Monetary Policy in the Data Economy

Zhaorui Li[*] Zhiwei Xu[†] Francesco Zanetti [‡]

This version: September 14, 2026

Abstract

We study the transmission mechanism and optimal monetary policy in an economy where data is produced as a byproduct of economic activity and improves firms' demand forecasts. In a data-intensive economy, monetary policy gains a new propagation channel that expands aggregate supply through a direct effect and a dynamic feedback loop effect between real activity and data generation. These effects amplify the output response and attenuate the inflation response to monetary policy, while relaxing the conditions under which monetary policy can generate self-fulfilling expectations. We present firm- and industry-level evidence of this mechanism and of the link between data generation and firms' forecasting accuracy. Data leaves the optimal targeting rule for output and prices unchanged but alters monetary policy implementation: a Taylor rule augmented with the value of data closely replicates the optimal commitment policy.

Keywords Data economy, Monetary policy, Information frictions, Optimal Policy

[*]Antai College of Economics and Management, Shanghai Jiao Tong University; Email: lizhaorui97@gmail.com.
[†]China Center for Economic Studies, School of Economics, Fudan University; Email: xuzhiwei09@gmail.com.
[‡]University of Oxford and CEPR; Email: francesco.zanetti@economics.ox.ac.uk.

1 Introduction

In the modern economy, data is a byproduct of economic activity: sales generate new data that improves forecast accuracy and supports additional sales. This *feedback loop* between economic activity and data is a hallmark of the data economy, yet it is absent from models of monetary transmission.[1] Our paper asks whether the data feedback loop matters for monetary policy: does it alter how monetary policy shocks propagate to output and inflation, and should it alter how the central bank conducts monetary policy? We answer these questions in three steps: a simple two-period model that isolates the mechanism, firm- and industry-level data that show its empirical relevance, and a New Keynesian model of the data economy that characterizes optimal monetary policy.

Simple model and mechanism. To isolate the mechanism, we develop a simple two-period model that clearly outlines the relationship between monetary policy and data. The model features firms facing demand uncertainty driven by a noisy signal about next-period demand, which they forecast to maximize profits. Nominal price rigidities generate real effects of monetary policy. Data is a byproduct of economic activity that reduces noise in the demand signal, thereby improving firms' forecasting accuracy and increasing sales. The data feedback loop generates two novel mechanisms in the transmission of monetary policy: (i) a direct effect and (ii) a dynamic feedback effect.

The *direct effect* arises from the immediate impact of monetary policy on real activity and production, which influences the generation of new data through the feedback loop. For instance, an expansionary monetary policy generates higher economic activity and increases the supply of data on impact, which, in turn, improves the precision of the demand signal and enhances the firm's forecast of demand in the next period. This reinforces the standard expansionary effect of monetary policy on real activity while containing the inflation increase resulting from the expansion of goods supply.

The *dynamic feedback effect* arises because higher output in the second period, driven by better forecasts from more data in the current period, makes data more valuable to the firm in the first period. Thus, expecting higher future profits from generating more data, the firm increases current demand for data to improve signal precision and take advantage of the expansionary policy, further increasing production and therefore the supply of data, thereby providing additional economic stimulus. Overall, the simple model shows that data plays a relevant role in monetary policy by strengthening the output response while curbing the inflation response to the policy action.

Empirical relevance. We empirically examine the premise of our theory, namely that data

[1] As discussed in Veldkamp and Chung (2024) and Baley and Veldkamp (2025), the feedback loop arises whenever data is a byproduct of activity.

generated by economic activity improves firms' forecasts. We test whether data-intensive firms exhibit a stronger output response and a muted price response to monetary policy shocks than low-data-intensive firms. We identify data-intensive firms by their capacity to collect and process data, as inferred from the presence of data scientists, using the Bureau of Labor Statistics' Occupational Employment and Wage Statistics. We construct novel measures of firms' data intensity and focus on the manufacturing sector, where firms use their own data to predict order demand, consistent with our theoretical framework.

We test whether data-intensive and non-data-intensive firms respond differently in output and prices to contemporaneous monetary policy shocks using the local projection method from Jordà (2005) and Cloyne et al. (2024). We also test the dynamic feedback channel by focusing on forward guidance to study whether expectations of future monetary easing generate differential effects consistent with our theory. We show that output at data-intensive firms responds more strongly to monetary policy (across both interest rate changes and forward guidance), corroborating the two novel channels of our theory. We also show that data-intensive firms have lower forecast errors—derived from one-quarter-ahead managerial forecast errors on earnings per share (EPS) from Thomson Reuters I/B/E/S—supporting the significant role of data in strengthening the firm's forecasting ability.

Monetary policy. To study monetary policy, we cast our model in a standard New Keynesian framework, which nests the canonical three-equation model comprising the dynamic IS curve, the New Keynesian Phillips Curve (NKPC), and the Taylor Rule (as in Woodford, 2003; Galí, 2015) as a special case. Unlike the standard NK model, we introduce a data intermediary that collects and processes raw data and provides firms with structured data to improve demand forecasting. We assume that creating new raw data as a byproduct of economic activity increases the data stock, which depreciates at a constant rate. These assumptions extend the standard model with two equations that represent the equilibrium price and quantity of data. We then derive the data-economy version of the NKPC, as data is critical to firms' pricing decisions, improving demand forecasts and including data acquisition costs as part of marginal costs.

The data reduce the indeterminacy region associated with the response of monetary policy to inflation and output in the standard Taylor rule. While a more-than-unitary response of the interest rate to inflation is needed for strict inflation-targeting monetary policy to raise the interest rate sufficiently to avoid self-fulfilling inflation expectations and indeterminacy (i.e., applying the Taylor principle in Woodford, 2001), the stronger response of output that mitigates the inflation dynamics in the data economy allows the Taylor rule to achieve determinacy with a weaker policy response to inflation, thereby reducing the indeterminacy region associated with monetary policy.

Data play a critical role in implementing optimal monetary policy. We study the optimal

policy by analytically deriving the welfare loss function using a second-order Taylor expansion around the constrained-efficient steady state, following Woodford (2003) and Galí (2015). We find that welfare in a data economy includes two additional countervailing components in the standard welfare function. On the one hand, data increases welfare by improving forecast accuracy, enhancing productivity, and enabling allocative efficiency. On the other hand, data decreases welfare because it requires larger labor inputs to sustain the output gains associated with the dynamic and feedback mechanisms arising from data.

We use the welfare function to analytically derive the optimal commitment (Ramsey) policy in the presence of a cost-push shock to inflation, directly linking our analysis to the standard New Keynesian model. The data feedback loop exacerbates the inflation-output tradeoff the central bank faces during supply-side disturbances. We show that the optimal policy should internalize the dynamics of both data stock and data value, optimizing their paths according to their marginal contributions to social welfare.

Finally, we study implementable interest-rate Taylor-type rules that replicate the optimal policy. Based on our results, we allow the standard Taylor rule targeting output and inflation to include, as an additional target, the value of data, given by the market price of data and the unit cost of data generation. We numerically study combinations of coefficients for inflation, the output gap, and data value that approach the optimal policy. The quantitative results indicate that the augmented Taylor rule implies a welfare loss of 0.1210% of steady-state consumption, compared with a consumption-equivalent welfare loss of 0.1196% under the Ramsey optimal policy. We show that including the data value with a positive coefficient in the augmented Taylor rule is critical to achieving welfare levels that approach the optimal rule. The intuition is as follows. Endogenous data generation amplifies the impact of cost-push shocks, reflected in the positive comovement between the data value and aggregate output. Therefore, positively targeting the data value in the monetary policy rule could effectively stabilize fluctuations in the data economy in response to supply-side disturbances.

Related Literature. Our analysis builds on the literature linking data to macroeconomics outlined in Veldkamp and Chung (2024) and the textbook Baley and Veldkamp (2025). We focus on data as a byproduct of economic activity that informs the firm's forecast and helps maximize profits, following Farboodi et al. (2019) and Farboodi and Veldkamp (2026). We are close to Groh et al. (2026), who study business cycle fluctuations related to monetary policy when data enters investment decisions and increases productivity. Our mechanism relates to models in which economic activity generates information, and uncertainty is endogenous to the business cycle (Van Nieuwerburgh and Veldkamp, 2006; Fajgelbaum et al., 2017; Argente and Yeh, 2022; Fernández-Villaverde et al., 2026). Relative to this literature, data in our model is a storable, tradable asset with a market price, allowing monetary policy to affect firms' incentives to generate information and making the value of data an observable policy target. Eeckhout

and Veldkamp (2022) show that data induces firms to lower markups to expand sales; the same force operates in our model and is the channel through which the value of data shifts aggregate supply. Asriyan and Kohlhas (2024) document evidence of a positive relationship between firm size and forecast accuracy and examine its implications for long-term allocation efficiency. Corhay et al. (2023) study the implications of data technology for firm dynamics and asset prices. Unlike these studies, we examine the role of data, as a byproduct of economic activity, in influencing the transmission and optimal conduct of monetary policy.

Our paper also relates to the broader literature on optimal monetary policy. Our model nests the workhorse New Keynesian framework of Woodford (2003) and Galí (2015), reproducing and enriching the standard three-equation structure (i.e., the IS curve, the New Keynesian Phillips Curve, and the Taylor rule) commonly used to study optimal policy. We maintain an equally simple framework and derive the analytical solution for the welfare function and the optimal commitment (Ramsey) policy in the presence of data.[2] Our results on optimal monetary policy are closely related to research focused on optimal policy in the presence of information frictions driven by exogenous noisy information (Ball et al., 2005; Adam, 2007; Ou et al., 2021), endogenous attention allocation (Paciello and Wiederholt, 2014), or imperfect information about production inputs (Angeletos and La'O, 2020). We complement those studies by focusing on optimal monetary policy when data, produced as a byproduct of economic activity, improves forecasts and mitigates uncertainty from noisy information.

The remainder of the paper is organized as follows. Section 2 presents a simple model to illustrate the monetary policy transmission mechanism in a data economy. Section 3 documents empirical evidence. Section 4 extends the simple model to an infinite-horizon New Keynesian dynamic general equilibrium model. Section 5 studies the transmission of monetary policy shocks. Section 6 examines the optimal monetary policy design. Section 7 concludes.

2 An Illustrative Two-period Model

In this section, we develop a simple two-period model to transparently illustrate the nexus between monetary policy and data. The economy is populated by a representative household, a continuum of monopolistically competitive firms indexed by $j \in [0, 1]$, and a central bank. Firms hire labor to produce differentiated goods and face information frictions, as they observe only a noisy signal of the demand for their product and use data to sharpen their inference. Economic activity generates data as a byproduct, which improves the firm's forecast of demand

[2]Our findings are also directly comparable to several studies that adopt the approach of Woodford (2003) to optimal monetary policy in different environments, including small open economies (Galí and Monacelli, 2005 and Soffritti and Zanetti, 2008), input-output networks (Tahbaz-Salehi and La'O, 2022, Qiu et al., 2026, and Rubbo, 2023), heterogeneous households or firms (Aoki, 2001; Nuño and Thomas, 2022; Cúrdia and Woodford, 2011 and Nisticò and Seccareccia, 2026), and financial frictions (Andrés et al., 2013 and Furlanetto et al., 2021).

and thereby alleviates information frictions. Nominal rigidity takes the form of a sticky nominal wage: the wage is preset and cannot adjust within the model's horizon, which is what gives monetary policy its real effects. The central bank conducts monetary policy by controlling the nominal interest rate. There is no aggregate uncertainty in this economy: households and firms perfectly foresee the aggregate states, and firms only face uncertainty regarding their idiosyncratic product demand.

2.1 Decentralized Decisions

We begin by describing the decentralized decisions of households, firms, and the monetary authority, which we then aggregate to derive the general equilibrium and derive the key intuitions in the next section.

Representative household. The representative household lives for two periods, $t \in \{1, 2\}$. The household chooses the consumption bundle $\{C_1, C_2\}$, the nominal bond holding B, equity holdings X_{j1} in each firm $j \in [0, 1]$ in period $t = 1$, and the period-2 labor supply L_2 to maximize lifetime utility:[3]

$$U(C_1, L_1) + \beta \mathbb{E}_1 U(C_2, L_2),$$

where $\beta \in (0, 1)$ is the discount factor and $U(C_t, L_t) = \frac{C_t^{1-\gamma}}{1-\gamma} - \chi L_t$ is the period utility, with $\gamma > 0$ the coefficient of relative risk aversion and $\chi > 0$ the weight on the disutility of labor. The consumption bundle C_t aggregates the differentiated goods C_{jt} produced by the continuum of firms $j \in [0, 1]$ through the CES aggregator

$$C_t = \left(\int_0^1 \epsilon_{jt}^{\frac{1}{\eta}} C_{jt}^{\frac{\eta-1}{\eta}} \mathrm{d}j \right)^{\frac{\eta}{\eta-1}},$$

where $\eta > 1$ is the elasticity of substitution across goods and ϵ_{jt} is an idiosyncratic demand shifter specific to good j. The associated price index is $P_t = \left(\int_0^1 \epsilon_{jt} P_{jt}^{1-\eta} \mathrm{d}j \right)^{\frac{1}{1-\eta}}$, where P_{jt} is the price of good j, and total expenditure satisfies $P_t C_t = \int_0^1 P_{jt} C_{jt} \mathrm{d}j$.

The household is endowed with one unit of each firm's equity ($X_{j0} = 1$). Its budget constraints in periods 1 and 2 are

$$\begin{aligned} \int_0^1 P_{j1} C_{j1} \mathrm{d}j + \int_0^1 \left(V_j - \Pi_{j1} \right) X_{j1} \mathrm{d}j + B &\leq W_1 L_1 + \int_0^1 V_j X_{j0} \mathrm{d}j, \\ \int_0^1 P_{j2} C_{j2} \mathrm{d}j &\leq W_2 L_2 + \int_0^1 \Pi_{j2} X_{j1} \mathrm{d}j + (1 + i) B, \end{aligned}$$

[3] In period 1, the household stands ready to supply whatever labor L_1 firms demand at the sticky wage $\bar{W}$.

where V_j is firm j's nominal value (stock price) in period 1, Π_{jt} denotes the nominal profit (dividend) distributed by firm j in period t, and B is the holding of nominal bonds.

The household's optimal allocation of expenditure across goods implies the demand function for each good $j \in [0,1]$:

$$C_{jt} = \epsilon_{jt} \left(\frac{P_{jt}}{P_t} \right)^{-\eta} C_t, \tag{1}$$

in which the idiosyncratic shock ϵ_{jt} acts as a demand shifter.

Nominal rigidity. The nominal wages are sticky: they are preset at $W_1 = W_2 = \bar{W}$. In period 1, the household labor supply is elastic at that wage; in period 2, the household supplies labor at the optimal condition:

$$C_2^{-\gamma} \frac{\bar{W}}{P_2} = \chi, \tag{2}$$

where the marginal utility gain from labor equals its marginal utility loss.

The household optimal bond holding decision yields the following Euler equation:

$$1 = \beta(1+i) \frac{C_2^{-\gamma}}{C_1^{-\gamma}} \frac{P_1}{P_2}. \tag{3}$$

Finally, the household's optimal equity choice yields the asset-pricing equation for the nominal firm value,

$$V_j = \Pi_{j1} + \beta \frac{C_2^{-\gamma}}{C_1^{-\gamma}} \frac{P_1}{P_2} \mathbb{E}[\Pi_{j2}], \tag{4}$$

where $\mathbb{E}[\Pi_{j2}]$ is the expected period-2 nominal profit, reflecting the assumption that the uncertainty only relates to idiosyncratic firm product demand ϵ_{jt}, and $\beta \frac{C_2^{-\gamma}}{C_1^{-\gamma}} \frac{P_1}{P_2}$ is the stochastic discount factor that firms use to value period-2 nominal profits. Combining (3) and (4) yields

$$V_j = \Pi_{j1} + \frac{\mathbb{E}[\Pi_{j2}]}{1+i}. \tag{5}$$

Monetary authority. A central bank conducts monetary policy by controlling the nominal interest rate i.

Producers. The production sector is central to the endogenous generation and the informational role of data. It consists of a continuum of monopolistically competitive producers indexed by $j \in [0,1]$. Each producer employs labor input, l_{jt}, to manufacture its differentiated good according to the linear production function $y_{jt} = l_{jt}$, hiring labor at the preset nominal wage $\bar{W}$. Each producer sells directly to households and therefore faces the demand curve (1),

which, using goods-market clearing $C_{jt} = y_{jt}$ and $C_t = Y_t$, can be written as the inverse demand $P_{jt} = \epsilon_{jt}^{1/\eta}(y_{jt}/Y_t)^{-1/\eta}P_t$. Each producer faces uncertainty about its product demand. To keep the model simple, we assume uncertainty on individual demand arises only in period 2.

Timeline. Figure 1 shows the timeline of events. In period 1, there is no uncertainty on individual product demand, and the idiosyncratic demand shock ϵ_{jt} in period 1 is equal to 1 for all $j \in [0,1]$. For each producer j, the data is generated as a by-product of production:

$$d_j = \theta y_{j1}, \tag{6}$$

where θ encapsulates the digitization level, i.e., the amount of data per unit of output that a firm generates.

