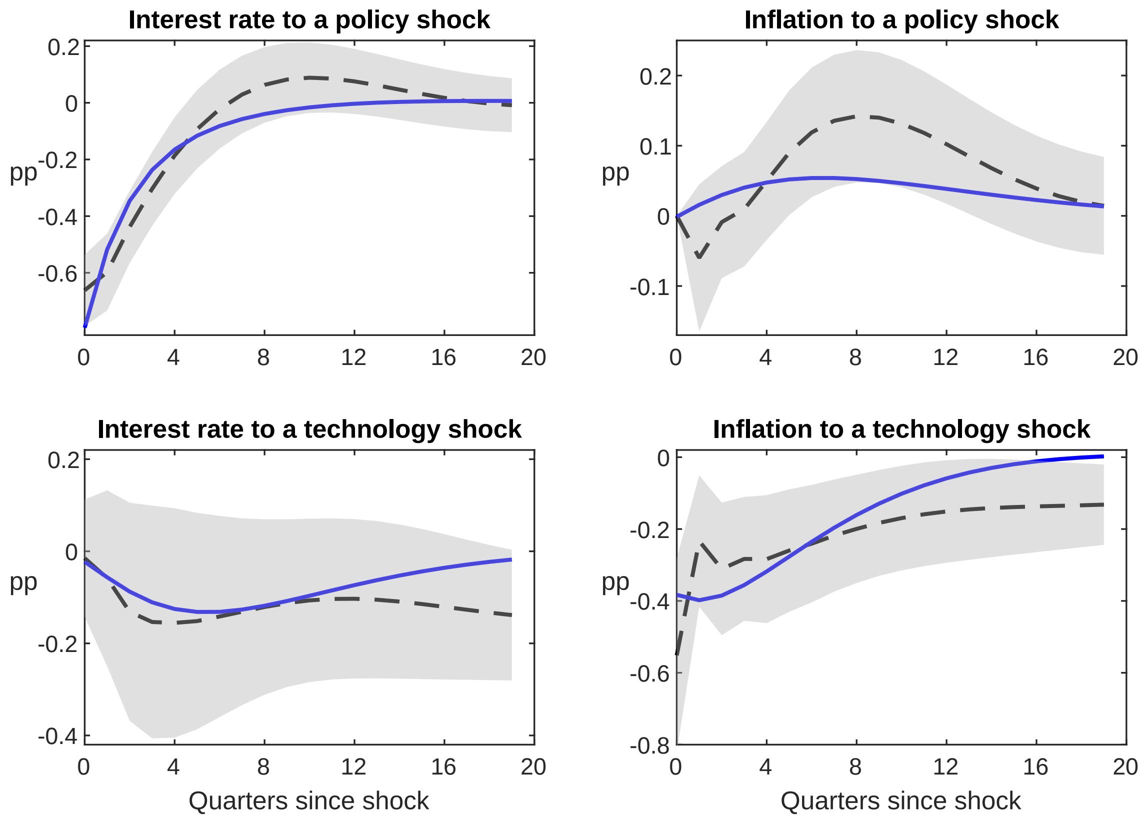
The extended model can match the empirical impulse responses of inflation reasonably well. The upper right panel of Figure 4 displays the response of inflation to an expansionary monetary policy shock in the model (the solid line) under the calibration of the degree of habit persistence $b = 0.7$, the monetary policy response to output growth $\phi_g = 1$, and other parameters reported in Table 1 except for the trend inflation rate $\pi = 1.036^{1/4}$ (i.e., 3.6 percent annually), which is the average inflation rate of the U.S. GDP deflator over the sample period 1951:Q1–2008:Q4 of Christiano et al. (2011). Following the shock, inflation rises gradually for seven quarters to a peak before declining gradually. To evaluate the empirical relevance of the extended model, the panel also reproduces the mean impulse response of inflation to a monetary policy shock (the dashed line) and its 95 percent probability interval (the gray band) in the VAR of Christiano et al. (2011). Although the empirical impulse response of inflation rises more steeply and continues to rise one more quarter than our model counterpart, the latter lies within the probability interval, thus providing an empirical validation of the extended model.[22]

The lower right panel of Figure 4 displays the impulse response of inflation to a one standard deviation positive technology shock. The standard deviation of the shock is set at $\sigma_a = 0.01$ in the extended model. Inflation declines on impact, in contrast with the delayed response to the monetary policy shock. The positive nonstationary technology shock induces a persistent decline in the real wage. This persistent decline lowers not only the real marginal cost of goods production but also expected future inflation through lower expected future real marginal costs. The joint decline in the real marginal cost and expected future inflation generates a substantial decline in inflation.[23] The impulse responses to the technology shock in the model almost lie within the 95 percent probability intervals of their empirical couterparts, thus providing an empirical validation of the extended model.

[22]In the case of the CES aggregator of goods (i.e., $\epsilon_p = 0$), the extended model shows that the response of inflation peaks three quarters after a policy shock.

[23]In the case of the CES aggregator of goods (i.e., $\epsilon_p = 0$), the technology shock has an even larger impact on inflation.

Figure 4: Impulse responses in the extended model and their VAR counterparts.



Sources: Christiano et al. (2015) and authors' calculations.

Notes: The upper and lower panels of the figure display impulse responses to a one standard deviation expansionary monetary policy shock and positive technology shock, respectively. The solid lines represent the impulse responses in the extended model under the calibration of the degree of habit persistence $b = 0.7$, the monetary policy response to output growth $\phi_g = 1$, and other parameters reported in Table 1 except for the trend inflation rate $\pi = 1.036^{1/4}$ (i.e., 3.6 percent annually). The dashed lines and the gray bands reproduce, respectively, the mean impulse responses and their 95 percent probability intervals in the VAR of Christiano et al. (2011), which are also reproduced in Christiano et al. (2015).

4 Effect of Trend Inflation on Inflation Persistence

Our model provides a microfoundation of intrinsic inertia in price and wage inflation by relating the degrees of intrinsic inertia to structural parameters of the model. A parameter of particular relevance for monetary policy is the rate of trend inflation, as it also represents the inflation target of the monetary authority in the model.[24] This section examines the effect of a decline in trend inflation on the degree of intrinsic inertia in inflation and more general measures of inflation persistence in the baseline model presented in Section 2.

A number of empirical studies indicate that inflation persistence has decreased in the U.S. since the early 1980s. Cogley and Sargent (2002) employ spectral analysis to estimate inflation persistence and find that the persistence displays a similar pattern to the level of inflation: both the level and the persistence of inflation increased in the 1970s and decreased gradually from the early 1980s onward. Cogley et al. (2010) use predictability as a measure of persistence, as shocks that are more persistent make time series more predictable. They show that the persistence of the inflation gap (i.e., the gap between actual and trend inflation) rose in the 1970s and fell during and after the Volcker disinflation in the early 1980s. Stock and Watson (2007) characterize inflation as consisting of a transitory and a permanent component and show empirically that the variance of the permanent component increased in the 1970s before declining in the mid 1980s. Fuhrer (2011) examines the persistence in various measures of inflation using different methods and finds that inflation persistence has decreased for headline inflation but less so for core inflation (which excludes food and energy prices). Consistent with those studies, estimated DSGE models indicate a decline in the degrees of intrinsic inertia in price and wage inflation from the period including the 1970s to the period since the mid 1980s; for instance, see the subsample estimates of the price and wage indexation parameters (ι_p, ι_w) by Smets and Wouters (2007) and the corresponding estimates by Hofmann et al. (2012).