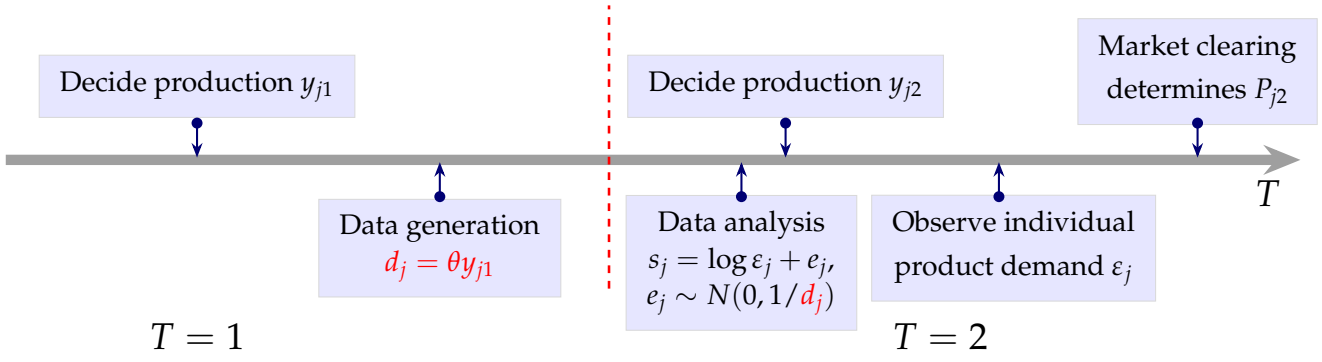


Figure 1: Timeline of the illustrative two-period model

In period 2, each producer j faces information frictions regarding demand. At the beginning of period 2, the idiosyncratic demand shock is drawn for each producer j. To ease notation, we omit the time subscript of the shock for brevity. The demand shock ϵ_j is independently and identically distributed, following a log-normal distribution $\log \epsilon_j \sim \mathcal{N}\left(-\frac{1}{2}\frac{1}{\tau_\epsilon}, \frac{1}{\tau_\epsilon}\right)$ with unit mean and variance of $\exp\left(\frac{1}{\tau_\epsilon}\right) - 1$. The distribution of ϵ_j is known to each producer, but they do not observe the exact realization of demand when setting the production plan. At the beginning of period 2, each producer j obtains a noisy signal s_j of the demand shock ϵ_j, whose precision increases with the amount of data. More formally, the signal is equal to:

$$s_j = \log \epsilon_j + e_j, \tag{7}$$

where the noise e_j is independent of the true demand ϵ_j and follows a Normal distribution with mean zero. The precision of the signal, defined as the inverse of the variance of the noise e_j, increases with the amount of data d_j, so data improve the quality of the signal the firm receives. Without loss of generality, as in Farboodi et al. (2019) and Farboodi and Veldkamp (2026), we assume that the variance of e_j is equal to $\frac{1}{d_j}$, such that $e_j \sim \mathbb{N}(0, 1/d_j)$. Based on

the received signal s_j, firm j optimizes its production plan for period 2, y_{j2}. Subsequently, the market-clearing condition determines the product price P_{j2}, after which the firm observes the true demand ϵ_j.

We solve the firm's optimal production by backward induction, starting with period 2 decisions.

Optimal decision in period 2. At the beginning of period 2, the stock of data d_j is a state variable for the firm, and each firm j chooses output y_{j2} to solve the profit maximization problem:

$$\Pi_{j2} = \max_{\{y_{j2}\}} \mathbb{E}\left[P_{j2}y_{j2} - \bar{W}y_{j2} \mid s_j\right],$$

subject to the demand function in equation (1) and the signal in equation (7), where the expectation operator $\mathbb{E}[\cdot|s_j]$ is conditional on signal s_j.[4] The next proposition summarizes the optimal decisions for the firm in period 2.

Proposition 1 (Optimal decision of the producing firm in period 2)**.** *In period 2, each firm j has optimal production equal to:*

$$y_{j2}(s_j) = \left(\frac{\eta-1}{\eta}\right)^{\eta} \mathbb{E}\left[\epsilon_j^{\frac{1}{\eta}} \mid s_j\right]^{\eta} \left(\frac{P_2}{\bar{W}}\right)^{\eta} Y_2, \tag{8}$$

and the expected real profit conditional on the signal s_j is equal to:

$$\Pi_{j2}(s_j) = \frac{1}{\eta-1}\left(\frac{\eta-1}{\eta}\right)^{\eta} \mathbb{E}\left[\epsilon_j^{\frac{1}{\eta}} \mid s_j\right]^{\eta} P_2^{\eta}\bar{W}^{1-\eta}Y_2, \tag{9}$$

where

$$\mathbb{E}[\epsilon_j^{\frac{1}{\eta}} \mid s_j]^{\eta} = \mathbb{E}\left[\epsilon_j \mid s_j\right] \exp\left(-\frac{1}{2}\frac{\eta-1}{\eta}\frac{1}{\tau_\epsilon + d_j}\right). \tag{10}$$

Proposition 1 shows that the firm's production decision and expected profit depend on the uncertainty of demand under imperfect information, as captured by the term $\mathbb{E}[\epsilon_j^{\frac{1}{\eta}} \mid s_j]^{\eta}$. Equation (10) shows that the conditional expectations of demand given the signal ($\mathbb{E}[\epsilon_j^{\frac{1}{\eta}} \mid s_j]^{\eta}$) comprise two parts. First, the conditional expectation of the demand shock after observing signal s_j, represented by $\mathbb{E}\left[\epsilon_j \mid s_j\right]$. Second, the inefficiency induced by the demand uncertainty, encapsulated by $\exp\left(-\frac{1}{2}\frac{\eta-1}{\eta}\frac{1}{\tau_\epsilon+d_j}\right) < 1$.

[4]For instance, under the log-normal distribution setup, the conditional expectation of ϵ_j is given by $\mathbb{E}[\epsilon_j|s_j] = \exp\left(\frac{d_j}{\tau_\epsilon+d_j}s_j\right)$. The ratio $\frac{d_j}{\tau_\epsilon+d_j}$ captures the weight on the signal in the conditional expectation. It increases with the data stock d_j. A higher data stock improves the precision of the signal s_j regarding the unobservable demand ϵ_j, implying a higher weight on the signal in the conditional expectation. Without any data, $d_j = 0$, the firm has no extra information about the demand for its product, so $\mathbb{E}[\epsilon_j \mid s_j]$ converges to the unconditional mean, i.e., $\mathbb{E}[\epsilon_j|s_j] = \mathbb{E}[\epsilon_j] = 1$.

As previously discussed, the data improves the precision of the signal and reduces uncertainty on individual product demand. As d_j increases, $\mathbb{E}\left[\epsilon_j \mid s_j\right]$ becomes, on average, closer to the true value of ϵ_j, thereby improving the ex-ante precision of production decisions. A higher d_j also increases the expected profit by improving the efficiency from the increased accuracy in the firm's forecast, captured by the term $\exp(-\frac{1}{2}\frac{\eta-1}{\eta}\frac{1}{\tau_\epsilon+d_j})$.[5]

Optimal decision in period 1. At the beginning of period 1, each firm j chooses output y_{j1} to solve the profit maximization problem:

$$\max_{y_{j1}} P_{j1}y_{j1} - \bar{W}y_{j1} + \frac{\mathbb{E}\left[\Pi_{j2}(s_j)\right]}{1+i},$$

subject to the demand function in equation (1) and data generation as a by-product of production in equation (6). The profit in period 2, based on signal s_j (i.e., $\Pi_{j2}(s_j)$), is obtained from the firm's optimal decisions in period 2, as given in equation (9). The expectation operator, $\mathbb{E}$, is taken on all the possible realizations of the signal s_j received in period 2.

From equation (9), the expected nominal profit in period 2 can be expressed as:

$$\mathbb{E}[\Pi_{j2}(s_j)] = \frac{1}{\eta-1}\left(\frac{\eta-1}{\eta}\right)^{\eta} P_2^{\eta}\bar{W}^{1-\eta}Y_2 \exp\left(-\frac{1}{2}\frac{\eta-1}{\eta}\frac{1}{\tau_\epsilon+d_j}\right). \tag{11}$$

It is straightforward to see that $\mathbb{E}[\Pi_{j2}(s_j)]$ increases with the data d_j. More data, by yielding more precise information, reduces the firm's future decision errors and improves ex-ante expected profit. Thus, the marginal benefit of accumulating data in period 1 is the increase in expected profit in period 2. This feedback loop incentivizes the firm to expand production, generating more data and improving the accuracy of its demand expectations. The next proposition summarizes the optimal decisions for the firm in period 1.

Proposition 2 (Optimal decision of the producing firm in period 1)**.** *In period 1, each firm j has optimal production equal to:*

$$y_{j1} = y_1 = \left(\frac{\eta-1}{\eta}\right)^{\eta}\left(\frac{P_1}{\bar{W}}\right)^{\eta} Y_1\left(1-Q\right)^{-\eta}, \tag{12}$$

and the optimal price equal to:

$$P_{j1} = \frac{\eta}{\eta-1}\left(1-Q\right)\bar{W}, \tag{13}$$

[5] In the case without information frictions where $d_j \to +\infty$ (i.e., perfect signal precision), the conditional expectation term $\mathbb{E}[\epsilon_j^{\frac{1}{\eta}} \mid s_j]^{\eta}$ equals ϵ_{jt} because $\mathbb{E}\left[\epsilon_j \mid s_j\right]$ in (10) approaches the true value of ϵ_{jt} and the information inefficiency term in equation (10) converges to 1.

where Q is the value of data, given by the net benefit of generating one unit of data, equal to:

$$Q \equiv \frac{1}{1+i}\frac{\partial \mathbb{E}\left[\Pi_{j2}(s_j)\right]}{\partial d_j} \Big/ \left(\frac{1}{\theta}\bar{W}\right) = \theta \frac{1}{\bar{W}}\frac{1}{1+i}\frac{\eta-1}{2\eta}\frac{\mathbb{E}\left[\Pi_{j2}(s_j)\right]}{(\tau_\epsilon + d_j)^2}, \tag{14}$$

where $\mathbb{E}_1\left[\Pi_{j2}(s_j)\right]$ is described by equation (11).

Proposition 2 characterizes the impact of the data on the production and prices of the producers in period 1. As shown in equation (14), the value of data is defined as the marginal benefit of generating one additional unit of data over its marginal cost, where the marginal benefit is the expected additional profit that the data generate in period two, discounted to period one ($\frac{1}{1+i}\frac{\partial \mathbb{E}\left[\Pi_{j2}(s_j)\right]}{\partial d_j}$), while the marginal cost is the additional labor needed to generate the data ($1/\theta$), multiplied by the nominal wage ($\bar{W}$). Equation (12) shows that the production in period 1 increases with the value of data, since the firm increases its production to generate more data when the data become more valuable.

Equation (13) shows that price in period one decreases with the value of data. Since higher economic activity generates data that enhances firms' forecasts, firms optimally set a lower price than the monopolistic level to increase sales and output, generate more data, and achieve greater forecasting precision. This mechanism is reflected in the lower markup over marginal cost in the data economy, captured by the term $\frac{\eta}{\eta-1}(1-Q)$ in equation (13).

2.2 Aggregation and Equilibrium

We obtain the aggregate equilibrium of the economy by aggregating the optimal decisions of the agents. We discuss each period backwardly, beginning with period 2.

Period 2. In period 2, goods-market clearing gives $C_2 = Y_2$. Using the optimal output of each firm j from equation (8) in Proposition 1, we obtain aggregate output in period 2:[6]

$$Y_2 = \left(\frac{\eta-1}{\eta}\right)^\eta \left(\frac{P_2}{\bar{W}}\right)^\eta Y_2 \left(\mathbb{E}\left[\mathbb{E}\left(\epsilon_j^{\frac{1}{\eta}} \mid s_j\right)^\eta\right]\right)^{\frac{\eta}{\eta-1}}. \tag{15}$$

From equation (10) and the definition of $\mathbb{E}\left[\epsilon_j \mid s_j\right]$, the term $\mathbb{E}[\epsilon_j^{\frac{1}{\eta}} \mid s_j]^\eta$ in equation (15) equals $\exp\left(\frac{d_j}{\tau_\epsilon + d_j}s_j - \frac{1}{2}\frac{\eta-1}{\eta}\frac{1}{\tau_\epsilon+d_j}\right)$.

Aggregate labor in period 2 is given by:

$$L_2 = \int l_{j2}\mathrm{d}j = \int y_{j2}(s_j)\mathrm{d}j = \left(\frac{\eta-1}{\eta}\right)^\eta \left(\frac{P_2}{\bar{W}}\right)^\eta Y_2 \mathbb{E}\left[\mathbb{E}(\epsilon_j^{\frac{1}{\eta}}|s_j)^\eta\right], \tag{16}$$

[6] We use the law of iterated expectations to aggregate individual output y_{j2}.

where the last equality follows from the law of iterated expectations.

Using equations (15) and (28), we express the aggregate output in period 2 as:

$$Y_2 = ZL_2, \tag{17}$$

where Z is the aggregate total factor productivity (TFP) equal to:

$$Z(D) = \left\{ \mathbb{E}\left[\mathbb{E}(\epsilon_j^{\frac{1}{\eta}} | s_j)^{\eta} \right] \right\}^{\frac{1}{\eta-1}} = \exp\left(-\frac{1}{2\eta} \frac{1}{\tau_\epsilon + D} \right). \tag{18}$$

The aggregate TFP (Z) in our model reflects the impact of information frictions on the aggregate output. A larger data stock (D) or greater signal precision (τ_ϵ) mitigates information frictions, thereby increasing TFP and output. Either more data or an increase in data precision enhances the firm's forecasting precision for demand in period 2, resulting in higher production.

From the labor aggregation equation (28), the production function (17), and the TFP equation (18), the period-2 price is a constant markup over the unit labor cost $\bar{W}/Z(D)$,

$$P_2 = \frac{\eta}{\eta - 1} \frac{\bar{W}}{Z(D)}. \tag{19}$$

Because the nominal wage is preset at $\bar{W}$, equation (19) is the period-2 aggregate supply curve: with a fixed wage, the price level is pinned down by the markup and by TFP, independently of output.[7] A larger data stock raises Z and shifts the supply curve downward, lowering the price.

Given the preset wage, the period-2 optimal labor decision (2) gives the downward-sloping aggregate demand curve,

$$C_2 = \left(\frac{1}{\chi} \frac{\bar{W}}{P_2} \right)^{\frac{1}{\gamma}}. \tag{20}$$

In equilibrium, by the law of large numbers and using (19), the expected period-2 nominal profit is[8]

$$\mathbb{E}_1[\Pi_{j2}(s_j)] = \frac{1}{\eta - 1} \bar{W} L_2. \tag{21}$$

[7] Compared with the period-1 price (22), the period-2 markup abstracts from the value of data Q, because data are no longer accumulated in the terminal period.

[8] By the law of large numbers, $\mathbb{E}_1[\Pi_{j2}(s_j)] = \int_0^1 \Pi_{j2}(s_j)\, \mathrm{d}j = \int_0^1 P_{j2} y_{j2} \mathrm{d}j - \int_0^1 \bar{W} l_{j2} \mathrm{d}j = P_2 Y_2 - \bar{W} L_2 = \frac{\eta}{\eta-1} \bar{W} L_2 - \bar{W} L_2 = \frac{1}{\eta-1} \bar{W} L_2$, where the third equality uses the fact that $\int_0^1 P_{j2} y_{j2} \mathrm{d}j = P_2 Y_2$ and the fourth equality uses equation (19).

Period 1. In period 1, the idiosyncratic shock is degenerate ($\epsilon_{j1} = 1$ for all j), so all producers are symmetric and set the common price in equation (13). The aggregate price level therefore equals the firms' price.

$$P_1 = \frac{\eta}{\eta - 1}(1 - Q)\bar{W}, \tag{22}$$

where Q is defined by equation (14). Because the nominal wage is preset at $\bar{W}$, equation (22) is the period-1 aggregate supply curve: the price is a markup over the unit labor cost $\bar{W}$, and the markup $\frac{\eta}{\eta-1}(1-Q)$ falls with the value of data Q. A higher Q thus shifts the aggregate supply curve downward, as firms cut their prices to sell more and generate more data.

Combining equations (2) and (3) yields the downward-sloping aggregate demand curve in period 1:

$$C_1 = \left(\frac{1}{\beta(1+i)}\frac{1}{\chi}\frac{\bar{W}}{P_1}\right)^{\frac{1}{\gamma}} \tag{23}$$

A cut in the nominal interest rate i shifts demand outward.

Since all firms are symmetric in period 1, aggregate labor equals aggregate output,

$$L_1 = \int l_{j1}\mathrm{d}j = \int y_1 \mathrm{d}j = C_1. \tag{24}$$

The supply of data. Each producing firm chooses output according to equation (12) in Proposition 2, so the same amount of data is produced across firms (i.e., $D = d_j$). Using the data-generating equation $d_j = \theta y_{j1}$ and the labor aggregation (24), the aggregate supply of data is

$$D = \theta \int y_{j1}\mathrm{d}j = \theta \int l_{j1}\mathrm{d}j = \theta C_1, \tag{25}$$

where final consumption C_1 increases with the data value Q, conditional on the monetary stance. The positive relationship between C_1 and the data stock D follows from the goods-market block (22) and (23): a higher value of data Q shifts the aggregate supply curve (22) downward, and along the downward-sloping demand curve (23), this raises consumption. Thus, we obtain an upward-sloping *data supply* curve

$$D = \theta\left[\frac{1}{\beta(1+i)}\frac{1}{1-Q}\frac{\eta-1}{\eta}\frac{1}{\chi}\right]^{\frac{1}{\gamma}}. \tag{26}$$

An interest rate cut, i.e., a reduction in i, shifts the data supply curve outward, as it directly increases firm production and data generation, given the value of data Q.

The demand for data. Following Proposition 2, the value of data Q equals the discounted marginal benefit of generating one additional unit of data, divided by its marginal cost. Substi-

tuting equation (21) and the equilibrium condition $d_j = D$ into the definition of Q in equation (14) yields

$$Q = \frac{\theta}{2\eta} \frac{1}{1+i} \frac{L_2}{(\tau_\epsilon + D)^2}, \tag{27}$$

where L_2 is the period-2 labor, which, by combining the aggregate production function (17), the aggregate supply curve (19), and the aggregate demand curve (20) in period 2, is

$$L_2 = \left(\frac{\eta - 1}{\eta} \frac{1}{\chi} \right)^{\frac{1}{\gamma}} Z(D)^{\frac{1}{\gamma} - 1}. \tag{28}$$

Equation (27) links the value of data to the data stock and future economic scale. On the one hand, the value of data decreases with the data stock D through the term $1/(\tau_\epsilon + D)^2$, as the marginal contribution of data to forecast precision diminishes when information is already abundant; this is the downward-sloping *data demand* curve. On the other hand, the value of data increases with the size of the period-2 economy, measured by aggregate labor L_2. L_2 acts as a demand shifter: a larger, more active economy makes accurate forecasts more valuable. In addition, an interest cut directly increases data value as it lowers the discount rate. In summary, (26) and (27) are the supply and demand schedules of the aggregate data market.

Equilibrium system. Given the monetary policy summarized by the nominal interest rate i, the equilibrium of the two-period model is fully characterized by the endogenous variables in the goods and data markets, $\{C_1, C_2, P_1, P_2, D, Q\}$. The goods-market aggregate supply and demand conditions in both periods, equations (19), (20), (22), and (23), imply that

$$C_1 = \left[\frac{1}{\beta(1+i)} \frac{1}{1-Q} \frac{\eta - 1}{\eta} \frac{1}{\chi} \right]^{\frac{1}{\gamma}}, \tag{29}$$

$$P_1 = \frac{\eta}{\eta - 1} (1 - Q) \bar{W}, \tag{30}$$

$$C_2 = \left[\frac{\eta - 1}{\eta} \frac{1}{\chi} Z(D) \right]^{\frac{1}{\gamma}}, \tag{31}$$

$$P_2 = \frac{\eta}{\eta - 1} \frac{\bar{W}}{Z(D)}, \tag{32}$$

where $Z(D)$ follows (18), while Q and D are solved by the data supply and demand schedules, equations (26) and (27).

2.3 The Transmission of Monetary Policy

In this section, we study how the production of data, as a byproduct of economic activity that improves the accuracy of the signal on demand, influences the transmission of monetary policy to prices and output. We highlight three distinct channels of monetary policy transmission: the standard demand channel present in any NK model, and two novel channels specific to the data economy, which we refer to as the *direct* and *dynamic feedback* channels. We present our results analytically, followed by graphical illustrations.

Define $\exp(e_m) \equiv \frac{1}{\beta}\frac{1}{1+i}$. Then $e_m > 0$ represents an expansionary monetary policy shock, i.e., an interest rate cut. For illustration, we take $e_m > 0$ in the following analysis.

Standard demand channel. An expansionary monetary policy first works through a standard demand channel in increasing period-1 aggregate output, as indicated by taking the partial derivative of (29) with respect to e_m:

$$\frac{\partial \log C_1}{\partial e_m} = \frac{1}{\gamma}. \tag{33}$$

Direct channel. Since data is the by-product of firm production, an increase in period-1 output directly leads to more data accumulation, as indicated by

$$\frac{d \log D}{d e_m} = \frac{d \log C_1}{d e_m}. \tag{34}$$

This, in turn, alleviates information frictions and raises period-2 TFP, increases period-2 output and reduces period-2 price by lowering unit labor cost. To see this, we have

$$\frac{d \log C_2}{d e_m} = \frac{1}{\gamma}\varepsilon_{ZD}\frac{d \log D}{d e_m}, \qquad \frac{d \log P_2}{d e_m} = -\varepsilon_{ZD}\frac{d \log D}{d e_m}, \tag{35}$$

where $\varepsilon_{ZD} \equiv \frac{\partial \log Z}{\partial \log D} = \frac{1}{2\eta}\frac{D}{(\tau_\epsilon + D)^2}$ measures the elasticity of period-2 aggregate TFP to the data stock. The direct channel, therefore, amplifies the output response while *mitigating* the price response.

Figure 2 illustrates the direct channel. The two upper panels are the goods-market equilibria in the two periods; with sticky wages the aggregate supply curves are flat at the price set as a markup over the unit labor cost. The bottom panel is the period-1 data market. An expansionary shock shifts aggregate demand outward in period 1, raising output. The higher period-1 output shifts the data supply curve outward, raising the data stock; the larger stock raises period-2 TFP, which shifts the period-2 supply curve *downward* (top-right panel). As a result, period-2 output rises by more and the period-2 price rises by less than in an economy without data.

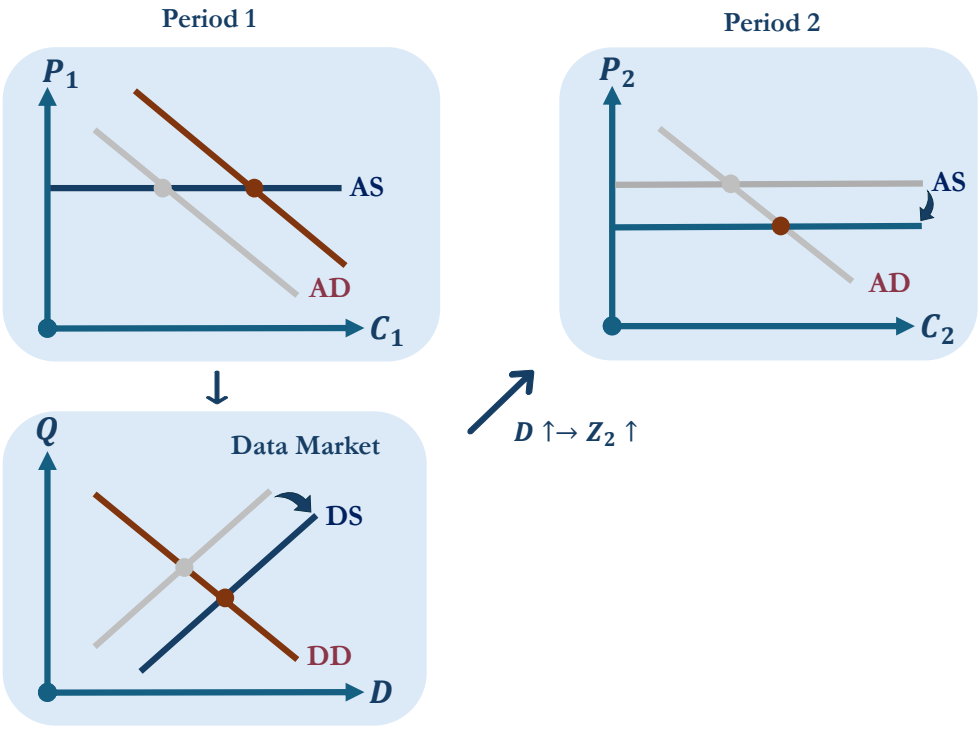


Figure 2: Monetary transmission in the data economy: direct channel

Dynamic feedback channel. A larger data stock raises period-2 TFP, $Z(D)$. Provided that period-2 employment comoves with TFP,[9] higher TFP also expands the scale of period-2 production, L_2. A larger production scale raises firms' expected profits and hence the value of data, as (27) shows, which shifts the data demand curve outward:

$$\frac{\partial \log Q}{\partial \log L_2} = 1, \quad \frac{d \log L_2}{de_m} = \left(\frac{1}{\gamma} - 1\right) \epsilon_{ZD} \frac{d \log D}{de_m}. \tag{36}$$

In addition, an expansionary monetary policy shifts the data demand curve outward directly, by lowering the nominal rate at which firms discount future profits. Since data supply also shifts outward, the data stock unambiguously increases, and the response of the value of data can be decomposed as

$$\frac{d \log Q}{de_m} = \underbrace{\frac{\partial \log Q}{\partial e_m}}_{\text{(i)}} + \underbrace{\frac{\partial \log Q}{\partial \log L_2} \frac{d \log L_2}{de_m}}_{\text{(ii)}} - \underbrace{\frac{2D}{\tau_\epsilon + D} \frac{d \log D}{de_m}}_{\text{(iii)}}. \tag{37}$$

Term (i), which equals one, is the exogenous outward shift of data demand caused by the discount-rate effect of monetary policy. Term (ii) is the endogenous outward shift of data demand arising from the larger future production scale. Term (iii) captures the movement along

[9] We impose the condition $\gamma < 1$, under which period-2 labor comoves with TFP by equation (28).

the downward-sloping data demand curve: data has diminishing marginal contribution in reducing uncertainty, and the strength of this effect is governed by the elasticity of data demand with respect to the data stock, $-2D/(\tau_\epsilon + D)$.

The value of data feeds back into both period-1 output and prices: a higher data value lowers the markup and hence the period-1 price, shifting the aggregate supply curve downward. We call the additional effects induced by the data value the dynamic feedback channel, as it reflects the forward looking nature of data value. The response of period-1 output and price can be decomposed as

$$\frac{d\log C_1}{de_m} = \underbrace{\frac{\partial \log C_1}{\partial e_m}}_{\text{Standard demand channel}} + \underbrace{\frac{\partial \log C_1}{\partial \log Q}\frac{d\log Q}{de_m}}_{\text{dynamic feedback}}, \tag{38}$$

where $\frac{\partial \log C_1}{\partial \log Q} = \frac{1}{\gamma}\frac{Q}{1-Q}$ by (29), and

$$\frac{d\log P_1}{de_m} = \underbrace{-\frac{Q}{1-Q}\frac{d\log Q}{de_m}}_{\text{dynamic feedback}}. \tag{39}$$

Combining (33), (36), (37), and (38), we obtain that

$$\frac{d\log D}{de_m} = \frac{d\log C_1}{de_m} = \frac{1}{\gamma(1-Q)+\mathcal{E}\,Q}, \qquad \frac{d\log Q}{de_m} = \frac{(1-Q)\,(\gamma-\mathcal{E})}{\gamma(1-Q)+\mathcal{E}\,Q}, \tag{40}$$

where $\mathcal{E} \equiv \frac{2D}{\tau_\epsilon+D} - \left(\frac{1}{\gamma}-1\right)\varepsilon_{ZD} = \frac{2D}{\tau_\epsilon+D} - \left(\frac{1}{\gamma}-1\right)\frac{1}{2\eta}\frac{D}{(\tau_\epsilon+D)^2}$ is the net elasticity of data demand with respect to the data stock, combining the diminishing marginal contribution of data (term (iii) in (37)) with the endogenous demand shift from the future production scale (term (ii) in (37)).

Equation (40) is the heart of the mechanism. In an economy without data ($Q = 0$), (29) and (30) collapses to $C_1 = \left[\exp(e_m)\frac{1}{1-Q}\frac{\eta-1}{\eta}\frac{1}{\chi}\right]^{\frac{1}{\gamma}}$ and $P_1 = \frac{\eta}{\eta-1}\bar{W}$: a monetary expansion raises period-1 output one-for-$1/\gamma$ and leaves the price unchanged. With data, the sign of the feedback hinges on $d\log Q/de_m$, which is positive whenever $\mathcal{E} < \gamma$, which holds under a sufficient condition

$$\tau_\epsilon + D^* \leq (\frac{1}{\gamma}-1)\frac{1}{4\eta}, \tag{41}$$

where $\gamma < 1$ and D^* is the equilibrium data stock. Condition (41) holds when the posterior precision is sufficiently small, i.e., information frictions are sufficiently severe. Intuitively, when information is scarce, the expansion in period-2 activity raises the marginal value of data by more than the dilution from a larger stock, so the data-demand curve shifts out by more than

the data-supply curve, and the equilibrium data value Q rises together with the stock D.

When condition (41) holds, $d \log Q / de_m > 0$ activates the *dynamic feedback channel* in period 1. From (38), the higher data value amplifies the output response and turns the price response negative: firms cut their markup to sell more and accumulate more data, so output expands by more and the price rises by less—indeed falls—relative to the data-free benchmark. The data economy thus amplifies the output response to monetary policy while dampening the price response, in both periods.

Figure 3 illustrates the dynamic feedback channel, using the same layout as Figure 2. The direct-channel expansion of period-2 supply raises firms' data demand in period 1 (red curve in the bottom panel), so the supply and demand for data both shift outward. When information frictions are severe (condition (41)), the outward shift in data demand dominates, raising both the equilibrium data value Q and the data stock D. The higher data value induces firms to cut their markup and expand production, shifting the period-1 supply curve downward (blue curve in the top-left panel). As a result, in period 1 the dynamic feedback channel enlarges the output expansion while dampening the price response.

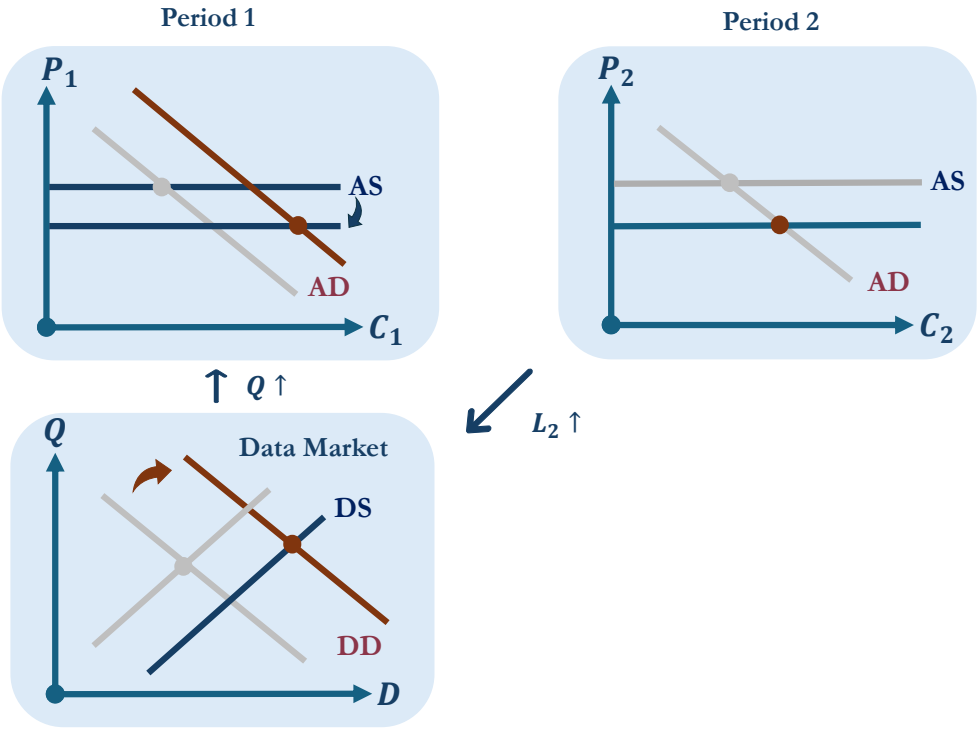


Figure 3: Monetary transmission in the data economy: dynamic feedback channel

In summary, the direct channel arises because data improve next-period production efficiency, while the dynamic feedback channel arises because data expand the next-period economy and thereby raise the marginal value of data that helps firms forecast demand. Proposition 3 summarizes the analysis.

Proposition 3 (Transmission of monetary policy in the data economy)**.** *When firms face sufficiently large information frictions, endogenous data accumulation induces a direct channel and a dynamic feedback channel through which monetary policy amplifies output responses and mitigates price responses, in both periods.*

3 Empirical Evidence

In this section, we test the theory's implications for the transmission of monetary policy in a data economy, exploiting sector-level heterogeneity in digitalization and firm-level heterogeneity in information frictions. We first describe the data and the construction of our measures of sectoral digitalization and firm-level forecast errors. We then assess the empirical relevance of the mechanism through which the data economy amplifies the response of output and dampens the response of inflation to monetary policy. To do so, we estimate the responses of output and prices to both conventional monetary policy shocks and forward guidance shocks. Forward guidance shocks allow us to test whether expectations of future monetary easing affect data-intensive firms differently. Finally, we test the premise of the theory: the additional production induced by a monetary expansion generates data, and these data reduce the forecast errors of data-intensive firms.