Our model provides a new perspective on the measured decrease in inflation persistence from a high level in the 1970s to a lower level beginning in the 1980s, around the time of

[24] Benati (2008) conducts an empirical analysis of inflation persistence across countries and time periods and finds that the degree of inflation persistence varies depending on monetary policy regimes. He therefore argues against the assumption that intrinsic inflation persistence is policy invariant, as is embedded in the NKPC (46) and in many existing DSGE models. Based on an empirical analysis of wage dynamics, Hofmann et al. (2012) make a similar argument concerning intrinsic persistence in wage inflation.

the Volcker disinflation. Most previous studies, such as Benati and Surico (2008), Carlstrom et al. (2009), and Davig and Doh (2014), attribute the decrease in inflation persistence to a more active monetary policy response to inflation, sometimes in combination with declines in the volatility of shocks to the U.S. economy.[25]

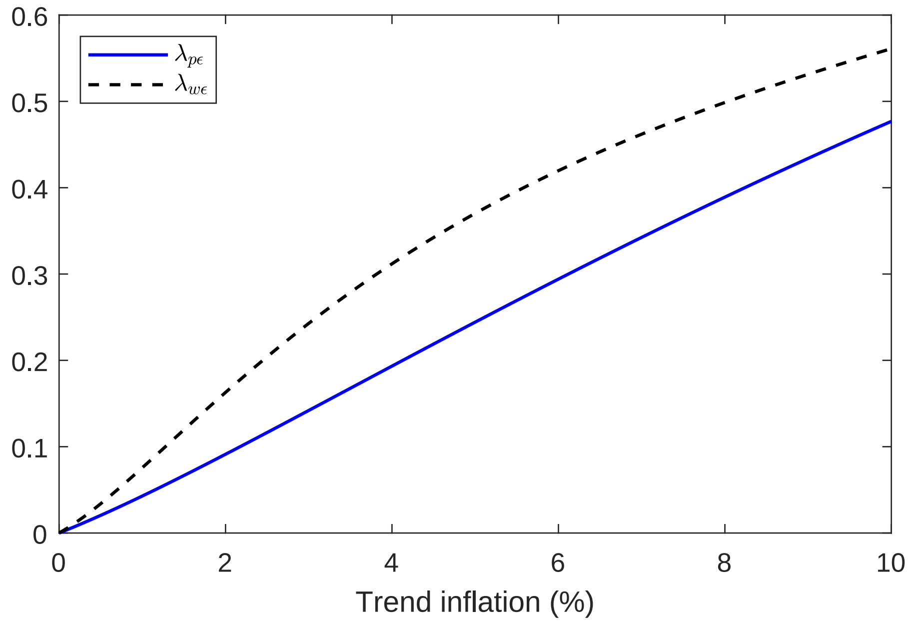
4.1 Degree of intrinsic inertia in inflation

In our model the degrees of intrinsic inertia in price and wage inflation $(\lambda_{p\epsilon}, \lambda_{w\epsilon})$ give a sense of the effect of trend inflation on inflation persistence. Recall from Section 2.5 that $\lambda_{p\epsilon}$ and $\lambda_{w\epsilon}$ are defined as the sum of the coefficients on lagged rates of price inflation in the GNKPC (33) and on those of wage inflation in the wage-GNKPC (39), respectively. Figure 5 plots $\lambda_{p\epsilon}$ and $\lambda_{w\epsilon}$ for values of the annualized trend inflation rate $\bar{\pi}$ ranging from zero to 10 percent, using the calibration of other model parameters reported in Table 1. For instance, at the baseline value for the trend inflation rate of 3.2 percent annually, we have $\lambda_{p\epsilon} = 0.15$ and $\lambda_{w\epsilon} = 0.26$. To compare the degrees of intrinsic inertia in price and wage inflation in our model with the estimates of Smets and Wouters (2007), we can consider the values in the figure at the annualized trend inflation rate of 4.1 percent, which is the average inflation rate of the U.S. GDP deflator over the sample period 1966:Q1–2004:Q4 of Smets and Wouters (2007). Those values are $\lambda_{p\epsilon} = 0.20$ and $\lambda_{w\epsilon} = 0.32$. They are close to the degrees of intrinsic inertia in the NK model with indexation, $\iota_p/(1 + \beta\iota_p) = 0.19$ in the NKPC (46) and $\iota_w/(1 + \beta\iota_w) = 0.37$ in the wage-NKPC (47), based on the estimates of Smets and Wouters (2007).

In our calibrated model, lower trend inflation reduces the degree of intrinsic inertia in price and wage inflation, as illustrated in Figure 5. For example, intrinsic inertia falls from $(\lambda_{p\epsilon}, \lambda_{w\epsilon}) = (0.33, 0.45)$ to $(0.09, 0.16)$ for a decline in the trend inflation rate $\bar{\pi}$ from 6.7 percent to 2 percent annually. The former value is the average inflation rate of the U.S. GDP deflator during the 1970s, while the latter is the average for the three decades from 1990:Q1 to 2019:Q4 and coincides with the Federal Reserve's target for the inflation rate of the personal

[25] Cogley et al. (2010) attribute the decrease in inflation-gap persistence primarily to a decline in the volatility of shocks to the inflation target (or trend inflation), with a secondary role for the monetary policy response to inflation. A shock to the inflation target in their estimated model is reminiscent of the credible disinflation examined in Section 5, although in their model a decline in the inflation target leads inflation to undershoot the new inflation target initially.

Figure 5: Degree of intrinsic inertia in price and wage inflation.



Notes: The figure presents the degree of intrinsic inertia in price inflation $\lambda_{p\epsilon}$ and that of intrinsic inertia in wage inflation $\lambda_{w\epsilon}$ in the model for a range of values of the annualized trend inflation rate under the calibration of parameters reported in Table 1. The degree of intrinsic inertia in price inflation is defined as the sum of the coefficients on lagged inflation rates in the GNKPC (33), and that of intrinsic inertia in wage inflation is its analogue in the wage-GNKPC (39).

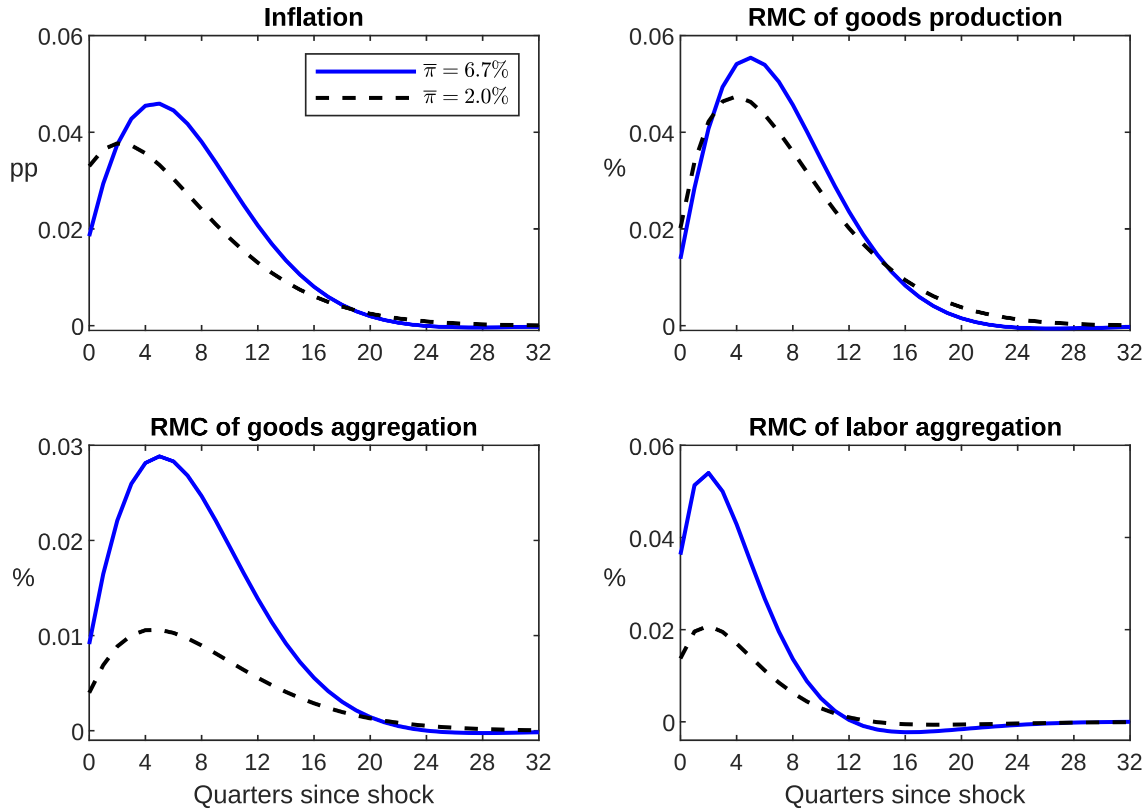
consumption expenditure price index. Thus, the model suggests an alternative explanation: the decline in trend inflation caused the decrease in inflation persistence.

4.2 Effect on impulse responses

The calibrated model shows that lower trend inflation reduces inflation persistence. Figure 6 illustrates impulse responses to an expansionary monetary policy shock at a trend inflation rate of 6.7 percent annually (the solid lines) and 2 percent annually (the dashed lines).