3.1 Data and measurements

We measure conventional monetary policy shocks using Jarociński and Karadi (2020), who isolate policy surprises from central bank information shocks, and forward-guidance shocks using Swanson (2021), who uses asset-price changes around FOMC announcements.

The theoretical effects of monetary policy hinge on the premise that data originates from economic activity and provides information that enhances firms' forecast accuracy. Consistent with our production-based data generation framework, we focus on the manufacturing sector (NAICS 31–33), where firms generate their own data rather than providing data services. We construct two samples: a firm-level sample to examine the impact of monetary policy on production and forecast accuracy, and an industry-level sample to analyze its effect on inflation.

Based on our theory, the effects of the data economy should be significant for firms with greater capability to collect and process data (i.e., higher levels of digitalization) and with a greater need for information (i.e., subject to stronger information frictions). To measure the firms' data collection and processing capabilities, we use data-scientist employment across 61 four-digit industries within the manufacturing sector from the Bureau of Labor Statistics' Occupational Employment and Wage Statistics (OEWS) program, which constructs estimates of the

stock of data scientists employed in each industry.[10] We construct the dummy variable $Digital_i$ that equals one if a four-digit industry i has a nonzero stock of data scientist employment in 2024, capturing greater reliance on data-driven decision-making and processing capabilities.[11]

We appraise the degree of information frictions at the firm level from the forecast error of each firm from the industry median. We construct firms' forecast errors from one-quarter-ahead managerial guidance on earnings per share (EPS) using Thomson Reuters Institutional Brokers' Estimate System (IBES) database.[12] For each *firm* j, we construct a dummy variable $High_InfoFric_{j,t}$ that equals one if the absolute forecast error is above the quarterly median, indicating a higher degree of information frictions faced by firm j in quarter t. Similarly, for each *industry* i, we construct a dummy variable $High_InfoFric_{i,t}$ that equals one if the average absolute forecast error within industry i is above the quarterly median across industries.

We merge the IBES with Compustat data, which provides firm-level balance sheet information, including sales, assets, profitability ratios, and leverage ratios. Our merged firm-level sample includes, on average, 130 firms per quarter, spanning from 2010:Q1 to 2023:Q4. We obtain sector-level prices from the Bureau of Labor Statistics' Producer Price Index (PPI) at the four-digit NAICS level. Our industry-level sample period includes 61 four-digit manufacturing industries, spanning from 2010:M1 to 2023:M12.[13]

3.2 Testing the mechanisms

Direct effect. We test the direct effect of a monetary policy shock using estimates of conventional monetary policy shocks constructed by Jarociński and Karadi (2020). To test the effects on firm production, we estimate the dynamic effects of monetary policy on firm-level production, which we measure using sales deflated by the 4-digit NAICS-level producer price index. We estimate a quarterly panel dataset using the local projections of Cloyne et al. (2024), with

[10] By the Standard Occupational Classification (SOC), a data scientist is defined as an occupation that "Develops and implements a set of techniques or analytics applications to transform raw data into meaningful information using data-oriented programming languages and visualization software; applies data mining, data modeling, natural language processing, and machine learning to extract and analyze information from large structured and unstructured datasets; visualizes, interprets, and reports data findings; and may create dynamic data reports."

[11] As of 2024, 26 out of 61 manufacturing industries have a nonzero stock of data scientist employment. Our premise is that industries that still employ no data scientists by 2024 are likely to have limited reliance on data or low data-processing capabilities.

[12] The forecast error is defined as the difference between one-quarter-ahead managerial guidance on EPS and its actual realization, normalized by the beginning-of-quarter stock price. Normalization by stock price mitigates scale effects, since absolute forecast errors in EPS tend to rise with the level of EPS.

[13] The IBES managerial guidance dataset begins in 1994, but we select 2010 as the starting point, coinciding with the emergence of big data technologies.

the following empirical specification:

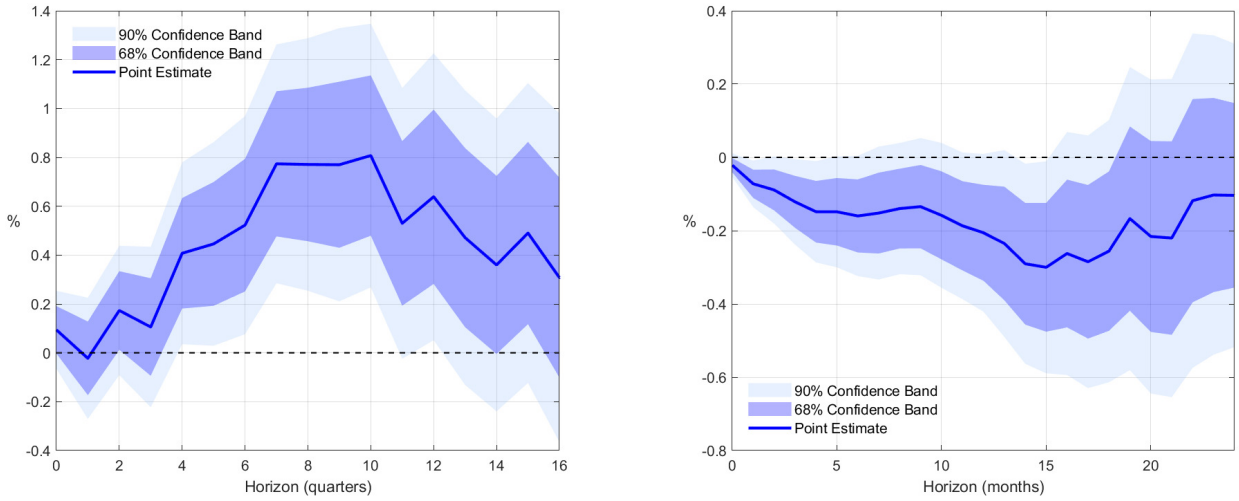
$$\begin{aligned}\Delta^k \log y_{j,t+k} &= \beta^y_{0,k} + \beta^y_{1,k}\Delta \log y_{j,t-1} + \beta^y_{2,k} MP_t \times High_Digital_j \times High_InfoFric_{j,t-1} \\ &+ \beta^y_{3,k} MP_t \times High_Digital_j + \beta^y_{4,k} MP_t \times High_InfoFric_{j,t-1} \\ &+ \beta^y_{5,k} High_Digital_j \times High_InfoFric_{j,t-1} + \beta^y_{6,k} High_InfoFric_{j,t-1} \\ &+ \gamma^y_{1,k}\Phi_{j,t-1} + \zeta_t + \mu_{j,t} + \epsilon^y_{j,t,k}, \end{aligned} \tag{42}$$

where $k = 0, 1, 2, ...$ denotes the projection horizon. The dependent variable, $\Delta^k \log y_{j,t+k} \equiv \log y_{j,t+k} - \log y_{j,t-1}$, is defined as the cumulative change in the natural logarithm of firm j's output from the pre-shock quarter $(t-1)$ to k quarters after the shock. MP_t is the measure of monetary policy shocks. The variable $High_Digital_j$ is a dummy equal to one if the NAICS 4-digit sector of the firm j has non-zero stock of data scientists in 2024, and zero otherwise, indicating the level of digitalization. $High_InfoFric_{j,t-1}$ is a dummy variable that equals one for firms with above-median absolute forecast errors in the quarter $t-1$ and zero otherwise, capturing the degree of firm-level information frictions. The regression includes controls for lagged firm output growth ($\Delta y_{j,t-1} \equiv \log y_{j,t-1} - \log y_{j,t-2}$) and lagged firm-level characteristics ($\Phi_{j,t-1}$), including firm size (logarithm of book assets), leverage ratio (debt-to-asset ratio), and profitability (return on assets, ROA).[14] ζ_t denotes time fixed effects and μ_t denotes firm-by-quarter fixed effects, included to control for firm-specific seasonal heterogeneity. The term $\epsilon^y_{j,t,k}$ denotes the regression error term.

The coefficient of interest, $\beta^y_{2,k}$, captures the marginal effect of monetary policy (MP_t) on firms characterized by both high levels of digitization ($High_Digital_j$) and high degrees of information frictions ($High_InfoFric_{j,t-1}$), that is, firms with greater capacity to process data and stronger reliance on data to overcome informational constraints. Our theory predicts that the output of firms in data-intensive industries responds more strongly to monetary policy shocks, particularly those facing higher information frictions, implying a negative coefficient $\beta^y_{2,k}$. This reflects a more pronounced output increase following an expansionary monetary policy shock (i.e., a negative MP_t).

Figure 4a reports the estimated values of $-1\% \times \beta^y_{2,k}$ for $k = 0, 1, \ldots, 16$, which captures the differential responses of output to a 1% expansionary monetary policy shock. The figure shows that an expansionary monetary policy shock generates larger output responses among firms in data-intensive industries and with greater information frictions. This finding provides empirical support for the prediction of our model that the data economy amplifies output responses

[14] Because the response of firms to monetary policy varies systematically with their balance sheets (Ottonello and Winberry, 2020; Cloyne et al., 2023), we also include the interactions of MP_t with each element of $\Phi_{j,t-1}$, so that the coefficient of interest is identified from variation in data intensity conditional on the financial characteristics that this literature shows to matter for transmission.



(a) Production (b) Price

Figure 4: Impulse responses to a 1% monetary policy expansion for highly digitalized firms with greater information frictions

to monetary policy shocks through an informational channel. For instance, for a firm with a high level of digitalization and greater information frictions (i.e., uncertainty in the previous quarter above the median), a 1% unexpected monetary loosening leads to an additional 0.77% increase in output growth at the two-year horizon ($k = 8$).[15] The effect is persistent, suggesting that the data economy generates sustained amplification of output responses to monetary policy shocks.

We then examine how the data economy influences the response of industry-level inflation to monetary policy shocks, with our theory predicting an attenuated response among digitalized firms. We estimate the dynamic effects of monetary policy on industry-level prices, measured by the Producer Price Index (PPI) at the 4-digit NAICS level from the BLS. Specifically, we estimate the following monthly panel-data local projections:

$$\begin{aligned}\Delta^k \log p_{i,t+k} &= \beta^p_{0,k} + \beta^p_{1,k}\Delta \log p_{j,t-1} + \beta_{2,k} MP_t \times High_Digital_i \times High_InfoFric_{i,t-1} \\ &+ \beta^p_{3,k} MP_t \times High_Digital_i + \beta^p_{4,k} MP_t \times High_InfoFric_{i,t-1} \\ &+ \beta^p_{5,k} High_Digital_i \times High_InfoFric_{i,t-1} + \beta^p_{6,k} High_InfoFric_{i,t-1} \\ &+ \zeta_t + \xi_{j,t} + \epsilon^p_{i,t,k}, \end{aligned} \tag{43}$$

where the dependent variable $\Delta^k \log p_{j,t+k} \equiv \log p_{j,t+k} - \log p_{j,t-1}$ is defined as the cumulative change in the natural logarithm of industry i's price level from the pre-shock month $(t-1)$ to k months after the shock. $High_Digital_j$ measures industry-level digitalization, same as in

[15] The estimated value of $\beta^y_{2,8} = -0.77$.

equation (42). $High_InfoFric_{i,t-1}$ is a dummy variable which equals one if the average absolute forecast error within industry i is above the median level across industries in month $t-1$, measuring the industry-level information frictions. The regression includes controls for lagged industrial inflation ($\Delta p_{j,t-1} \equiv \log p_{j,t-1} - \log p_{j,t-2}$), time fixed effects ($\zeta_t$), and industry-by-quarter fixed effects ($\xi_{i,t}$) to control industry-specific seasonality. The variable $\varepsilon^p_{i,t,k}$ denotes the regression error term.

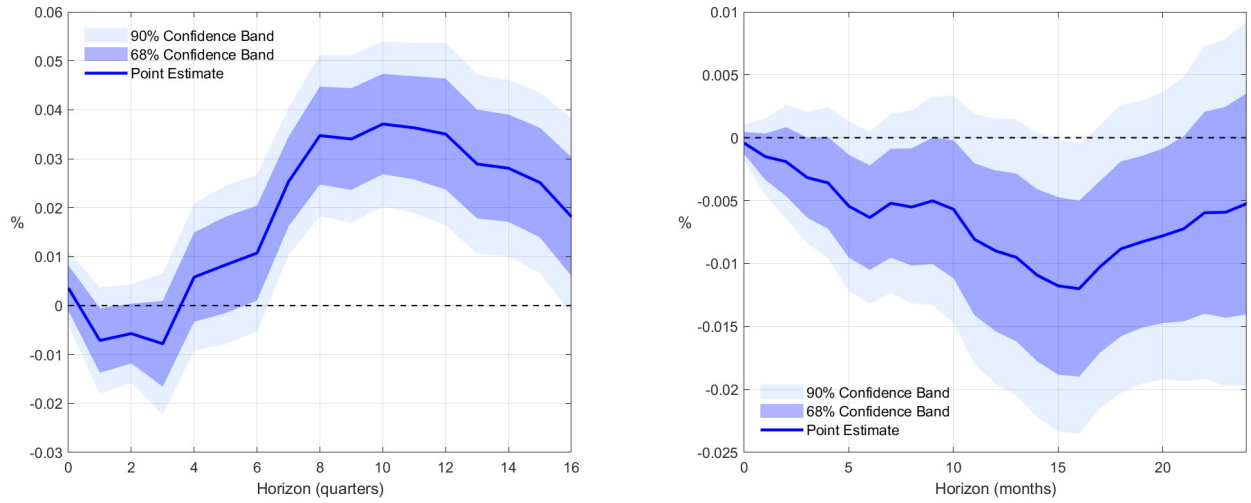
The coefficient of interest, $\beta^p_{2,k}$, captures the marginal effect of monetary policy (MP_t) on industries characterized by both high levels of digitization ($High_Digital_i$) and high degrees of information frictions ($High_InfoFric_{i,t-1}$). Our theory predicts that prices in data-intensive industries respond less strongly to monetary policy shocks, especially those firms with higher information frictions, implying a positive coefficient $\beta^p_{2,k}$, since an expansionary monetary policy (negative MP_t) entails a more muted inflation response. Figure 4b reports the estimated values of $-1\% \times \beta^p_{2,k}$ for $k = 0, 1, \ldots, 25$, which, according to equation (43), represent the impulse responses of prices to a 1% expansionary monetary policy shock for highly digitalized industries with greater information frictions. The figure shows that loosened monetary policy leads to smaller increases in the prices of industries with both high need for data as information and high ability to process data. The data economy mitigates the price responses to monetary policy shocks, consistent with our theoretical prediction. For instance, for a highly digitalized industry with above-median degree of information frictions, the effect of a 1% unexpected monetary expansion to cumulative inflation growth is mitigated by 0.206 % at the 1-year horizon (i.e., $k = 12$).[16] The effects remain significant for about 1.5 years, suggesting that the data economy generates persistent mitigation effects on the responses of inflation to monetary policy shocks.

Dynamic feedback effect. The dynamic feedback effect arises from firms adjusting current output and prices in anticipation of *future* changes in output, rather than in response to contemporaneous effects of monetary policy. Thus, we test this mechanism by examining the responses of output and inflation to forward guidance shocks —based on announcements of future policy actions— constructed by Swanson (2021). Our theory predicts that firms in more digitized industries exhibit stronger output responses and weaker price responses to forward guidance shocks, with these effects being more pronounced among firms facing greater information frictions.

We perform the analysis by replacing the conventional monetary policy shock in regressions (42) and (43) with the forward guidance shock. Figure 5a shows the dynamic effect of forward guidance on firm output. The estimated values of $\beta^y_{2,k}$ on different horizons capture the differential dynamic output responses of digitalized firms with information frictions to forward

[16] The estimated value of $\beta^p_{2,12} = -0.206$.

guidance shocks. The figure shows that the anticipated future monetary expansion leads to a larger increase in output for digitalized firms facing information frictions. This result provides empirical support for the dynamic channel of our theory: Expectations about future monetary policy easing increase output more significantly in data-intensive firms that face large information frictions. For instance, for a firm with a high digitalization level and greater information frictions (last-quarter uncertainty is above the median level), a 1% expansion in forward guidance leads to an additional 0.35% increase in production growth at the two-year horizon (i.e., $k = 8$).[17] The effects remain significant for about 4 years, suggesting that the data economy generates persistent amplification effects on the responses of production to forward guidance shocks.



(a) Production (b) Price

Figure 5: Impulse responses to a 1% expansionary forward guidance shock for highly digitalized firms with greater information frictions

Figure 5b shows the dynamic effects of forward guidance on industrial inflation for data-intensive firms. The estimates for $\beta^p_{2,k}$ on different horizons represent the differential dynamic price response of industries which are data-intensive and with greater information frictions. The figure shows that expected monetary expansion induces smaller increases in the prices of high digitalized industries with greater information frictions. When future aggregate demand is expected to be higher as a result of forward guidance and firms expand output, the response of prices diminishes due to the increase in goods supply. For instance, for a highly digitalized industry with above-median degree of information frictions, the influence of a 1% expansionary forward guidance shock on cumulative inflation growth is mitigated by 0.009%

[17] The estimated coefficient at 8-quarter horizon is -0.035.

at the 1-year horizon (i.e., $k = 12$).[18] The effects remain persistently significant, suggesting that the data economy generates persistent mitigation effects on the responses of inflation to forward guidance shocks.