The upper left panel of Figure 6 shows that the response of inflation to the shock is more persistent at the higher trend inflation rate of 6.7 percent annually than at the lower rate of 2 percent annually under the calibration of other model parameters reported in Table 1. Indeed, the half-life of the inflation response is 15 quarters at the higher rate, and it declines to 10 quarters at the lower rate. At the same time, the duration between the occurrence of

Figure 6: Impulse responses at high and low trend inflation rates.



Notes: The figure presents the impulse responses to a one standard deviation expansionary monetary policy shock under the calibration of model parameters reported in Table 1. The inflation rate is displayed in annualized terms. The solid and the dashed lines assume a trend inflation rate $\bar{\pi}$ of 6.7 percent and 2 percent annually, respectively.

the shock and the peak response of inflation shortens from five quarters at the higher rate to two quarters at the lower rate. Under positive trend inflation and staggered price- and wage-setting, the Kimball-type non-CES aggregators of goods and labor cause the real marginal cost (RMC) of goods aggregation $\hat{d}_{p,t}$ to generate intrinsic inertia in price inflation and the RMC of labor aggregation $\hat{d}_{w,t}$ to generate intrinsic inertia in wage inflation, which enhances the persistence of price inflation through the RMC of goods production $\hat{mc}_t$, that is, the real wage $\hat{w}_t$. The RMCs of goods and labor aggregation increase with the level of trend inflation, as displayed in the lower panels of the figure, and therefore the lower trend inflation rate leads to lower persistence of price inflation. Thus, our model provides a new explanation for the evidence that inflation persistence decreased around the time of the Volcker disinflation. According to our explanation, the decreases in trend inflation and in inflation persistence are no coincidence; the decline in trend inflation led to the decrease in inflation persistence.

4.3 Effect on autocorrelation and predictability

How well can the calibrated model account for the decline in unconditional empirical measures of inflation persistence since the mid 1980s? Two often used measures are the sum of the autoregressive coefficients in an estimated AR(p) process for inflation and the predictability of inflation. Following Cogley et al. (2010), the h-quarter predictability is the ratio of the conditional and the unconditional variance of inflation. That is,

$$R_h^2 = 1 - \frac{\mathrm{var}_t\left(\mathbf{e}_\pi X_{t+h}\right)}{\mathrm{var}\left(\mathbf{e}_\pi X_{t+h}\right)} \approx 1 - \frac{\mathbf{e}_\pi \left[\sum_{j=0}^{h-1} \left(A^j\right) \mathrm{var}(u_t) \left(A^j\right)'\right] \mathbf{e}_\pi'}{\mathbf{e}_\pi \left[\sum_{j=0}^{\infty} \left(A^j\right) \mathrm{var}(u_t) \left(A^j\right)'\right] \mathbf{e}_\pi'}, \tag{48}$$

where A is the companion-form matrix of the AR(p) process, $X_t = (\pi_t, \pi_{t-1}, \ldots, \pi_{t-1+p})'$, u_t is the residual vector of the companion form, and $\mathbf{e}_\pi$ is a selector vector.

Table 2: Autocorrelation and predictability of inflation.

		Sum AR(p)	R_1^2	R_4^2	R_8^2
US data	1959:Q1–1984:Q4	0.915	0.786	0.529	0.318
	1985:Q1–2019:Q4	0.788	0.458	0.235	0.084
	Difference	−0.127	−0.328	−0.294	−0.234
Model	$\bar{\pi} = 6.7\%$	0.961	0.978	0.719	0.218
	$\bar{\pi} = 2\%$	0.924	0.894	0.492	0.099
	Difference	−0.037	−0.084	−0.227	−0.119

Source: US Bureau of Economic Analysis.
Notes: The second column shows the sum of the autoregressive coefficients of an estimated AR(p) process with order $p = 4$. The third to fifth columns display the predictability of inflation based on the estimated autoregressive process. The upper and lower panels present statistics for the inflation rate of, respectively, the US GDP deflator and the simulated model under the calibration of parameters reported in Table 1.

A decline in trend inflation in the calibrated model can account for a substantial portion of the decline in the unconditional empirical measures of inflation persistence since the mid 1980s. The upper panel of Table 2 presents the measures of inflation persistence for the GDP deflator during two periods obtained by splitting the sample 1959:Q1–2019:Q4 in 1985, as well as their changes. Each of the measures displays a moderate decline since 1985, in line with existing evidence.[26] To generate their model counterparts, the model is simulated for 10,000 quarters, of which the first 10 percent are discarded. Although the model generates too much autocorrelation and predictability in inflation, except for the eight-quarter predictability, it can account for a substantial portion of the declines in the autocorrelation and predictability. The simulation results show that a decline in the trend inflation rate from 6.7 percent to 2 percent annually accounts for between 1/4 and 3/4 of the observed declines in the sum of autoregressive coefficients and the predictability.

5 Disinflation

Another approach for assessing inflation persistence is to examine the response of inflation to a credible disinflation. In this section, our model is used to analyze a transition from one steady state to another one with lower positive trend inflation. We also assess whether the effect of a disinflation on steady-state output in our model is consistent with empirical evidence.

5.1 Credible disinflation in the model

Historical evidence shows that disinflations tend to be gradual and recessionary (for the evidence, see, e.g., Gordon, 1982; Ball, 1994b; Cecchetti and Rich, 2001). To account for these dynamics, the existing literature has stressed that intrinsic inertia in inflation plays a key role in NK models. As Fuhrer (2011) points out, when intrinsic inertia of inflation is absent in an NK model, a credible permanent reduction in trend inflation causes inflation to jump to its new trend rate and output to remain at its steady-state value. Once the intrinsic inertia is embedded in the model, the credible disinflation generates a gradual adjustment of inflation to its new trend rate and a temporary decline in output.[27]

Our model also accounts for the gradual adjustment of inflation and the temporary decline in output after a credible disinflation, even though price-setting decisions are purely

[26] Cogley and Sbordone (2008) and Fuhrer (2011) present similar evidence on the decline in the autocorrelation of inflation.

[27] Another explanation for the gradual decline in inflation and the output loss during a disinflation emphasizes imperfect credibility (see, e.g., Erceg and Levin, 2003; Goodfriend and King, 2005).

forward-looking. To show this, we simplify the model by abstracting from nominal wage rigidity and policy inertia (i.e., $\alpha_w = \rho = 0$), so that the lagged real marginal cost of goods aggregation and the lagged relative price distortion are the only endogenous state variables that shape the inflation response. Using the simplified model, we carry out the following experiment. In period 0, the economy is in the steady state with a trend inflation rate of 6.7 percent annually. At the start of period 1, trend inflation is reduced suddenly and credibly to 2 percent annually.[28] Denote the vector of endogenous state variables in the log-linearized models by $\hat{k}_{t-1} = \log k_{t-1} - \log k(\pi)$; for instance, $k_{t-1} = [d_{p,t-1}, \Delta_{t-1}]'$ in our model, $k_{t-1} = \pi_{t-1}$ in the NK counterpart with partial indexation to past inflation, and there is no endogenous state variable in the counterpart with full indexation to trend inflation.[29] Here $k(\pi)$ denotes the vector of steady-state values of k_{t-1}, which stresses that some of these values are functions of trend inflation π. Because in period 0 all variables are in the steady state, in period 1 the lagged endogenous state variables under the new trend inflation rate are given by $\log k(\pi^0) - \log k(\pi^1)$, where $\pi^0 = 1.067^{1/4}$ and $\pi^1 = 1.02^{1/4}$. Then, the solution of the log-linearized model under the trend inflation rate π^1 is used to compute inflation and output in period $t = 1, 2, 3, \ldots$