3.3 Testing the link between data intensity and future forecast precision

The premise of our theory is that data is a by-product of economic activity that improves the forecast precision, allowing firms to better match demand and expand profits. We now provide empirical support for this key assumption by studying whether data accumulated through the increase in production from expansionary monetary policy reduces the forecast errors of data-intensive firms, testing this link for both conventional and forward-guidance monetary policy.

Forecast error and conventional monetary policy shocks. We study the relationship between changes in firms' forecast errors and output growth resulting from the expansionary monetary policy. Our working hypothesis is that data-intensive firms with high information frictions experience reduced forecast errors following expansionary monetary policy, as higher production enhances their forecasting precision.

We isolate the effect of monetary policy on the growth of economic activity in the quarter of the shock by using the fitted value of production growth from regression (42) at $k = 0$, i.e., $\widehat{\Delta \log y_{j,t}} = \widehat{\log y_{j,t} - \log y_{j,t-1}}$. $\widehat{\Delta \log y_{j,t}}$ captures the variation in production driven by exogenous demand-side changes related to monetary policy shocks. We estimate the dynamic effects of $\widehat{\Delta \log y_{j,t}}$ on subsequent firm-level forecast precision, measured by the absolute value of one-quarter-ahead managerial forecast errors. Specifically, we estimate the following panel-data local projection:

$$\begin{aligned}
\Delta^k \log|FE_{j,t+k}| &= \beta^f_{0,k} + \beta^f_{1,k}\Delta \log|FE|_{j,t-1} + \beta^f_{2,k}\widehat{\Delta \log y_{j,t}} \times High_Digital_j \times High_Uncer_{j,t-1} \\
&+ \beta^f_{3,k}\widehat{\Delta \log y_{j,t}} \times High_Digital_j + \beta^f_{4,k}\widehat{\Delta \log y_{j,t}} \times High_Uncer_{j,t-1} + \beta^f_{5,k}\widehat{\Delta \log y_{j,t}} \\
&+ \beta^f_{6,k} High_Digital_j \times High_Uncer_{j,t-1} + \beta^f_{7,k} High_Uncer_{j,t-1} \\
&+ \gamma^f_{1,k}\Phi_{j,t-1} + \gamma^f_{2,k}\xi_t + \gamma^f_{3,k}\mu_{j,t} + \epsilon^f_{j,t,k}, \qquad (44)
\end{aligned}$$

where $k = 1, 2, ...$ denotes the projection horizon. The dependent variable $\Delta^k \log|FE_{j,t+k}| \equiv \log|FE_{j,t+k}| - \log|FE_{j,t-1}|$ is the cumulative change in the natural logarithm of firm j's 1-quarter ahead forecast error from the pre-shock quarter $(t-1)$ to k quarters after the shock. The term $High_Digital_j$ and $High_Uncer_{j,t-1}$ are defined same as in regression (42). The regression includes controls for lagged change in the forecast precision ($\Delta|FE_{j,t-1}| \equiv \log|FE_{j,t-1}| -$

[18] The estimated value of $\beta^p_{2,12} = -0.009$.

$\log|FE_{j,t-2}|$). The other control variables ($\Phi_{j,t-1}$, ζ_t, and $\mu_{j,t}$) are the same as in regression (42). The term $\varepsilon^f_{j,t,k}$ denotes the regression error term.

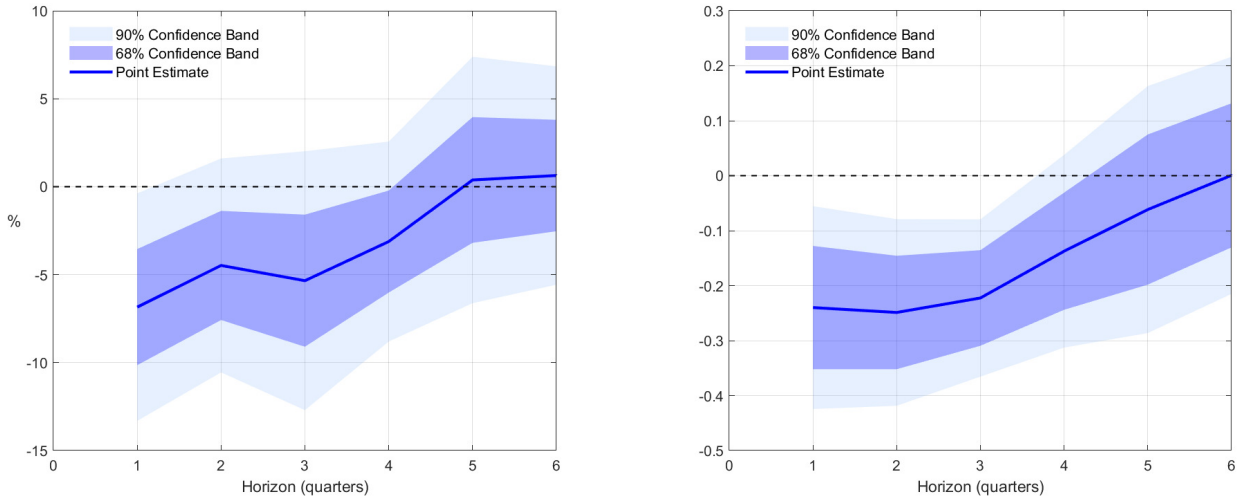
The coefficient of interest, $\beta^f_{2,k}$, captures the marginal effect of monetary-policy-induced output growth on subsequent forecast errors for firms characterized by high levels of digitization and high degrees of information frictions. The premise of our theory is that production generates data, which enhances forecasting precision, thereby implying a negative coefficient $\beta^f_{2,k}$.

Figure 6a reports the estimated values of $-1\% \times \beta^y_{2,0} \times \beta^f_{2,k}$ for $k = 1, \ldots, 6$, which represent the dynamic effects of a 1% monetary policy expansion on forecast errors through its impact on contemporaneous production. The figure shows that production growth in the period of monetary policy easing is associated with a reduction in subsequent forecast errors. For example, for a firm with a high level of digitalization and high information frictions (absolute forecast error in the previous quarter above the median), production growth induced by a 1% expansionary monetary policy shock is accompanied by a 6.85% reduction in next-quarter forecast error, with this improvement in forecast precision persisting for about one year.[19]

Forecast error and forward-guidance shocks. Analogous to the case of conventional monetary policy, we test the effects of forward guidance shocks on forecast errors by replacing $\widehat{\Delta \log y_{j,t}}$ in regression (44) with the predicted production growth in the quarter of the forward guidance shock. Figure 6b displays the estimated dynamic effects of a 1% expansionary forward guidance shock on future forecast errors through its impact on contemporaneous production. Consistent with our theory, the production growth induced by forward guidance shocks in the impact period improves future forecast precisions. For example, for a firm with a high level of digitalization and high information frictions (with uncertainty in the previous quarter above the median), a 1% expansionary forward guidance shock reduces the one-quarter-ahead forecast error by 0.24% through its impact on contemporaneous production.[20] This effect persists for about 1 year, consistent with the estimates in Figure 6a.

[19] The estimated coefficients for the conventional monetary policy are $\beta^f_{2,1} = -72.86$ and $\beta^y_{2,0} = -0.094$. Taken together, these imply that a 1% expansionary monetary policy shock reduces the next-period forecast error by approximately 6.85% ($-1\% \times -72.86 \times -0.094$) through its effect on impact-period production.

[20] The estimated coefficients for forward guidance are $\beta^f_{2,1} = -67.40$ and $\beta^y_{2,0} = -0.0036$. Taken together, these imply that a 1% expansionary forward guidance shock reduces the next-period forecast error by approximately 0.24% ($-1\% \times -67.40 \times -0.0036$) through its effect on impact-period production. The effect of forward guidance is weaker than that of conventional monetary policy shocks, consistent with the "forward guidance puzzle" in the literature, which highlights the relatively muted impact of forward guidance compared with contemporaneous monetary policy.



(a) Conventional monetary policy

(b) Forward guidance shock

Figure 6: Dynamic effects of 1% expansionary monetary policy shock on subsequent forecast errors for highly digitalized firms with greater information frictions

4 A Full-Fledged Data-Economy New Keynesian Model

This section extends the simple two-period model to an infinite horizon model. The dynamic general equilibrium framework enables us to study a novel New Keynesian Phillips Curve (NKPC) for the data economy and analyze the dynamic propagation of monetary policy and the changes in the stability of the system. In the next section, we use the extended framework to analytically derive the optimal monetary policy and study implementable interest rate rules.

The basic setup of the data-economy DSGE model is similar to that in the two-period model. One distinction is that it explicitly introduces a data market and data prices. In particular, we distinguish between raw data and structured data. The raw data is generated as a by-product of the firm's production activities, as in the two-period model. Firms cannot utilize the raw data directly to analyze their product demand. In each period, firms sell their raw data to a representative data intermediary at a price of Q_t^r. The data intermediary transforms the raw data collected from the firms into structured data and sells it back to firms at a price of Q_t^d. The firm can utilize the structured data processed by the data intermediary to directly analyze the market demand, reducing the uncertainty it faces.[21] In addition, different from the two-period model, we introduce nominal rigidity by the standard Calvo-type of price stickiness and assume that monetary authority implements a Taylor rule, where the nominal interest rate targets the output gap and the inflation rate.

[21] Technically, the introduction of the data intermediary helps degenerate the distribution of data across firms at the end of each period. Such simplification allows us to derive a tractable linearized data economy New Keynesian dynamic system and a transparent quadratic welfare loss function.

The model economy comprises households, retailers, wholesalers, intermediate goods producers, and data intermediaries. Households supply labor to the intermediate goods producers and purchase a continuum of retail goods from retailers to form a final consumption bundle. Retailers resemble wholesale goods into retail goods. The price setting of retail goods follows the Calvo-type of nominal rigidities. The wholesalers combine a continuum of intermediate goods to produce wholesale goods. The intermediate goods producers employ labor to produce and face demand uncertainty. Their production activities automatically generate raw data. The data intermediaries collect raw data from these firms, transform it into structured data, and sell it to the intermediate goods producers. Figure 7 gives a graphic elaboration of the model's framework.

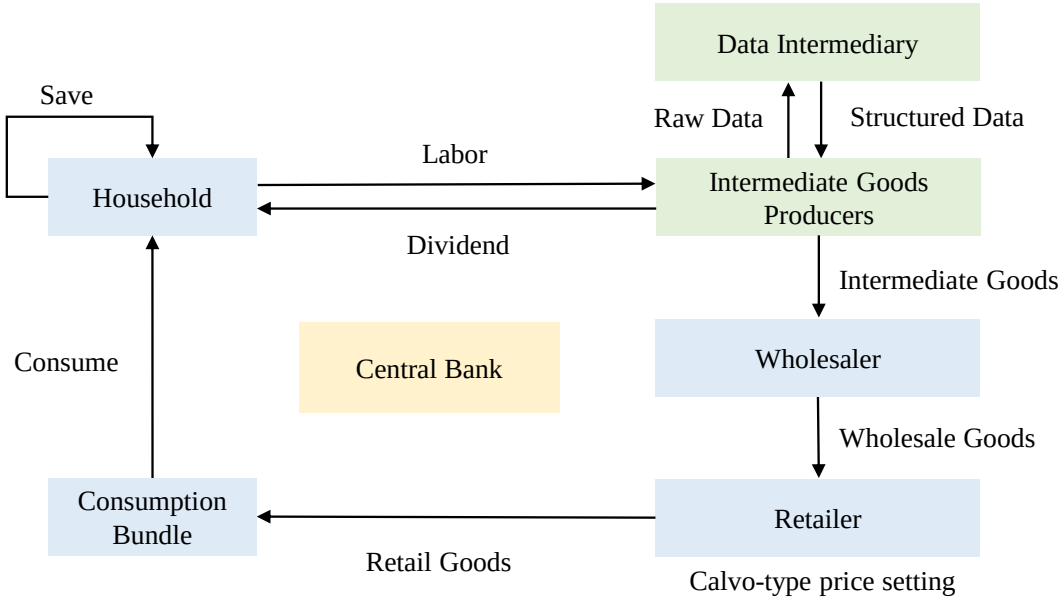


Figure 7: Framework of the infinite-horizon NK model

4.1 Households

We first briefly introduce the household's problem. The representative household derives utility from final consumption C_t and labor L_t. The life-time utility is $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\gamma}}{1-\gamma} - \chi L_t \right]$, where $\beta \in (0,1)$ is the discount factor, $\gamma > 0$ is the coefficient of relative risk aversion, and χ measures labor disutility. Unlike the two-period model, we do not explicitly model the money demand for the household problem. The monetary policy is introduced by an interest rate rule. Analogous to the two-period model, the final consumption is a CES aggregate of a continuum of differentiated retail goods $C_t = \left(\int_0^1 C_{it}^{\frac{\sigma-1}{\sigma}} \mathrm{d}i \right)^{\frac{\sigma}{\sigma-1}}$, where C_{it} is the consumption of retail goods $i \in [0,1]$ and $\sigma > 1$ is the elasticity of substitution among retail goods. The budget constraint

follows

$$\int_0^1 \tilde{P}_{it} C_{it} \mathrm{d}i + \frac{B_{t+1}}{R_t} + \int_{\iota \in \mathcal{I}} \left(V_{\iota t} - \Pi_{\iota t} \right) X_{\iota t} \mathrm{d}\iota \leq W_t L_t + \int_{\iota \in \mathcal{I}} V_{\iota t} X_{\iota t-1} \mathrm{d}\iota + B_t - T_t, \tag{45}$$

where $\tilde{P}_{it}$ is the retail price of variety i, W_t is the nominal wage rate, R_t is the risk-free nominal interest rate, $V_{\iota t}$ and $\Pi_{\iota t}$ denote the stock value and dividend payment of firm $\iota \in \mathcal{I}$ at period t, where $\mathcal{I}$ includes both retailers and intermediate goods producers. T_t is the government transfer, B_{t+1} is bond holding, and $X_{\iota t}$ is the equity holding of firm ι. Household chooses C_{it}, B_{t+1} and $X_{\iota t+1}$ to maximize her life-time utility subject to the budget constraint (45).

Let $\Lambda_t = C_t^{-\gamma}$ denote the marginal utility of consumption in period t. The optimal condition for the bond demand implies

$$\frac{1}{R_t} = \beta \mathbb{E}_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \frac{1}{\Pi_{t+1}} \right], \tag{46}$$

where $\Pi_{t+1} \equiv \tilde{P}_{t+1} / \tilde{P}_t$ is the price inflation, and $\tilde{P}_t$ is the aggregate price level for the final consumption bundle in period t. The optimal labor supply decision is

$$\Lambda_t \frac{W_t}{\tilde{P}_t} = \chi, \tag{47}$$

where $\frac{W_t}{\tilde{P}_t}$ is the real wage.

4.2 New Keynesian sector

A continuum of retailers indexed by $i \in [0,1]$ sells consumption goods to the household. A typical retailer i purchases wholesale goods C_{it} at the price P_t and transforms them into the same amount of retail goods as in the two-period model. In each period, retailers face Calvo-type price stickiness: with probability $1-\omega$, they can set prices, and with probability ω, prices are fixed at the previous period's level. A retailer i that can adjust its price in period t chooses the optimal price $\tilde{P}_{it}^*$ to maximize the discounted sum of expected profit $\mathbb{E}_t \sum_{\kappa=0}^{\infty} (\beta\omega)^\kappa \frac{\Lambda_{t+\kappa}}{\Lambda_t} \left[(1+\mu_r) \frac{\tilde{P}_{it}^*}{\tilde{P}_{t+\kappa}} - \frac{P_{t+\kappa}}{\tilde{P}_{t+\kappa}} \right] C_{it+\kappa}$, where the parameter $\mu_r = \frac{1}{\sigma - 1}$ is the subsidy rate to the retailers that corrects the distortion caused by the market power. A representative wholesaler combines a continuum of intermediate goods y_{jt} to produce wholesale goods Y_t through a CES production function $Y_t = \left(\int_0^1 \epsilon_{jt}^{\frac{1}{\eta}} y_{jt}^{\frac{\eta-1}{\eta}} \mathrm{d}j \right)^{\frac{\eta}{\eta-1}}$, where ϵ_{jt} is the idiosyncratic shock. The wholesaler's decision problem is exactly the same as that in the two-period model. To save space, we do not give detailed illustrations.

4.3 Production sector

The setup of the intermediate goods sector is similar to the two-period model. There is a continuum of heterogeneous firms indexed by $j \in [0,1]$. These firms operate under monopolistic competition, using labor input l_{jt} to produce intermediate goods y_{jt} via a linear production function $y_{jt} = l_{jt}$.