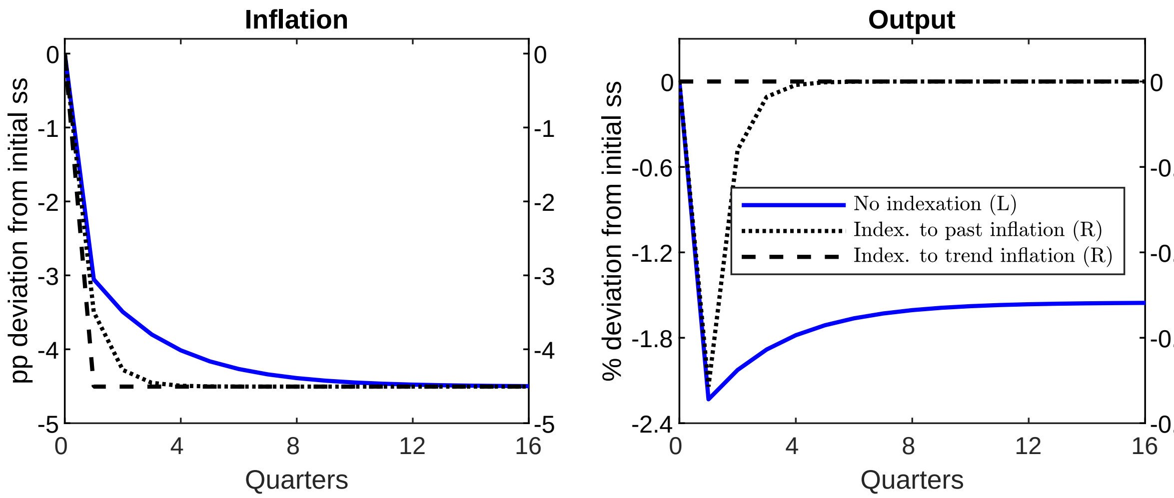
Figure 7 displays the responses of inflation and output to the sudden and credible reduction in trend inflation from 6.7 percent to 2 percent annually, using the calibration of other model parameters reported in Table 1, except for $\alpha_w = 0$ and $\rho = 0$. In the figure the dotted lines represent the responses in the NK counterpart with partial indexation to past inflation $\iota_p = 0.24$. In this model, inflation declines gradually toward its new trend rate, while output falls temporarily and then rebounds gradually to the initial steady-state value, in line with the responses indicated by Fuhrer (2011).[30] Similar responses are obtained in our model, as illustrated by the solid lines in Figure 7. This is because our model has intrinsic inertia of

[28]The disinflation is sudden in that agents did not anticipate the possibility of a change in trend inflation before period 1. The disinflation is credible in that agents believe that the new rate of trend inflation is permanent.

[29]Nominal wage rigidity adds two endogenous state variables w_{t-1} and $d_{w,t-1}$ in our model and w_{t-1} and $\pi_{w,t-1}$ in the NK counterpart with partial indexation to past inflation, respectively. Assuming $\alpha_w = 0.75$ would not qualitatively affect the results illustrated in Figure 7.

[30]In the NK counterpart with full indexation to trend inflation, the responses of inflation and output to the sudden and credible reduction in trend inflation are displayed by the dashed lines in Figure 7. In this model, inflation drops instantly to the new rate of trend inflation, while output remains at its steady-state value.

Figure 7: Credible disinflation.



Notes: The figure displays the responses of inflation and output to a sudden and credible reduction in trend inflation of -4.7 percentage points (from 6.7 percent to 2 percent annually), using the calibration of other model parameters reported in Table 1, except for $\alpha_w = 0$ and $\rho = 0$. The solid lines show the responses in our model, while the dotted and the dashed lines illustrate those in the NK counterparts with partial indexation to past inflation $\iota_p = 0.24$ and with full indexation to trend inflation (i.e., $\iota_p = 0$), respectively.

inflation through the real marginal cost of goods aggregation, as shown in the GNKPC (33).

5.2 Evidence on the long-run response of output

One notable difference between our model and the NK counterpart with partial indexation to past inflation is that output in our model rebounds to its new steady-state value associated with the new rate of trend inflation, which is lower than the initial value of steady-state output.[31] The effect of a change in trend inflation on steady-state output runs counter to the classical dichotomy, which is the view that real variables evolve independently from nominal variables in the long run. In the remainder of this section, we assess the empirical evidence for the classical dichotomy and conclude that the prediction of our calibrated model of a moderate steady-state output response to a disinflation is consistent with the evidence.

Evidence supporting the classical dichotomy is mixed. Early work by King and Watson

[31] Kurozumi and Van Zandweghe (2016) show that the Kimball-type non-CES aggregator of goods can lead a decline in trend inflation to lower steady-state output by increasing the steady-state average markup, in contrast with the case of the CES aggregator. Such an increase in the average markup is consistent with empirical evidence for the U.S. that markups have risen since 1980 (see, e.g., De Loecker et al., 2020).

(1994) finds evidence of a negative long-run relationship between inflation and unemployment, indicating a positive long-run relationship between inflation and output, using an estimated bivariate structural VAR with short-run identifying restrictions. However, their finding is not robust to adding another variable to the VAR, as shown by Evans (1994). Moreover, Bullard and Keating (1995) find no significant long-run response of output to a permanent increase in inflation in a bivariate structural VAR with long-run identifying restrictions estimated on the U.S. inflation and output growth data (although they find a positive long-run response for a few other countries). Benati (2015) revisits the long-run Phillips curve by estimating a six-variable structural VAR with long-run identifying restrictions using Bayesian methods. Using such a larger set of variables can avoid issues of informational deficiency discussed by Forni and Gambetti (2014). Although the estimation results of Benati (2015) do not allow to reject the null hypothesis of a vertical long-run Phillips curve, they provide limited support for the notion of a negative long-run relationship between inflation and unemployment in the U.S., as the response of unemployment to a positive permanent inflation shock is negative for more than 80 percent of draws from the posterior distribution. Indeed, uncertainty about the long-run response is sufficiently large that the null hypothesis of specific values of a steep Phillips curve cannot be rejected.[32]

To further investigate the long-run response of output to a permanent change in inflation, we estimate a cointegrated structural VAR with long-run identifying restrictions on six variables.[33] The variables are the inflation rate, log output, log consumption, log investment, a short-term nominal interest rate, and M1 money velocity for the sample period 1951:Q1–2019:Q4.[34] The money velocity may contain useful information about the trend inflation rate because, as pointed out by Benati (2020), M1 velocity approximately captures the permanent component of the short-term interest rate. The cointegrated VAR includes

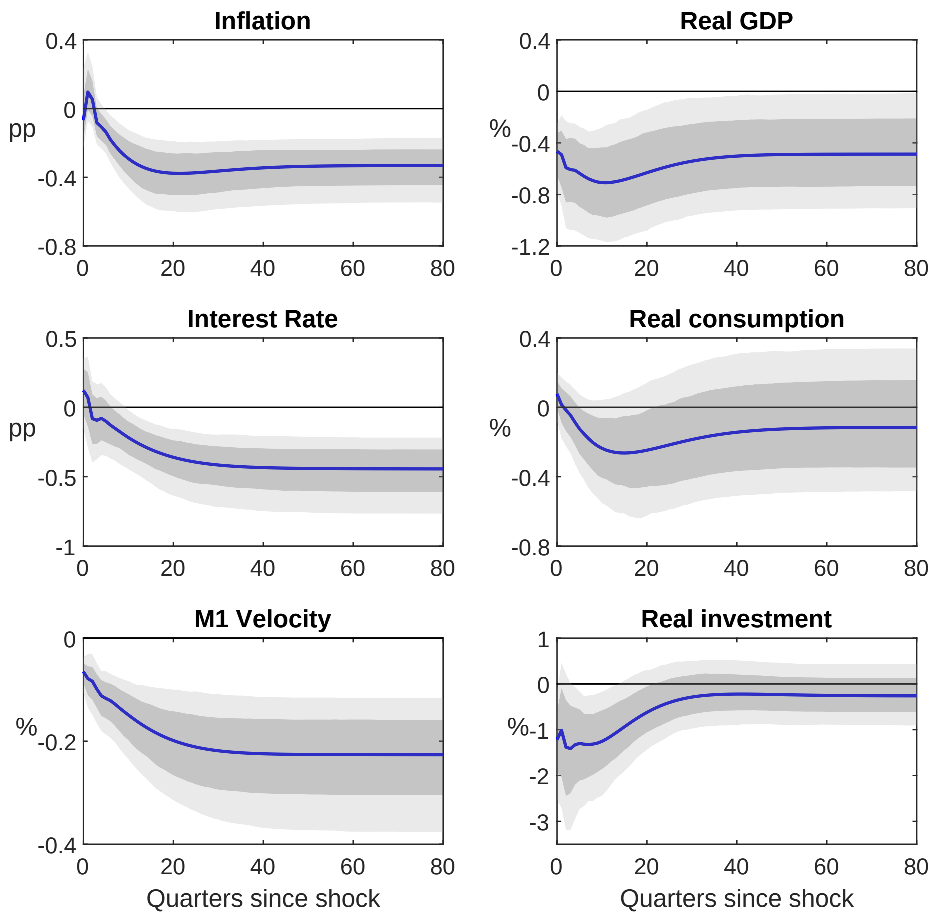
[32] Benati (2015) reports a median response of unemployment of -0.27 percentage point to a one percentage point inflation shock, which, using a simple Okun's Law, indicates a permanent output effect of 0.54 percentage point.

[33] We are grateful to an anonymous referee for generously sharing the code and data for estimating the VAR.

[34] Specifically, the data for output, consumption, and investment are, respectively, real GDP, personal consumption expenditures on nondurable goods and services deflated by the GDP deflator, and real gross private domestic investment. The short-term nominal interest rate is the 3-month Treasury bill rate. M1 velocity is calculated as the ratio of nominal GDP and nominal M1, where M1 is augmented with money market deposit accounts as in the New M1 measure of Lucas and Nicolini (2015). As before, the inflation rate is measured by that of the GDP deflator.

three cointegration vectors, based on Johansen's maximum eigenvalue test. The lag order of the VAR is set at two quarters, which is the maximum of the lag orders chosen by the Schwarz and Hannan-Quin information criteria. A permanent inflation shock is identified based on the restriction that no other shocks affect inflation in the long run.

Figure 8: Empirical impulse responses to a permanent inflation shock.



Notes: The figure presents the impulse responses to a one standard deviation negative permanent inflation shock from an estimated cointegrated structural VAR. The interest and inflation rates are displayed in annualized terms. The dark and light gray shaded areas represent, respectively, one and two standard deviation error bands obtained from 1,000 bootstrap replications.

Figure 8 plots the impulse responses to a one standard deviation negative permanent inflation shock in the estimated cointegrated structural VAR. The responses level out between 40 and 80 quarters, indicating that this window approximates the long-run responses. Two results stand out. First, the null hypothesis that a permanent inflation shock has no permanent impact on GDP can barely be rejected at the two standard deviation level. Second, uncertainty is substantial and is compatible with a considerable range of nonzero permanent responses. Consider the response of steady-state output to a disinflation in the calibrated model. The magnitude of the response in Figure 7 (-1.6 percentage points) is about one third as large as the change in trend inflation (-4.7 percentage points). Therefore, a decline in trend inflation in the model equal to the median long-run inflation response in Figure 8 of -0.33 percentage point generates a response of steady-state output of -0.11 ($= -0.33/3$) percentage point.[35] Then, in the estimated VAR, the null hypothesis of a long-run decline in real GDP of 0.11 percentage point cannot be rejected at the one or two standard deviation level. Thus, given the estimated uncertainty about the long-run response of real GDP to a permanent inflation shock, the moderate predicted response of output in our model is within the range of empirically plausible responses.

6 Conclusion

This paper has proposed a novel theory of intrinsic inflation persistence by introducing trend inflation and Kimball-type non-CES aggregators of individual differentiated goods and labor in a model with Calvo-style staggered price- and wage-setting. Under nonzero trend inflation, the Kimball-type non-CES aggregator of individual goods and staggered price-setting give rise to a variable real marginal cost of goods aggregation, which becomes a driver of inflation. The real marginal cost then consists of an aggregate of the goods' relative prices, which depends on current and past inflation rates, thereby generating intrinsic inertia in inflation. Likewise, the non-CES aggregator of labor and staggered wage-setting lead to intrinsic inertia in wage inflation, which enhances the persistence of price inflation. The model provides a microfoundation of intrinsic inflation persistence without relying on ad hoc backward-looking price-setting behavior. In a plausibly calibrated version of the model, inflation exhibits a persistent response to an expansionary monetary policy shock, with a hump shape and

[35] The responsiveness of steady-state output to a change in trend inflation in the calibrated model depends on the values of parameters α_p, θ_p, and ϵ_p. The responsiveness is less at smaller values of α_p or θ_p. The response of steady-state output to trend inflation is also less at somewhat larger values of ϵ_p, although it changes sign at values of ϵ_p near zero. In the case of the CES aggregator (i.e., $\epsilon_p = 0$), the decline in trend inflation generates an increase in steady-state output.

a gradual decline. The model has also shown that lower trend inflation reduces inflation persistence, providing a new explanation for the measured decrease in inflation persistence around the time of the Volcker disinflation. Moreover, the paper has demonstrated that a credible permanent reduction in trend inflation leads to a gradual decline in inflation and a fall in output.

Appendix A Derivation of the Generalized New Keynesian Phillips Curve

In this appendix we outline the steps for deriving the GNKPC. The derivation of the wage-GNKPC follows analogous steps when a unit elasticity of labor supply is assumed.

First, under assumption (28), log-linearizing price-adjusting firms' first-order condition (9) yields

$$
\begin{aligned}
&\frac{\gamma_p - 1 - \epsilon_{p2}(1+\gamma_p)}{\gamma_p - 1}\hat{p}_t^* \\
&= \frac{\gamma_p - 1 - \epsilon_{p2}}{\gamma_p - 1}\left(1 - \alpha_p\beta\pi^{\gamma_p}\right)\sum_{j=0}^{\infty}(\alpha_p\beta\pi^{\gamma_p})^j\left(E_t\hat{mc}_{t+j} + \gamma_p\hat{d}_{p,t+j}\right) \\
&\quad + \frac{\gamma_p - 1 - \epsilon_{p2}}{\gamma_p - 1}\gamma_p\sum_{j=1}^{\infty}(\alpha_p\beta\pi^{\gamma_p})^j E_t\hat{\pi}_{t+j} - \gamma_p\left(1 - \alpha_p\beta\pi^{\gamma_p - 1}\right)\sum_{j=0}^{\infty}(\alpha_p\beta\pi^{\gamma_p-1})^j E_t\hat{d}_{p,t+j} \\
&\quad - (\gamma_p - 1)\sum_{j=1}^{\infty}(\alpha_p\beta\pi^{\gamma_p-1})^j E_t\hat{\pi}_{t+j} - \frac{\epsilon_{p2}}{\gamma_p - 1}\sum_{j=1}^{\infty}(\alpha_p\beta\pi^{-1})^j E_t\hat{\pi}_{t+j}. \qquad (49)
\end{aligned}
$$

Iterating this equation forward one period and multiplying both sides by $\alpha_p\beta\pi^{\gamma_p}$, we obtain

$$
\begin{aligned}
&\alpha_p\beta\pi^{\gamma_p}\frac{\gamma_p - 1 - \epsilon_{p2}(1+\gamma_p)}{\gamma_p - 1}E_t\hat{p}_{t+1}^* \\
&= \frac{\gamma_p - 1 - \epsilon_{p2}}{\gamma_p - 1}\left(1 - \alpha_p\beta\pi^{\gamma_p}\right)\sum_{j=0}^{\infty}(\alpha_p\beta\pi^{\gamma_p})^{j+1}\left(E_t\hat{mc}_{t+1+j} + \gamma_p\hat{d}_{p,t+1+j}\right) \\
&\quad + \frac{\gamma_p - 1 - \epsilon_{p2}}{\gamma_p - 1}\gamma_p\sum_{j=1}^{\infty}(\alpha_p\beta\pi^{\gamma_p})^{j+1} E_t\hat{\pi}_{t+1+j} - \alpha_p\beta\pi^{\gamma_p}\gamma_p\left(1 - \alpha_p\beta\pi^{\gamma_p - 1}\right)\sum_{j=0}^{\infty}(\alpha_p\beta\pi^{\gamma_p-1})^j E_t\hat{d}_{p,t+1+j} \\
&\quad - \alpha_p\beta\pi^{\gamma_p}(\gamma_p - 1)\sum_{j=1}^{\infty}(\alpha_p\beta\pi^{\gamma_p-1})^j E_t\hat{\pi}_{t+1+j} - \alpha_p\beta\pi^{\gamma_p}\frac{\epsilon_{p2}}{\gamma_p - 1}\sum_{j=1}^{\infty}(\alpha_p\beta\pi^{-1})^j E_t\hat{\pi}_{t+1+j}. \qquad (50)
\end{aligned}
$$