The timeline is exactly the same as that in the two-period model. At the beginning of period t, firm j has a data stock D_{jt-1} accumulated from the last period. So, the data stock is the state variable in the firm's problem. The demand for firm j's product is $P_{jt} = \epsilon_{jt}^{\frac{1}{\eta}} \left(\frac{y_{jt}}{Y_t}\right)^{-\frac{1}{\eta}} P_t$, where the idiosyncratic shock ϵ_{jt} is unobservable; P_{jt} and P_t are the individual price and aggregate price index, respectively. The firm receives a signal s_{jt} about ϵ_{jt} with the information structure $s_{jt} = \log \epsilon_{jt} + e_{jt}$, where the noise e_{jt} follows a normal distribution with zero mean and variance $1/D_{jt-1}$. Thus, the data can improve the precision of the signal. After receiving the signal, the firm decides the amount of output it aims to produce, y_{jt}. Then the individual demand shock ϵ_{jt} realizes, and the market clearing condition pins down the price P_{jt}.

The raw data is generated along with the production procedure via $d_{jt}^{raw} = \theta y_{jt}$. Firms cannot directly analyze and utilize the raw data. They rely on the data intermediary to transform the raw data into structured data. The firm sells its raw data to the data intermediary at the price Q_t^r. Meanwhile, the firm purchases d_{jt} amount of structured data from the data intermediary at the price of Q_t^d. The data accumulates according to the following law of motion

$$D_{jt} = (1-\delta) D_{jt-1} + d_{jt}, \tag{48}$$

where $\delta \in (0,1)$ is the depreciation rate of the data stock that measures how fast the data becomes obsolete.

Given the data stock D_{jt-1} and the signal s_{jt}, firm j chooses the output y_{jt} and the demand for structured data d_{jt} to maximize the sum of the discounted expected profit. Define firm j's profit from the production side in period t as $\Pi_{jt} \equiv \left[(1+\mu_p)\frac{P_{jt}}{\tilde{P}_t} - \frac{W_t}{\tilde{P}_t}\right] y_{jt}$. The Bellman equation in period t can be written as

$$\mathbf{V}_t\left(D_{jt-1}, s_{jt}\right) = \max_{y_{jt}} \mathbb{E}_t \left\{ \Pi_{jt} + \frac{Q_t^r}{\tilde{P}_t} d_{jt}^{raw} + \max_{\{D_{jt}, d_{jt}\}} \left[-\frac{Q_t^d}{\tilde{P}_t} d_{jt} + \beta \mathbb{E}_t \frac{\Lambda_{t+1}}{\Lambda_t} \mathbf{V}_{t+1}(D_{jt}, s_{jt+1}) \right] | s_{jt} \right\},$$

subject to individual demand curve $P_{jt} = \epsilon_{jt}^{\frac{1}{\eta}} \left(\frac{y_{jt}}{Y_t}\right)^{-\frac{1}{\eta}} P_t$, data generation equation $d_{jt}^{raw} = \theta y_{jt}$ and data accumulation (48). The parameter $\mu_p = \frac{1}{\eta - 1}$ is the subsidy rate to the intermediate goods producers that corrects the distortion caused by the monopolistic competition.

Akin to the two-period model, let Q_t denote the value of the data, defined as $Q_t \equiv \theta Q_t^r / W_t$.

The firm's optimal production decision y_{jt} satisfies $y_{jt} = \mathbb{E}\left[\epsilon_{jt}^{\frac{1}{\eta}}|s_{jt}\right]^{\eta} (\frac{P_t}{W_t})^{\eta} (1-Q_t)^{-\eta} Y_t$. The term $\mathbb{E}\left[\epsilon_{jt}^{\frac{1}{\eta}}|s_{jt}\right]^{\eta} = \exp\left\{\frac{D_{jt-1}s_{jt}-\frac{1}{2}}{\tau_\epsilon+D_{jt-1}} + \frac{1}{2\eta}\frac{1}{\tau_\epsilon+D_{jt-1}}\right\}$ is exactly the same as that in the two-period model, capturing the benefit of data on improving the signal precision and the expected profit. The optimal output y_{jt} strictly increases with the value of the data Q_t, reflecting the positive impact on output due to the firm's motivation to collect more data.

The optimal stock of structured data D_{jt} satisfies the following Euler equation

$$\frac{Q_t^d}{\tilde{P}_t} = \beta \mathbb{E}_t \frac{\Lambda_{t+1}}{\Lambda_t}\left\{\frac{\partial \mathbb{E}_{t+1}[\Pi_{jt+1}|s_{jt+1}]}{\partial D_{jt}} + (1-\delta)\frac{Q_{t+1}^d}{\tilde{P}_{t+1}}\right\}, \tag{49}$$

where the expected marginal effect of data on the next-period profit satisfies $\frac{\partial \mathbb{E}_{t+1}[\Pi_{jt+1}|s_{jt+1}]}{\partial D_{jt}} = \frac{1}{2}\frac{\eta-1}{\eta}\frac{\mathbb{E}_{t+1}\left[\Pi_{jt+1}|s_{jt+1}\right]}{\left(\tau_\epsilon+D_{jt}\right)^2}$. Under the assumption that the signal s_{jt} is independent over time and across individuals, the above Euler equation indicates that the optimal stock of data D_{jt} only depends on the aggregate states and is irrelevant to the individual states. Therefore, we denote the optimal data stock as $D_{jt} = D_t$ for any $j \in [0,1]$. As $\frac{\partial \mathbb{E}_{t+1}[\Pi_{jt+1}|s_{jt+1}]}{\partial D_{jt}}$ is determined by the aggregate states, the above optimal condition is essentially the asset pricing equation for the data. We will elaborate on this issue more in the later equilibrium analysis.

Notice that the data stock and the demand for the structured data are identical across firms. The raw data generated in the production procedure varies across firms as d_{jt}^{raw} is proportional to the individual firm's output y_{jt}. For a firm that receives a low signal s_{jt}, it produces a small quantity of goods and generates a limited amount of raw data. The existence of the data intermediary allows the firm to purchase structured data with a larger size than its raw data, i.e., $d_t > d_{jt}^{raw}$. Conversely, a firm that receives a high signal will produce a large quantity of output, generating a substantial amount of raw data. This firm's demand for structured data will be less than the raw data it has, and it will sell abundant data to the data intermediary. The trading of data is beneficial for both types of firms.

4.4 Data intermediary

In the intermediate goods producer's problem, we derive the demand for the structured data. We now discuss the supply side of the data. A continuum of representative data intermediaries exists, with unit measure. The data intermediaries are competitive. A representative data intermediary collects raw data from firms at the price Q_t^r, transforms it into structured data, and sells it to firms at the price Q_t^d. The total amount of raw data the data intermediary collects is $d_t^{raw} = \int_0^1 d_{jt}^{raw} \mathrm{d}j$, and the total amount of structured data supplied is $d_t \equiv \int_0^1 d_{jt} \mathrm{d}j = D_t - (1-\delta)D_{t-1}$. We assume one unit of raw data can be transformed into one unit of structured

data, implying $d_t^{raw} = d_t$. As the data market is competitive, we must have $Q_t^r = Q_t^d$.

Notice that the one-for-one basis of data transformation does not contradict the non-rivalry feature of the data as emphasized in Jones and Tonetti (2020). To see this, consider a more general scenario where the number of firms that demand data is N^d and the number of digitalized firms that can generate data is N^s. The former is larger than the latter, i.e., $N^d > N^s$. The non-rivalry of data means that the structured data can be duplicated and sold to N^d firms. The maximum amount of data each firm can purchase is $N^s d_t$. So, the data intermediary's profit is $Q_t^d N^d N^s d_t - Q_t^r N^s d_t^{raw}$. The competitive market structure implies $Q_t^d N^d = Q_t^r$. In our model, the number of firms is fixed, i.e., $N^d = N^s = 1$, implying $Q_t^d = Q_t^r$. Thus, the data's non-rivalry automatically holds. To model non-rivalry explicitly, one could endogenize the firm numbers by introducing firm entry.

4.5 Equilibrium

The definition of aggregate variables and the general equilibrium are similar to the two-period model in Section 2.2.

Aggregate production function The aggregate production function of the intermediate goods sector takes exactly the same form as that in the two-period model, which is $Y_t = Z_t L_t$. The total factor productivity is defined as $Z_t \equiv \exp\left(-\frac{1}{2\eta}\frac{1}{\tau_\epsilon + D_{t-1}}\right)$, which is strictly increasing in the data stock D_{t-1}. The final output (consumption) satisfies $C_t = Y_t/\Xi_t$, where $\Xi_t \equiv \int \left(\frac{\tilde{P}_{it}}{\tilde{P}_t}\right)^{-\sigma} \mathrm{d}i$ is the price dispersion.

New Keynesian Phillips Curve. The inflation satisfies the following New Keynesian Phillips curve

$$1 = \omega(\Pi_t)^{\sigma-1} + (1-\omega)\left(\frac{\sigma}{\sigma-1}\frac{\mathcal{A}_t}{\mathcal{B}_t}\right)^{1-\sigma}, \tag{50}$$

where

$$\mathcal{A}_t = C_t^{1-\gamma}\frac{P_t}{\tilde{P}_t} + \beta\omega\mathbb{E}_t\Pi_{t+1}^{\sigma}\mathcal{A}_{t+1}, \quad \mathcal{B}_t = C_t^{1-\gamma} + \beta\omega\mathbb{E}_t\Pi_{t+1}^{\sigma-1}\mathcal{B}_{t+1}, \tag{51}$$

and $\frac{P_t}{\tilde{P}_t}$ is the marginal cost. It is the real price of wholesale goods, which, under price subsidies μ_r and μ_p satisfies

$$\frac{P_t}{\tilde{P}_t} = \frac{1}{Z_t}\frac{W_t}{\tilde{P}_t}(1-Q_t), \tag{52}$$

where $\frac{W_t}{\tilde{P}_t}$ is the real wage given by (47); Q_t is the value of the data that measures the net benefits of collecting data. The marginal cost of the retailer decreases with the value of data Q_t. This relationship arises because increased data value enhances the marginal returns from data, incentivizing firms to boost output to collect more data. This, in turn, reduces the wholesale goods price and lowers the marginal cost for retailers.

Data accumulation. In the equilibrium, the aggregate amount of raw data is $d_t^{raw} = \int_0^1 d_{jt}^{raw} \mathrm{d}j = \theta \int_0^1 y_{jt} \mathrm{d}j = \theta L_t$, where the last equality is from the aggregation of output and labor in the intermediate goods sector. According to the previous analysis of the data intermediary's decisions, the structured data in each period has the same amount as the raw data due to the one-to-one data transformation, implying $d_t = d_t^{raw}$. Thus, the data stock accumulation in the aggregate level follows

$$D_t = (1-\delta) D_{t-1} + \theta L_t. \tag{53}$$

This equation encapsulates the dynamics of data accumulation over time, accounting for both the depreciation of existing data stock and the contribution of new data generated through production.

Data value. From the Euler equation of data stock, we can write the equilibrium value of the data $Q_t \equiv \frac{Q_t^r}{W_t/\theta} = \frac{Q_t^d}{W_t/\theta}$ into an asset pricing formula

$$Q_t = \beta \mathbb{E}_t \left[\frac{\theta}{2\eta} \frac{(1-Q_{t+1}) L_{t+1}}{(\tau_\epsilon + D_t)^2} + (1-\delta) Q_{t+1} \right]. \tag{54}$$

The value of data is the discounted sum of its future cash flows, where $\frac{\theta}{2\eta} \frac{(1-Q_{t+1}) L_{t+1}}{(\tau_\epsilon + D_t)^2}$ represents the cash flow generated by the structured data. The current value of data is positively correlated with the overall economic prosperity of the next period, as measured by L_{t+1}. Future macroeconomic prosperity increases the marginal cash flow returns of the data, which in turn boosts the current value of the data.

Monetary policy To close the model, we introduce a central bank that controls the nominal interest rate R_t following a standard Taylor rule

$$R_t = \left(\frac{C_t}{C^{ss}} \right)^{\psi_C} \left(\frac{\Pi_t}{\Pi^{ss}} \right)^{\psi_\Pi} \exp(u_t), \tag{55}$$

where C^{ss} and Π^{ss} are the steady-state levels of consumption (final output) and inflation. Coefficients $\psi_C > 0$ and $\psi_\Pi > 0$ in the policy rule govern the responsiveness of monetary policy to the deviations in final output and inflation; and u_t is the monetary policy shock, following

an AR(1) process $u_t = \rho_r u_{t-1} + \varepsilon_t^r$, where ε_t is independently and identically distributed over time, following a normal distribution $\mathbb{N}(0, \sigma_r^2)$.

A competitive equilibrium is a sequence $\{R_t, \Pi_t, C_t, Q_t, D_t\}_{t\geq 0}$ consistent with (46), (50), (53), (54), and (55), given D_{-1} and exogenous series monetary policy shocks $\{u_t\}$.

5 Dynamics

In this section, we examine the transmission of monetary policy shocks in the dynamic general equilibrium data economy. We first derive a linearized equilibrium dynamic system around the steady state. The system consists of five equations, akin to the three-equation linear system in a standard NK model. The linearization exercise offers a clear exposition of the transmission mechanism of monetary policy when the dynamic general equilibrium features are considered. In the quantitative analysis, we further calibrate the model and investigate the local determinacy of the equilibrium system around the steady state under a conventional Taylor rule. We also document the dynamic properties of the model under monetary policy shocks and compare it with a traditional NK DSGE model, where the endogenous data generation procedure is muted.

5.1 Linearized equilibrium system

Let $\hat{x}_t$ denote the percentage deviation of variable x_t from its steady state level x^{ss}, i.e., $\hat{x}_t \equiv (x_t - x^{ss})/x^{ss}$. The data-economy version of the New Keynesian dynamic general equilibrium framework can be summarized by the following five-equation system:

$$\hat{R}_t = \psi_\Pi \hat{\Pi}_t + \psi_C \hat{C}_t + u_t, \tag{56}$$

$$\hat{C}_t = -\frac{1}{\gamma}\left(\hat{R}_t - \mathbb{E}_t \hat{\Pi}_{t+1}\right) + \mathbb{E}_t \hat{C}_{t+1}, \tag{57}$$

$$\hat{\Pi}_t = \beta \mathbb{E}_t \hat{\Pi}_{t+1} + \phi\left(\gamma \hat{C}_t - \frac{Q^{ss}}{1-Q^{ss}}\hat{Q}_t - \Theta \hat{D}_{t-1}\right), \tag{58}$$

$$\hat{D}_t = \left[1 - \delta\left(1+\Theta\right)\right]\hat{D}_{t-1} + \delta \hat{C}_t, \tag{59}$$

$$\hat{Q}_t = \left[1 - \beta\left(1-\delta\right)\right]\mathbb{E}_t \hat{C}_{t+1} + \zeta_Q \mathbb{E}_t \hat{Q}_{t+1} - \zeta_D \hat{D}_t, \tag{60}$$

where $\phi = \frac{(1-\omega)(1-\beta\omega)}{\omega}$, $\zeta_Q = 1 - \frac{1-\beta(1-\delta)}{1-Q^{ss}}$, $\zeta_D = \left[1-\beta\left(1-\delta\right)\right]\left(\frac{1}{2\eta}\frac{1}{\tau_\epsilon + D^{ss}} + 2\right)\frac{D^{ss}}{\tau_\epsilon + D^{ss}}$, $\Theta = \frac{1}{2\eta}\frac{D^{ss}}{(\tau_\epsilon + D^{ss})^2}$; Q^{ss} and D^{ss} are steady-state levels of value of data and data stock, respectively.

Equations (56)–(58) constitute the standard New Keynesian block —namely, the Taylor rule, the dynamic IS curve, and the data-economy version of the NKPC. The term $\gamma \hat{C}_t$ in NKPC captures the marginal cost of production ($W_t/\tilde{P}_t$), which is obtained from the labor supply curve $W_t/\tilde{P}_t = \chi C_t^\gamma$. The term $\frac{Q^{ss}}{1-Q^{ss}}\hat{Q}_t$ captures the contemporaneous effect of data on the aggregate

supply. A higher value of data induces firms to expand production through the dynamic feedback effect, and this expansion reduces prices, resulting in a downward shift of the NKPC. In addition, data accumulated from the previous period, D_{t-1}, enhances firms' forecasting accuracy and raises aggregate TFP, further shifting the NKPC downward. The influence of data depends on the severity of information frictions, captured by the precision of the demand signal, τ_ϵ. When information frictions are absent and signal precision tends to infinity (i.e., $\tau_\epsilon \to +\infty$) the steady-state value of data, Q^{ss}, is zero, as implied by the asset pricing equation (54), and the coefficient Θ also converges to zero. Thus, the NKPC in (58) nests the standard NKPC of the New Keynesian model. Equations (59)-(60) characterize the data market conditions.

Equation (59) is the data supply curve, derived from the data accumulation in equation (54). The term $\delta \hat{C}_t$ reflects the effects of aggregate demand on the data accumulation. An economic boom increases the data stock, as higher output generates more data through the production process, thereby shifting the data supply curve outward.

Equation (60) represents the downward-sloping data demand curve, derived from the pricing equation (54). An increase in expected future consumption, $\mathbb{E}_t \hat{C}_{t+1}$, shifts the data demand curve outward. This is because an economic boom in the next period (i.e., a rise in $\hat{C}_{t+1}$) increases the cash flow generated by each unit of data, thereby stimulating firms' demand for data. In this way, monetary easing raises demand for both output and data supply, since data is a byproduct of output.