Subtracting (50) from (49) yields

$$
\begin{aligned}
&\frac{\gamma_p - 1 - \epsilon_{p2}(1+\gamma_p)}{\gamma_p - 1}\hat{p}_t^* - \alpha_p\beta\pi^{\gamma_p}\frac{\gamma_p - 1 - \epsilon_{p2}(1+\gamma_p)}{\gamma_p - 1}E_t\hat{p}_{t+1}^* \\
&= \frac{\gamma_p - 1 - \epsilon_{p2}}{\gamma_p - 1}\left(1 - \alpha_p\beta\pi^{\gamma_p}\right)\hat{mc}_t - \gamma_p\left(\alpha_p\beta\pi^{\gamma_p-1}(\pi - 1) + \frac{\epsilon_{p2}}{\gamma_p - 1}(1 - \alpha_p\beta\pi^{\gamma_p})\right)\hat{d}_{p,t} \\
&\quad + \frac{\gamma_p - 1 - \epsilon_{p2}(1+\gamma_p)}{\gamma_p - 1}\alpha_p\beta\pi^{\gamma_p}E_t\hat{\pi}_{t+1}
\end{aligned}
$$

$$+ (\pi - 1) \sum_{j=1}^{\infty} (\alpha_p \beta \pi^{\gamma_p - 1})^j \left(\gamma_p (1 - \alpha_p \beta \pi^{\gamma_p - 1}) E_t \hat{d}_{t+j} + (\gamma_p - 1) E_t \hat{\pi}_{t+j} \right)$$

$$+ \left(\pi^{\gamma_p + 1} - 1 \right) \frac{\epsilon_{p2}}{\gamma_p - 1} \sum_{j=1}^{\infty} (\alpha_p \beta \pi^{-1}) E_t \hat{\pi}_{t+j}. \tag{51}$$

Next, log-linearizing the law of motion for the real marginal cost of goods aggregation (10) yields

$$\hat{p}_t^* = \frac{\alpha_p \pi^{\gamma_p - 1}}{1 - \alpha_p \pi^{\gamma_p - 1}} \hat{\pi}_t + \frac{1}{1 - \alpha_p \pi^{\gamma_p - 1}} \hat{d}_{p,t} - \frac{\alpha_p \pi^{\gamma_p - 1}}{1 - \alpha_p \pi^{\gamma_p - 1}} \hat{d}_{p,t-1}. \tag{52}$$

Finally, using (52) to substitute for $\hat{p}_t^*$ in (51), iterating forward (52) to substitute for $E_t \hat{p}_{t+1}^*$ in (51), and rearranging terms yields the GNKPC (29) with the conditions (30) and (31) for the auxiliary variables $\varphi_{p,t}$ and $\psi_{p,t}$, respectively.[36]

Appendix B Composite Coefficients in Log-Linearized Equilibrium Conditions of the Model

The composite coefficients in log-linearized equilibrium conditions (29)–(32) are given by

$$\kappa_p \equiv \frac{(1 - \alpha_p \pi^{\gamma_p - 1})(1 - \alpha_p \beta \pi^{\gamma_p})}{\alpha_p \pi^{\gamma_p - 1} [1 - \epsilon_{p2} \gamma_p / (\gamma_p - 1 - \epsilon_{p2})]}, \quad \kappa_{pd} \equiv \gamma_p (\kappa_p - \tilde{\kappa}_{pd}) - \alpha_p \beta \pi^{\gamma_p} - \frac{1}{\alpha_p \pi^{\gamma_p - 1}},$$

$$\kappa_{p\varphi} \equiv \frac{\beta (\pi - 1)(1 - \alpha_p \pi^{\gamma_p - 1})}{1 - \epsilon_{p2} (1 + \gamma_p) / (\gamma_p - 1)}, \quad \kappa_{p\epsilon\psi} \equiv \frac{\epsilon_{p2} \beta (\pi^{1 + \gamma_p} - 1)(1 - \alpha_p \pi^{\gamma_p - 1})}{\pi^{\gamma_p} [\gamma_p - 1 - \epsilon_{p2} (1 + \gamma_p)]},$$

where

$$\tilde{\kappa}_{pd} \equiv \frac{(1 - \alpha_p \pi^{\gamma_p - 1})(1 - \alpha_p \beta \pi^{\gamma_p - 1})}{\alpha_p \pi^{\gamma_p - 1} [1 - \epsilon_{p2} (1 + \gamma_p) / (\gamma_p - 1)]}, \quad \epsilon_{p2} \equiv \epsilon_{p1} \frac{1 - \alpha_p \beta \pi^{\gamma_p - 1}}{1 - \alpha_p \beta \pi^{-1}} \leq 0$$

for $\epsilon_p \leq 0$ under assumption (28).

The composite coefficients in log-linearized equilibrium conditions (34)–(38) are given by

$$\kappa_w \equiv \frac{(1 + \epsilon_{w1})(1 - \alpha_w \pi^{\gamma_w - 1})(1 - \alpha_w \beta \pi^{2\gamma_w})}{\alpha_w \pi^{\gamma_w - 1} \{(1 + \epsilon_{w3})[1 - \epsilon_{w2} \gamma_w / (\gamma_w - 1 - \epsilon_{w2})] + \gamma_w\}},$$

$$\tilde{\kappa}_{wd} \equiv \frac{(1 + \epsilon_{w3})(1 - \alpha_w \pi^{\gamma_w - 1})(1 - \alpha_w \beta \pi^{\gamma_w - 1})}{\alpha_w \pi^{\gamma_w - 1} \{(1 + \epsilon_{w3})[1 - \epsilon_{w2} (1 + \gamma_w) / (\gamma_w - 1)] + \gamma_w [1 - \epsilon_{w2} / (\gamma_w - 1)]\}},$$

[36] Kurozumi and Van Zandweghe (2016) present an alternative representation of the GNKPC. Their representation can be obtained by multipling the forward iterated (49) by a factor $\alpha_p \beta \pi^{\gamma_p - 1}$ instead of $\alpha_p \beta \pi^{\gamma_p}$ and following the same steps as outlined above.

$$\kappa_{w\epsilon} \equiv \frac{\epsilon_{w2}(1+\epsilon_{w3})(1-\alpha_w \pi^{\gamma_w - 1})(1-\alpha_w \beta \pi^{-1})}{\alpha_w \pi^{\gamma_w - 1}\{(1+\epsilon_{w3})[\gamma_w - 1 - \epsilon_{w2}(1+\gamma_w)] + \gamma_w(\gamma_w - 1 - \epsilon_{w2})\}},$$

$$\kappa_{wd} \equiv \gamma_w \left[\kappa_w \left(1 + \frac{1}{1+\epsilon_{w1}}\right) - \tilde{\kappa}_{wd}\right] - \alpha_w \beta \pi^{2\gamma_w} - \frac{1}{\alpha_w \pi^{\gamma_w - 1}},$$

$$\kappa_{w\epsilon\zeta} \equiv -\frac{\epsilon_{w3}\beta\pi(\pi^{\gamma_w} - 1)(1-\alpha_w \pi^{\gamma_w - 1})}{(1+\epsilon_{w3})[1 - \epsilon_{w2}\gamma_w/(\gamma_w - 1 - \epsilon_{w2})] + \gamma_w},$$

$$\kappa_{w\varphi} \equiv \frac{\beta(\pi^{\gamma_w + 1} - 1)(1+\epsilon_{w3})(1-\alpha_w \pi^{\gamma_w - 1})}{(1+\epsilon_{w3})[1 - \epsilon_{w2}(1+\gamma_w)/(\gamma_w - 1)] + \gamma_w[1 - \epsilon_{w2}/(\gamma_w - 1)]},$$

$$\kappa_{w\epsilon\psi} \equiv -\frac{\epsilon_{w2}\beta(\pi^{2\gamma_w + 1} - 1)(1+\epsilon_{w3})(1-\alpha_w \pi^{\gamma_w - 1})}{\pi^{\gamma_w}\{(1+\epsilon_{w3})[\gamma_w - 1 - \epsilon_{w2}(1+\gamma_w)] + \gamma_w(\gamma_w - 1 - \epsilon_{w2})\}},$$

where

$$\epsilon_{w2} \equiv \epsilon_{w1}\frac{1-\alpha_w \beta \pi^{\gamma_w - 1}}{1-\alpha_w \beta \pi^{-1}} \leq 0, \quad \epsilon_{w3} \equiv \epsilon_{w1}\frac{1-\alpha_w \beta \pi^{2\gamma_w}}{1-\alpha_w \beta \pi^{\gamma_w}} \leq 0$$

for $\epsilon_w \leq 0$ under assumption (28).

Appendix C Log-Linearized Equilibrium Conditions of the Extended Model

In Section 3.4, the extended model is augmented with working capital, habit formation in consumption preferences, the monetary policy response to output growth, and (nonstationary) technology shocks. The production technology (6) is extended so that it includes total factor productivity A_t: $Y_t(f) = A_t\, N_t(f)$. The productivity follows a random walk

$$\log A_t = \log A_{t-1} + \varepsilon_{a,t},$$

where $\varepsilon_{a,t} \sim N(0, \sigma_a^2)$ is an i.i.d. technology shock. The spending Euler equation (40), the Taylor-type monetary policy rule (41), the real marginal cost of goods production (42), the definition of wage inflation (43), and the aggregate production equation (44) are extended to, respectively,

$$\hat{\lambda}_t = E_t \hat{\lambda}_{t+1} + \hat{\imath}_t - E_t \hat{\pi}_{t+1} - E_t \varepsilon_{a,t+1},$$

$$\hat{\imath}_t = \rho\, \hat{\imath}_{t-1} + (1-\rho)[\phi_\pi \hat{\pi}_t + \phi_Y \hat{y}_t + \phi_g\, (\hat{y}_t - \hat{y}_{t-1} + \varepsilon_{a,t})] + \varepsilon_{i,t},$$

$$\hat{mc}_t = \hat{\omega}_t + \hat{\imath}_t,$$

$$\hat{\pi}_{w,t} = \hat{\omega}_t - \hat{\omega}_{t-1} + \varepsilon_{a,t} + \hat{\pi}_t,$$

$$\hat{y}_t = \hat{N}_t - \hat{\Delta}_t,$$

where $\hat{\lambda}_t = -\left(\hat{y}_t - b\hat{y}_{t-1} + b\varepsilon_{a,t}\right) / (1-b)$, $y_t \equiv Y_t/A_t$ is detrended output, $\omega_t \equiv w_t/A_t$ is the detrended real wage, $\lambda_t \equiv \Lambda_t A_t$ is the associated real value of the Lagrange multiplier on the household's budget constraint (16), and $b \in [0,1)$ is the degree of habit persistence in consumption preferences. The GNKPC and the conditions for its auxiliary variables (29)–(31) are extended to, respectively,

$$\hat{\pi}_t = \beta\pi E_t\hat{\pi}_{t+1} + \kappa_p \hat{mc}_t - \kappa_\lambda \left(\hat{y}_t + \hat{\lambda}_t\right) + \kappa_{pd}\hat{d}_{p,t} + \hat{d}_{p,t-1} + \beta\pi E_t\hat{d}_{p,t+1} + \varphi_{p,t} + \psi_{p,t},$$

$$\varphi_{p,t} = \alpha_p\beta\pi^{\gamma_p-1}E_t\varphi_{p,t+1} + \kappa_{p\varphi}\Big[(1-\alpha_p\beta\pi^{\gamma_p-1})\left(\gamma_p E_t\hat{d}_{p,t+1} + E_t\hat{y}_{t+1} + E_t\hat{\lambda}_{t+1}\right) + (\gamma_p - 1)E_t\hat{\pi}_{t+1}\Big],$$

$$\psi_{p,t} = \alpha_p\beta\pi^{-1}E_t\psi_{p,t+1} + \kappa_{p\epsilon\psi}\left(E_t\hat{\pi}_{t+1} + E_t\hat{y}_{t+1} + E_t\hat{\lambda}_{t+1}\right),$$

while the wage-GNKPC and the conditions for its auxiliary variables (34)–(37) are extended to, respectively,

$$\begin{aligned}
\hat{\pi}_{w,t} = {} & \beta\pi^{\gamma_w+1}E_t\hat{\pi}_{w,t+1} + \kappa_w\left(2\hat{N}_t - \hat{\omega}_t\right) - (\tilde{\kappa}_{wd} - \kappa_{w\epsilon})\left(\hat{N}_t + \hat{\lambda}_t\right) + \kappa_{wd}\hat{d}_{w,t} + \hat{d}_{w,t-1} \\
& + \beta\pi^{\gamma_w+1}E_t\hat{d}_{w,t+1} + \zeta_{w,t} + \varphi_{w,t} + \psi_{w,t}, \\
\zeta_{w,t} = {} & \alpha_w\beta\pi^{\gamma_w}E_t\zeta_{w,t+1} \\
& + \kappa_{w\epsilon\zeta}\Big[(1-\alpha_w\beta\pi^{\gamma_w})\left(2E_t\hat{N}_{t+1} - E_t\hat{\omega}_{t+1} + \gamma_w E_t\hat{d}_{w,t+1}\right) + E_t\hat{\omega}_{t+1} - \hat{\omega}_t + \gamma_w E_t\hat{\pi}_{w,t+1}\Big], \\
\varphi_{w,t} = {} & \alpha_w\beta\pi^{\gamma_w-1}E_t\varphi_{w,t+1} \\
& + \kappa_{w\varphi}\Big[(1-\alpha_w\beta\pi^{\gamma_w-1})\left(E_t\hat{N}_{t+1} + E_t\hat{\lambda}_{t+1} + \gamma_w E_t\hat{d}_{w,t+1}\right) + \gamma_w E_t\hat{\pi}_{w,t+1} - E_t\hat{\pi}_{t+1} - E_t\varepsilon_{a,t+1}\Big], \\
\psi_{w,t} = {} & \alpha_w\beta\pi^{-1}E_t\psi_{w,t+1} + \kappa_{w\epsilon\psi}\Big[(1-\alpha_w\beta\pi^{-1})\left(E_t\hat{N}_{t+1} + E_t\hat{\lambda}_{t+1}\right) - E_t\hat{\pi}_{t+1} - E_t\varepsilon_{a,t+1}\Big].
\end{aligned}$$

In these log-linearized equilibrium conditions, the composite coefficients κ_p, κ_{pd}, $\kappa_{p\varphi}$, $\kappa_{p\epsilon\psi}$, κ_w, $\kappa_{w\epsilon\zeta}$, $\tilde{\kappa}_{wd}$, $\kappa_{w\epsilon}$, κ_{wd}, $\kappa_{w\varphi}$, and $\kappa_{w\epsilon\psi}$ remain unchanged from the baseline model, while the additional one is given by

$$\kappa_\lambda = \frac{1-\alpha_p\pi^{\gamma_p-1}}{\alpha_p\pi^{\gamma_p-1}\left[1-\epsilon_{p2}(1+\gamma_p)/(\gamma_p-1)\right]}\left[\alpha_p\beta\pi^{\gamma_p-1}(\pi-1) - \frac{\epsilon_{p2}}{\gamma_p-1}\alpha_p\beta\pi^{-1}\left(\pi^{\gamma_p+1}-1\right)\right].$$

The laws of motion of the real marginal costs of goods and labor aggregation (32) and (38) and the law of motion of the relative price distortion (45) remain unchanged.

References

Ascari, Guido and Tiziano Ropele (2009). "Trend inflation, Taylor principle and indeterminacy." *Journal of Money, Credit and Banking*, 41(8), pp. 1557–1584.

Ascari, Guido and Argia M. Sbordone (2014). "The macroeconomics of trend inflation." *Journal of Economic Literature*, 52(3), pp. 679–739.

Ball, Laurence (1994a). "Credible disinflation with staggered price-setting." *American Economic Review*, 84(1), pp. 282–289.

Ball, Laurence (1994b). "What determines the sacrifice ratio?" In N. Gregory Mankiw, editor, *Monetary Policy*, chapter 5, pp. 155–182.

Barattieri, Alessandro, Susanto Basu, and Peter Gottschalk (2014). "Some evidence on the importance of sticky wages." *American Economic Journal: Macroeconomics*, 6(1), pp. 70–101.

Beck, Günther and Sarah M. Lein (2020). "Price elasticities and demand-side real rigidities in micro data and in macro models." *Journal of Monetary Economics*, 115, pp. 200–212.

Benati, Luca (2008). "Investigating inflation persistence across monetary regimes." *Quarterly Journal of Economics*, 123(3), pp. 1005–1060.

Benati, Luca (2015). "The long-run Phillips curve: A structural VAR investigation." *Journal of Monetary Economics*, 76, pp. 15–28.

Benati, Luca (2020). "Money velocity and the natural rate of interest." *Journal of Monetary Economics*, 116, pp. 117–134.

Benati, Luca and Paolo Surico (2008). "Evolving U.S. monetary policy and the decline of inflation predictability." *Journal of the European Economic Association*, 6(2–3), pp. 634–646.