An expected economic boom increases the marginal value of data, thereby increasing current data demand, and the shift in demand is proportional to the degree of information frictions (Proposition 3). For a sufficiently high degree of information frictions, the increase in demand outweighs the expansion of data supply, so that both the price and the stock of data rise in equilibrium. A larger stock and higher value of data induce firms to increase production, thereby stimulating aggregate supply.

5.2 Quantitative analysis

To study the dynamic properties of the system –namely, system stability and the dynamic propagation of monetary policy– we calibrate the model.

Calibration We calibrate the model at quarterly frequency and divide the parameters into two subsets, $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$. Subset $\mathbf{\Omega}_1 = \{\beta, \sigma, \eta, \chi, \gamma, \omega, \psi_C, \psi_\Pi, \rho_m, \sigma_m\}$ comprises standard parameters extensively used in the literature. The discount factor β is set to be 0.99, corresponding to a quarterly real interest rate of 1%. Following the standard NK literature, we set the elasticities of substitution in the retail and the wholesale sectors to $\sigma = \eta = 6$, implying a steady-state markup of 20%. From the labor supply curve, the real wage satisfies $w_t \equiv \frac{W_t}{\tilde{P}_t} = \chi C_t^\gamma$, implying

that the parameter γ captures the relative volatility of real wages and total output at the logarithmic level, i.e., $\gamma = \frac{\text{Std}[\log(w_t)]}{\text{Std}[\log(C_t)]}$. In Stock (1999), the volatilities of real wage and total output are 0.64 and 1.66, respectively, implying $\gamma = 0.39$. Following Hansen (1985), we set the labor disutility parameter χ to be 1.432 to ensure a steady-state labor of $1/3$. Following Galí (2015), we set the price stickiness parameter ω to 0.75, the interest rate rule parameters $\psi_C = 0.125$ and $\psi_\Pi = 1.5$. Finally, we set the first-order autocorrelation coefficient for monetary policy shocks ρ_m to 0.9. We normalize the standard deviation of the monetary policy shock σ_m to 1.

Subset $\mathbf{\Omega}_2 = \{\theta, \delta, \tau_\epsilon\}$ comprises parameters specific to the data economy. The parameter θ reflects the extent of a firm's digitalization, i.e., the efficiency of data generation from one unit of output. The parameter δ captures the depreciation rate of data. To ensure that the data renders non-negligible effects on the real economy, we need reasonably large θ and δ. As we do not have a particular reference for the values of θ and δ, we set θ to 0.1 and δ to 0.4. We then conduct a sensitivity analysis on the values of these two parameters. The model's dynamic pattern remains robust. We set the prior precision of individual demand shocks τ_ϵ to 0.06, ensuring that the ratio of the prior precision to the posterior precision in steady state, $\frac{\tau_\epsilon}{\tau_\epsilon + D^{ss}}$, is 0.42, consistent with estimates from David et al. (2016). Table 1 summarizes the calibrated parameter values.

Table 1: Calibrated values of parameters

Parameter	Description	Value
β	discount factor	0.99
σ	elasticity of substitution of retail goods	6
η	elasticity of substitution of intermediate goods	6
ω	probability of not adjusting price	0.75
γ	relative risk aversion	0.39
χ	labor disutility	1.432
ψ_C	Taylor rule coefficient on output gap	0.125
ψ_Π	Taylor rule coefficient on inflation	1.5
ρ_m	persistence of monetary policy shocks	0.9
θ	digitalization level	0.1
δ	data depreciation rate	0.4
τ_ϵ	prior precision of individual demand shocks	0.06

Determinacy. We discuss the local determinacy of the dynamic system under a conventional Taylor rule. Figure 8 shows that the region of indeterminacy in the data economy is smaller compared to the traditional economy. For a given value of ψ_C, the required level of ψ_Π to

ensure determinacy is lower in the data economy than in the traditional economy; see the dashed and solid lines in Figure 8. The underlying mechanism hinges on the supply-side effect of the data in response to a monetary policy shock.

In the standard New Keynesian model, low monetary policy responsiveness to inflation can cause self-fulfilling dynamics in inflation. Positive sentiment about expected inflation raises current inflation. With an insufficient response in the nominal interest rate, the real interest rate declines, further stimulating the aggregate economy, as shown in the dynamic IS curve. A boom in the aggregate economy, in turn, pushes up inflation from the supply side, as shown in the NKPC. This confirms the initial positive sentiment on inflation, resulting in self-fulfilling dynamics.

In contrast, in the data economy, the supply-side reinforcement channel of inflation is mitigated. A positive sentiment on inflation stimulates the aggregate demand; meanwhile, the economic boom raises the value of data, motivating firms to accumulate more data. This, in turn, increases the aggregate supply and weakens the reinforcement effect on inflation. The endogenous data accumulation and its real impacts compress the region of self-fulfilling equilibrium. Therefore, the interactions between the data market and the real economy strengthen the stability of the monetary policy.

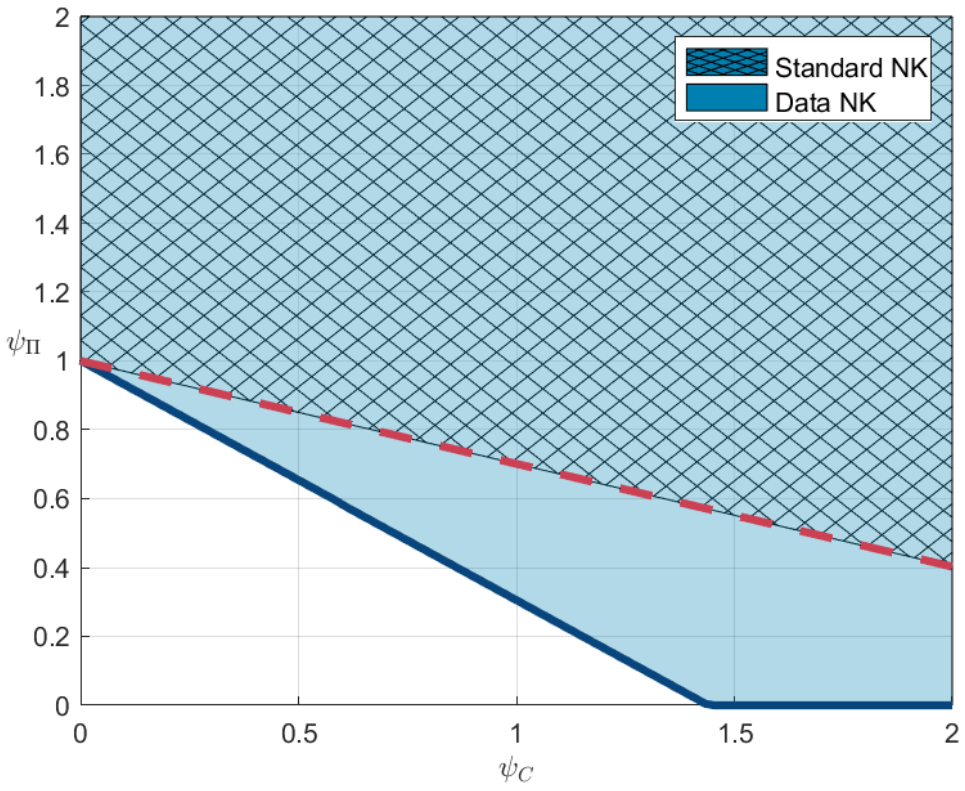


Figure 8: Local determinacy in the data economy

Notes: The dashed and solid lines represent the minimum values of ψ_Π required to ensure determinacy under different levels of ψ_C, in the standard NK economy and the data NK economy, respectively. The area of cross lines indicates the parameter values that support the local determinacy in the standard NK model. The blue area, including the part of the cross lines, is for the data economy. Parameters other than ψ_Π and ψ_C are set according to the calibrations in Table 1.

Dynamic responses to a monetary policy shock. Figure 9 illustrates the transmission mechanism of monetary policy by comparing the impulse responses to an expansionary monetary policy shock in the standard NK model (dashed line) and in the data-economy NK model (solid line).

In the standard NK model, an expansionary monetary increases demand and expands production (measured by consumption, C_t), raising inflation (Π_t), as shown in the dashed lines in the figure. Labor (N_t) increases to sustain the expansion in output resulting from higher demand. Total factor productivity (Z_t) is exogenous and unresponsive to the monetary policy.

In the data-economy New Keynesian model, expansionary monetary policy increases output and labor more strongly while inflation exhibits a smaller rise than in the standard NK model. An expansionary monetary policy that increases output also generates new data, which enhances signal precision and enables firms to better meet demand, leading to a larger output

expansion which attenuates the rise in inflation compared to the standard NK model. These differences are generated by the direct and indirect feedback effects arising from the link between data and monetary policy, outlined in the simple model in Section 2.2. An expansionary policy increases the data stock and price (D_t and Q_t, respectively) on impact. However, the data price falls below its steady-state value two quarters after the shock, reflecting the sustained expansion in data supply driven by the increase in economic activity induced by monetary policy.

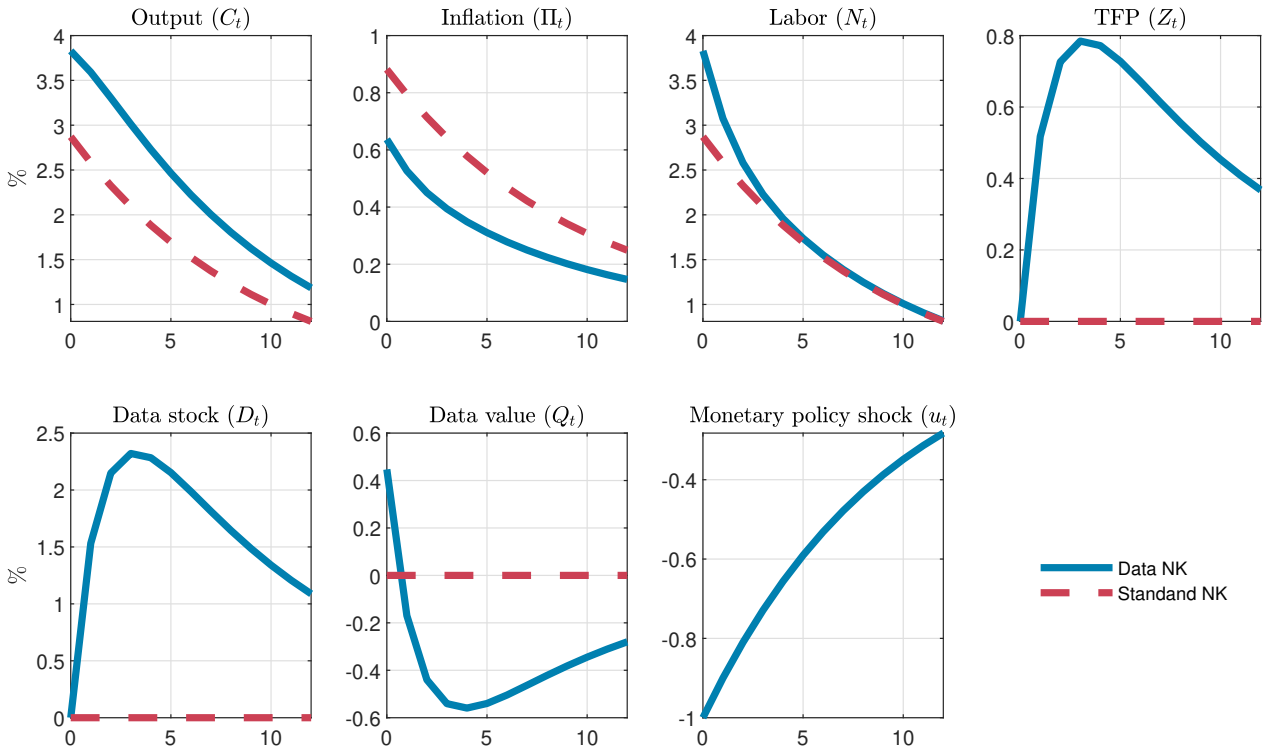


Figure 9: Impulse responses to an expansionary monetary policy shock

Notes: The figure shows the dynamic responses of key macroeconomic variables under a 1% expansionary monetary policy shock, i.e., $u_0 = -1\%$. The horizontal axis represents quarters, and the vertical axis shows the percentage change of each variable from its steady-state level. The solid and dashed lines represent the responses in the data economy and the standard NK economy, respectively. Parameter values are based on the calibration summarized in Table 1.

6 Optimal Monetary Policy in the Data Economy

In this section, we analyze the optimal monetary policy in the data economy. We analytically derive the welfare loss function from the second-order (local) approximation to households' utility around the constrained-efficient steady state. We then determine the optimal commitment policy in the face of cost-push shocks to inflation and show that a simple interest rate rule, which also considers the value of data, closely replicates the allocations of the optimal policy.

6.1 Welfare loss function in the data economy

Social welfare in the model is equal to the expected lifetime utility of the representative household $\mathcal{W} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_{it}^{1-\gamma}}{1-\gamma} - \chi L_{it} \right)$. We use the conventional definition of the welfare loss function: $\mathcal{L} = \frac{\mathcal{W}^{ss} - \mathcal{W}}{C^{ss} U_C^{ss}}$, where $\mathcal{W}^{ss}$ represents the level of welfare at the deterministic steady state absent external shocks, and C^{ss} and U_C^{ss} are steady-state consumption and the marginal utility of consumption, respectively, allowing to express the welfare loss in consumption units.

Constrained-efficient equilibrium with flexible prices. We begin by deriving the constrained-efficient steady state under information frictions with flexible prices (thus abstracting from distortions arising from nominal rigidities). The social planner sets aggregate consumption C_t and labor supply L_t, as well as the allocation of labor inputs l_{jt} and data stock D_{jt} for each firm j. As in the decentralized economy, the planner cannot observe the realization of individual demand shock ϵ_{jt} for each product variety j, and relies on a set of noisy signals $\{s_{jt}\}$ to allocate labor across firms. The signal extraction problem is the same as in the decentralized economy, with the data stock D_{jt} improving the signal's precision.

The social planner's optimization problem maximizes aggregate welfare in period t ($\mathcal{W}_t$) according to the Bellman equation:

$$\mathcal{W}_t(D_{t-1}) = \max_{\{C_t, L_t, D_t, D_{jt}, l_{jt}\}} \mathbb{E}_{\{\epsilon_{jt}\}} \left[U(C_t, L_t) | \{s_{jt}\} \right] + \beta \mathbb{E}_t \mathcal{W}_{t+1}(D_t), \tag{61}$$

subject to aggregate production function $C_t = \left(\int_0^1 \epsilon_{jt}^{\frac{1}{\eta}} l_{jt}^{\frac{\eta-1}{\eta}} dj \right)^{\frac{\eta}{\eta-1}}$, the labor resource constraint $\int_0^1 l_{jt} \mathrm{d}j = L_t$, the information structure equation $s_{jt} = \log \epsilon_{jt} + e_{jt}$, $e_{jt} \sim \mathbb{N}(0, 1/D_{jt})$, the law of motion of data: $D_{jt} = (1-\delta) D_{jt-1} + \theta l_{jt}$, where the aggregation of the data $D_t = \int_0^1 D_{jt} \mathrm{d}j$. The term $\mathbb{E}_{\{\epsilon_{jt}\}}[\cdot \mid \{s_{jt}\}]$ is the expectation operator with respect to the demand shocks $\{\epsilon_{jt}\}$, conditional on signals $\{s_{jt}\}$. The following proposition outlines the constrained-efficient allocation.

Proposition 4. *In the constrained-efficient equilibrium with flexible prices, the social planner's optimal labor allocation satisfies* $l_{jt} = \frac{g(s_{jt})}{\mathbb{E}[g(s_{jt})]} L_t$, *where* $g(s_{jt}) = \mathbb{E}\left[\epsilon_{jt}^{\frac{1}{\eta}} | s_{jt} \right]^{\eta}$. *The optimal data stock is identical across firms, i.e.,* $D_t = D_{jt}$, *for* $j \in [0,1]$. *The aggregate production function is given by* $C_t = Z_t L_t$, *and the aggregate TFP satisfies* $Z_t \equiv \mathbb{E}\left[g(s_{jt}) \right]^{\frac{1}{\eta-1}} = \exp\left(-\frac{1}{2\eta} \frac{1}{\tau_\epsilon + D_{t-1}} \right)$. *The optimal consumption–labor tradeoff is characterized by:*

$$-\frac{U_N(C_t, L_t)}{U_C(C_t, L_t)} = \frac{Z_t}{1 - Q_t}, \tag{62}$$

where U_N *and* U_C *are the marginal (dis)utilities of labor and consumption, respectively, and* $Q_t \equiv$

$-\frac{\beta\theta\mathbb{E}t\frac{\partial\mathcal{W}t+1}{\partial D_t}}{U_N(C_t,L_t)}$ *captures the value of data for the social planner, and is the same as in the decentralized economy in equation* (54).

Proof. See Appendix A.1. □

As shown in equation (62), the constrained-efficient allocation in a flexible-price equilibrium accounts for the additional benefits of labor supply in data generation. This effect is captured by the term Q_t, representing the marginal value of data to future welfare, which is the same as in the decentralized economy since the optimal condition for Q_t under the social planner is identical to the data pricing equation (54) in the decentralized equilibrium.