Bullard, James and John W. Keating (1995). "The long-run relationship between inflation and output in postwar economies." *Journal of Monetary Economics*, 36(3), pp. 477–496.

Calvo, Guillermo A. (1983). "Staggered prices in a utility-maximizing framework." *Journal of Monetary Economics*, 12(3), pp. 383–398.

Carlstrom, Charles T., Timothy S. Fuerst, and Matthias Paustian (2009). "Inflation persistence, monetary policy, and the Great Moderation." *Journal of Money, Credit and Banking*, 41(4), pp. 767–786.

Cecchetti, Stephen G. and Robert W. Rich (2001). "Structural estimates of the U.S. sacrifice ratio." *Journal of Business and Economic Statistics*, 19(4), pp. 416–427.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005). "Nominal rigidities and the dynamic effects of a shock to monetary policy." *Journal of Political Economy*, 113(1), pp. 1–45.

Christiano, Lawrence J., Martin S. Eichenbaum, and Mathias Trabandt (2015). "Understanding the Great Recession." *American Economic Journal: Macroeconomics*, 7(1), pp. 110–167.

Christiano, Lawrence J., Mathias Trabandt, and Karl Walentin (2011). "DSGE models for monetary policy analysis." In Benjamin M. Friedman and Michael Woodford, editors, *Handbook of Monetary Economics*, volume 3A, chapter 7, pp. 285–367. Elsevier.

Cogley, Timothy, Giorgio E. Primiceri, and Thomas J. Sargent (2010). "Inflation-gap persistence in the US." *American Economic Journal: Macroeconomics*, 2(1), pp. 43–69.

Cogley, Timothy and Thomas J. Sargent (2002). "Evolving post-World War II U.S. inflation dynamics." In Ben S. Bernanke and Kenneth Rogoff, editors, *NBER Macroeconomics Annual 2001*, pp. 331–373.

Cogley, Timothy and Argia M. Sbordone (2008). "Trend inflation, indexation, and inflation persistence in the New Keynesian Phillips curve." *American Economic Review*, 98(5), pp. 2101–2126.

Coibion, Olivier and Yuriy Gorodnichenko (2011). "Monetary policy, trend inflation, and the Great Moderation: An alternative interpretation." *American Economic Review*, 101(1), pp. 341–370.

Damjanovic, Tatiana and Charles Nolan (2010). "Relative price distortions and inflation persistence." *Economic Journal*, 120(547), pp. 1080–1099.

Davig, Troy and Taeyoung Doh (2014). "Monetary policy regime shifts and inflation persistence." *Review of Economics and Statistics*, 96(5), pp. 862–875.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger (2020). "The rise of market power and the macroeconomic implications." *Quarterly Journal of Economics*, 135(2), pp. 561–644.

Dossche, Maarten, Freddy Heylen, and Dirk Van den Poel (2010). "The kinked demand curve and price rigidity: Evidence from scanner data." *Scandinavian Journal of Economics*, 112(4), pp. 723–752.

Dotsey, Michael and Robert G. King (2005). "Implications of state-dependent pricing for dynamic macroeconomic models." *Journal of Monetary Economics*, 52(1), pp. 213–242.

Dupor, Bill, Tomiyuki Kitamura, and Takayuki Tsuruga (2010). "Integrating sticky prices and sticky information." *Review of Economics and Statistics*, 92(3), pp. 657–669.

Eichenbaum, Martin and Jonas D. M. Fisher (2007). "Estimating the frequency of price re-optimization in Calvo-style models." *Journal of Monetary Economics*, 54(7), pp. 2032–2047.

Erceg, Christopher J. and Andrew T. Levin (2003). "Imperfect credibility and inflation persistence." *Journal of Monetary Economics*, 50(4), pp. 915–944.

Evans, Charles L. (1994). "The post-war U.S. Phillips curve: A comment." *Carnegie Rochester Conference Series on Public Policy*, 41, pp. 221–230.

Forni, Mario and Luca Gambetti (2014). "Sufficient information in structural VARs." *Journal of Monetary Economics*, 66, pp. 124–136.

Fuhrer, Jeffrey and George Moore (1995). "Inflation persistence." *Quarterly Journal of Economics*, 110(1), pp. 127–159.

Fuhrer, Jeffrey C. (2000). "Habit formation in consumption and its implications for monetary-policy models." *American Economic Review*, 90(3), pp. 367–390.

Fuhrer, Jeffrey C. (2011). "Inflation persistence." In Benjamin M. Friedman and Michael Woodford, editors, *Handbook of Monetary Economics*, volume 3A, chapter 9, pp. 423–486. Elsevier.

Galí, Jordi and Mark Gertler (1999). "Inflation dynamics: A structural econometric analysis." *Journal of Monetary Economics*, 44(2), pp. 195–222.

Goodfriend, Marvin and Robert G. King (2005). "The incredible Volcker disinflation." *Journal of Monetary Economics*, 52(5), pp. 981–1015.

Gordon, Robert J. (1982). "Why stopping inflation may be costly: Evidence from fourteen historical episodes." In Robert E. Hall, editor, *Inflation: Causes and Effects*, chapter 1, pp. 11–40.

Guerrieri, Luca, Christopher Gust, and J. David López-Salido (2010). "International competition and inflation: A New Keynesian perspective." *American Economic Journal: Macroeconomics*, 2(4), pp. 247–280.

Hofmann, Boris, Gert Peersman, and Roland Straub (2012). "Time variation in U.S. wage dynamics." *Journal of Monetary Economics*, 59, pp. 769–783.

Kimball, Miles S. (1995). "The quantitative analytics of the basic neomonetarist model." *Journal of Money, Credit and Banking*, 27(4), pp. 1241–1277.

King, Robert G. and Mark W. Watson (1994). "The post-war U.S. Phillips curve: A revisionist econometric history." *Carnegie-Rochester Conference Series on Public Policy*, 41, pp. 157–219.

Klenow, Peter J. and Oleksiy Kryvtsov (2008). "State-dependent or time-dependent pricing: Does it matter for recent U.S. inflation?" *Quarterly Journal of Economics*, 123(3), pp. 863–904.

Kreps, David M. (1998). "Anticipated utility and dynamic choice." In Donald P. Jacobs, Ehud Kalai, and Morton I. Kamien, editors, *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*, volume 3A, chapter 9, pp. 242–274. Cambridge University Press.

Kurozumi, Takushi and Willem Van Zandweghe (2016). "Kinked demand curves, the natural rate hypothesis, and macroeconomic stability." *Review of Economic Dynamics*, 20, pp. 240–257.

Levin, Andrew T., J. David López-Salido, Edward Nelson, and Tack Yun (2008). "Macroeconomic equivalence, microeconomic dissonance, and the design of monetary policy." *Journal of Monetary Economics*, 55(Supplement), pp. S48–S62.

Lucas, Jr., Robert E. and Juan Pablo Nicolini (2015). "On the stability of money demand." *Journal of Monetary Economics*, 73, pp. 48–65.

Mankiw, N. Gregory and Ricard Reis (2002). "Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve." *Quarterly Journal of Economics*, 117(4), pp. 1295–1328.

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

Nakamura, Emi, Jón Steinsson, Patrick Sun, and Daniel Villar (2018). "The elusive costs of inflation: Price dispersion during the U.S. Great Inflation." *Quarterly Journal of Economics*, 133(4), pp. 1933–1980.

Phaneuf, Louis, Eric Sims, and Jean Gardy Victor (2018). "Inflation, output and markup dynamics with purely forward-looking wage and price setters." *European Economic Review*, 105, pp. 115–134.

Pivetta, Frederic and Ricardo Reis (2007). "The persistence of inflation in the United States." *Journal of Economic Dynamics and Control*, 31(4), pp. 1326–1358.

Sheedy, Kevin D. (2010). "Intrinsic inflation persistence." *Journal of Monetary Economics*, 57(8), pp. 1049–1061.

Smets, Frank and Rafael Wouters (2007). "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review*, 97(3), pp. 586–606.

Stock, James H. and Mark W. Watson (2007). "Why has U.S. inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39(Supplement s1), pp. 3–33.

Taylor, John B. (1993). "Discretion versus policy rules in practice." *Carnegie-Rochester Conference Series on Public Policy*, 39, pp. 195–214.

Woodford, Michael (2007). "Interpreting inflation persistence: Comments on the conference on 'Quantitative Evidence on Price Determination'." *Journal of Money, Credit and Banking*, 39(Supplement s1), pp. 203–210.