The steady-state allocations of the NK data economy align with those of the flexible-price equilibrium, characterized by an increased labor supply governed by the marginal value of data. This reflects the recognition that additional labor not only contributes to current production but also enhances data generation, which in turn improves future welfare through better demand forecasting and more efficient resource allocation.

Welfare loss function. Following Woodford (2003), Galí (2015) and Benigno and Woodford (2005), we approximate the welfare loss function through a second-order Taylor expansion around the constrained-efficient steady state.

Proposition 5. *The welfare loss function in the data economy, $\mathcal{L}_{data}$, can be decomposed into three components, resulting from the second-order approximation around the constrained-efficient steady state presented in Proposition 4:*

$$\mathcal{L}_{data} = \frac{1}{2}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\gamma X_t^2 + \frac{\sigma}{\phi}\hat{\Pi}_t^2 + \xi_D \hat{D}_{t-1}^2\right], \tag{63}$$

where $\hat{X}_t = \hat{C}_t - \frac{1}{\gamma}\Theta\hat{D}_{t-1}$ is the output gap, $\phi = \frac{(1-\omega)(1-\beta\omega)}{\omega}$, $\Theta = \frac{1}{2\eta}\frac{D^{ss}}{(\tau_\epsilon + D^{ss})^2}$, $\xi_D = \Theta^2\frac{\gamma-1}{\gamma} + \Theta\frac{2D^{ss}}{\tau_\epsilon + D^{ss}}$, and D^{ss} is the steady-state level of data stock, respectively.

Proof. See Appendix A.2. □

The first term captures the volatility in the output gap, which is defined as the gap between output $\hat{C}_t$ and its standard-NK "desired" level implied by the TFP ($\hat{Z}_t = \Theta\hat{D}_{t-1}$), which is $\frac{1}{\gamma}\hat{Z}_t$. The second term captures inflation volatility, as in the standard NK. The third term captures the volatility in the data stock. Data improves decision accuracy and enhances aggregate productivity. More data reduces forecast errors and improves allocative efficiency. More volatile data leads to more volatile TFP, which, given the output gap, makes output more volatile and harms welfare.

6.2 Optimal monetary policy

We use the welfare loss function in equation (63) to study optimal monetary policy in the data economy by solving the Ramsey problem. We then study simple Taylor-type interest rate rules and determine the specification that replicates the allocations under the optimal policy.

Cost-push shocks. To facilitate direct comparison of our results with the existing literature, we study the monetary policy tradeoffs in the presence of cost-push shocks, e_t, such that the NKPC in equation (58) is enriched with the error term:

$$\hat{\Pi}_t = \beta \mathbb{E}_t \hat{\Pi}_{t+1} + \phi \left(\gamma \hat{X}_t - \frac{Q^{ss}}{1 - Q^{ss}} \hat{Q}_t \right) + e_t \tag{64}$$

where the exogenous shock $\varepsilon_t^e \sim \mathbb{N}(0, \sigma_e^2)$ is identically and independently distributed.[22]

Ramsey policy. The Ramsey problem is solved by minimizing the welfare loss (63), subject to the linearized dynamic system (56)-(60).

Appendix B shows that the optimal output-inflation stabilization tradeoff in the data economy satisfies

$$\hat{X}_t + \frac{1}{\gamma} \Theta \hat{D}_{t-1} = -\sigma \hat{P}_t, \tag{65}$$

where $\hat{P}_t \equiv \log P_t - \log P_{-1}$ is the price-level deviation. Compared to the standard NK model, the output-inflation tradeoff is $\hat{X}_t = -\sigma \hat{P}_t$. The positive term $\frac{1}{\gamma} \Theta \hat{D}_{t-1}$ in the above relationship indicates that the endogenous data decisions exacerbate the output-inflation tradeoff in optimal monetary policy in the data economy. The intuition is that the production-data feedback channel amplifies output responses to cost-push shocks and mitigates inflation responses in the data economy. Therefore, compared to the standard NK model, optimal monetary policy in the data economy should place more weight on output stabilization.

Implementable interest rate rule. To implement the optimal policy in practice, we propose a simple interest rate rule that targets some specific indicators in the data economy. A key characteristic of our baseline data economy model is that new data is produced in conjunction with the production process. This incentivizes firms to expand their output to collect additional data. The incentive for data generation can be reflected in the value of data, Q_t, the ratio of the market price of structured data, Q_t^d, to the unit cost of generating data, W_t/θ. An increase in the value of data motivates firms' incentives to collect more data by producing more of it. Thus, Q_t could serve as an indicator of data market conditions. Thereby, we augment the standard

[22] Ireland (2004) shows that the cost-push shock e_t may arise from exogenous shocks to the elasticity of demand.

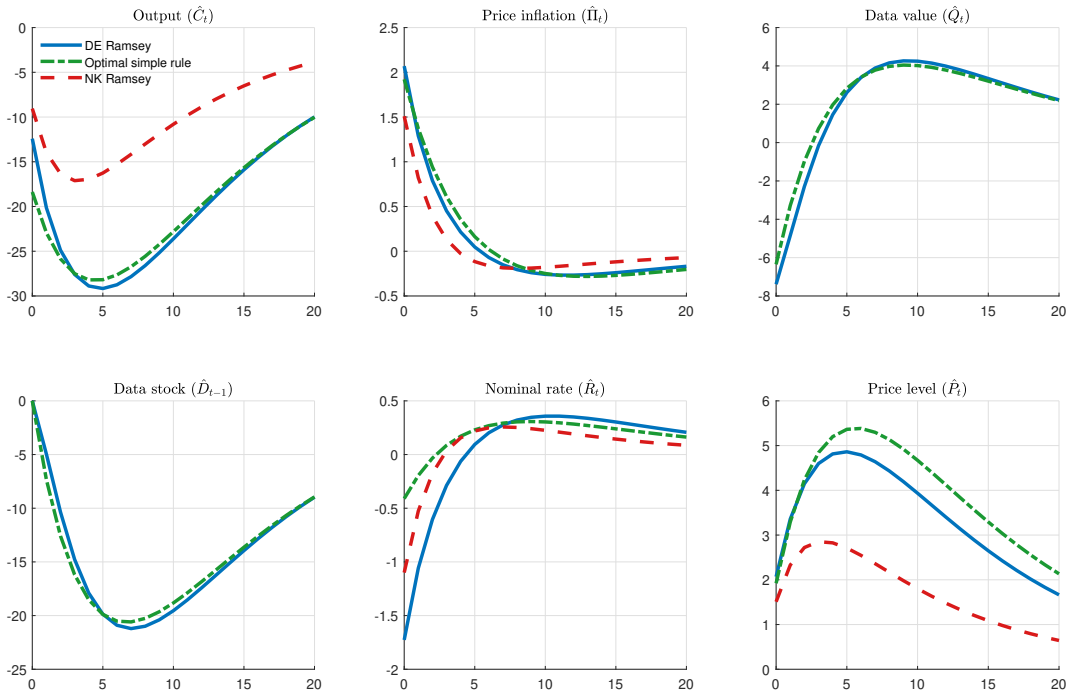


Figure 10: Dynamic responses to cost-push shocks under different monetary policies

Notes: This figure shows the dynamic responses of key macroeconomic variables to a 0.1% cost-push shock under different types of monetary policies. The horizontal axis represents time (in quarters), and the vertical axis shows the percentage change of each variable from its steady-state level (in %). The blue solid lines correspond to the optimal Ramsey policy in the data economy; the red dashed lines correspond to the Ramsey policy in the standard New Keynesian model; the green dotted lines correspond to the optimal simple rule.

Taylor rule by incorporating Q_t as an extra target. Specifically, the augmented Taylor rule in the data economy takes the form of

$$\hat{R}_t = \psi_C \hat{C}_t + \psi_\Pi \hat{\Pi}_t + \psi_Q \hat{Q}_t, \tag{66}$$

where $\psi_C, \psi_\Pi, \psi_Q > 0$.

The interpretation of this augmented Taylor rule is as follows. In the data economy, the central bank should also aim to stabilize the incentives for data generation, as endogenous data decisions create an amplification channel for business cycles. When the value of data declines, the central bank may implement monetary easing by lowering the nominal interest rate to stimulate data generation. Conversely, if the value of data increases, implying an overheated incentive for data generation, the central bank should tighten monetary policy to restore balance.

We evaluate the performance of the above augmented monetary policy by simulating the dynamic system and calculating welfare losses in response to cost-push shocks. We search over the parameter space to find optimal combinations of ψ_C, ψ_Π, and ψ_Q that minimize the welfare

loss as defined in (63). The parameters related to cost-push shocks are set to be $\sigma_e = 0.1\%$ and $\rho_e = 0.9$. Other parameters are calibrated according to the values presented in Table 1. The numerical results indicate that the optimal augmented Taylor rule implies $\psi_C = 0.08$, $\psi_\Pi = 2.97$, and $\psi_Q = 0.73$. Figure 10 reports the impulse responses to a 0.1% cost-push shock under different types of monetary policies. We consider three scenarios, including the optimal Ramsey policy (blue solid lines), the augmented Taylor rule with optimal coefficients (red dashed lines), and the standard Taylor rule (yellow dashed-dotted lines).

The endogenous supply-side effects arising from the interaction between data markets and goods markets amplify the impact of cost-push shocks. Under a standard Taylor rule, a positive cost-push shock reduces output and labor, pushing up inflation. A decline in real activity also depresses the data stock and equilibrium data value, reducing aggregate TFP. Under the augmented Taylor rule, which also targets the data value Q_t ($\psi_Q > 0$), monetary policy becomes more expansionary than under the standard Taylor rule in response to a decline in the data price. As a result, the adverse effects on the data market are mitigated, weakening the negative impacts of a cost-push shock on the real economy. The drop in output (and final consumption) is significantly smaller than that under the standard Taylor rule. Inflation is relatively volatile, exhibiting a larger increase in the impact period and a rapid decline afterward. The responses under the optimal Ramsey policy exhibit similar patterns to those under the augmented Taylor rule. This suggests that incorporating the data value into the monetary policy rule is an effective strategy for stabilizing the aggregate fluctuations in the data economy.

Welfare implications. Table 2 reports the welfare loss $\mathcal{L}$ in equation (63), expressed as the permanent decline in steady-state consumption that yields an equivalent loss, together with its decomposition into the output-gap, inflation, and data-stock components. Because losses are expressed in consumption units (i.e., $\mathcal{L} = (\mathcal{W}^{ss} - \mathcal{W})/(C^{ss}U_C^{ss})$), they are comparable across the data economy and the standard New Keynesian economy despite the different loss functions.

Three results stand out. First, cost-push shocks are more costly in the data economy. Under the Ramsey optimal policy, the welfare loss in the data economy (0.1196%) is close to twice the loss in the standard New Keynesian economy (0.0660%). The comparison is between optimal policies in the two economies, so the difference is not a consequence of a poorly designed rule but of the structure of the economy itself: endogenous data generation adds a supply-side amplification channel through which a cost-push shock lowers output, data generation, and hence future productivity, exacerbating the inflation-output tradeoff faced by the central bank.

Second, roughly half of the loss in the data economy (0.0587% of 0.1196%) is accounted for by the data-stock term $\xi_D \hat{D}_{t-1}^2$, which is absent in the standard model. This additional welfare loss arises from labor disutility caused by a distorted steady state characterized by increased

labor supply due to the incentives to accumulate more data by producing more. The data-stock term is therefore the part of the loss that monetary policy cannot fully undo, and it generates higher losses to the same shocks than an economy without data. The remaining half of the loss is split almost evenly between the output-gap (0.0341) and inflation (0.0268) components.

Third, the augmented Taylor rule with optimal coefficients closely replicates the optimal policy. Its total loss (0.1209%) slightly exceeds the Ramsey loss, and each of the three components is within a few thousandths of a percentage point of its Ramsey counterpart. The rule attains this outcome by responding to the value of data with a positive coefficient ($\psi_Q = 0.73$), easing when the value of data falls and firms' incentives to generate data weaken, alongside an aggressive response to inflation ($\psi_\Pi = 2.97$) and a small response to output ($\psi_C = 0.08$). A simple rule that accounts for the value of data recovers essentially all of the welfare attainable under commitment, underscoring the importance of stabilizing firms' data-generation incentives in the conduct of monetary policy.

Table 2: Welfare losses under alternative monetary policies

	Rule coefficients			Welfare loss (% of steady-state consumption)			
	ψ_Π	ψ_C	ψ_Q	Total $\mathcal{L}$	Output gap	Inflation	Data stock
Panel A: Standard New Keynesian economy							
Ramsey optimal policy	–	–	–	0.0660	0.0539	0.0121	–
Panel B: Data economy							
Ramsey optimal policy	–	–	–	0.1196	0.0341	0.0268	0.0587
Optimal augmented rule	2.97	0.08	0.73	0.1210	0.0350	0.0280	0.0580

Notes: The table reports the welfare loss $\mathcal{L}$ in equation (63), expressed as the permanent percentage decline in steady-state consumption that yields the same loss, in response to cost-push shocks with $\sigma_e = 0.1\%$ and $\rho_e = 0.9$. The three right-hand columns decompose the total loss into the contributions of the output-gap term $\gamma(\hat{C}_t - \Theta\hat{D}_{t-1}/\gamma)^2$, the inflation term $(\sigma/\phi)\hat{\Pi}_t^2$, and the data-stock term $\xi_D\hat{D}_{t-1}^2$; in the standard New Keynesian economy the data-stock term is absent and the output-gap term reduces to $\gamma\hat{C}_t^2$. "Optimal augmented rule" chooses $(\psi_\Pi, \psi_C, \psi_Q)$ to minimize $\mathcal{L}$ within the rule $\hat{R}_t = \psi_\Pi\hat{\Pi}_t + \psi_C\hat{C}_t + \psi_Q\hat{Q}_t$. "Ramsey optimal policy" is the commitment solution of Section 6. All other parameters follow Table 1.

Implementing an optimal commitment rule can effectively address this issue. Albeit a slight decrease in welfare gains, it manages to reduce both the standard welfare loss $\mathcal{L}_{NK}$ and the welfare loss due to the data $\mathcal{L}_{data}$ by enhancing output-inflation stabilization, as illustrated in Figure 10. This improvement leads to a decrease in total welfare loss. Furthermore, the augmented Taylor rule with optimal coefficients can also achieve a small welfare loss close to that of the optimal commitment policy, further underscoring the effectiveness and importance of stabilizing data generation incentives in the data economy.

7 Conclusion

This paper studies the transmission and the optimal conduct of monetary policy in an economy where data is a byproduct of economic activity and improves firms' forecasts of demand.

Data introduces a supply-side channel of monetary transmission that comprises direct and dynamic feedback effects. An expansionary monetary policy shock raises output and, hence, data generation, thereby improving future forecasts and productivity (direct effect). Anticipating higher future demand increases the value of data today, prompting firms to expand output to generate more sales and data (dynamic feedback effect). Both effects shift aggregate supply outward, so that monetary policy that acts through changes in demand has a larger effect on output and a smaller effect on inflation than in the standard NK model. At our calibration, the output response is about one percentage point larger and the inflation response about 0.2 percentage point smaller. The same supply-side response relaxes the Taylor principle, shrinking the policy region under which self-fulfilling inflation expectations can arise. Evidence from U.S. manufacturing, obtained by combining Compustat and IBES firm-level data with industry-level producer prices from the Bureau of Labor Statistics, corroborates our mechanism: output in data-intensive industries is more responsive to monetary policy shocks, whereas prices are less responsive; and policy-induced output expansions are followed by a persistent reduction in the forecast errors of data-intensive firms.

The normative implications are significant. The welfare loss function of the data economy includes, in addition to the output-gap and inflation terms, a term for the volatility of the data stock. The production-data feedback loop exacerbates the output-inflation stabilization tradeoff. Thus, the optimal policy should place more weight on output stabilization than that in the NK model. A Taylor rule augmented with the value of data as a target, with a positive coefficient, delivers almost the same welfare loss as that of the optimal commitment policy.

Several extensions are worth pursuing in future research. To keep the welfare criterion analytical, the model assumes transitory idiosyncratic demand shocks and a data market that equalizes the data stock across firms, thereby generating the same expectations across firms. Extending the model to firms with different degrees of digitization would allow cross-sectional differences in expectations, which recent research shows are relevant to business cycle fluctuations (Born et al., 2025; Ignaszak et al., 2026) and could therefore be an important dimension for monetary policy transmission. Extending the model to consider other roles of data, as an input to production and innovation (Jones and Tonetti, 2020; Cong et al., 2021), would enable a direct link between monetary policy and data, innovation policies, and economic growth.

References

Adam, Klaus, "Optimal Monetary Policy with Imperfect Common Knowledge," *Journal of Monetary Economics*, 2007, *54* (2), 267–301.

Andrés, Javier, Óscar Arce, and Carlos Thomas, "Banking Competition, Collateral Constraints, and Optimal Monetary Policy," *Journal of Money, Credit and Banking*, 2013, *45* (s2), 87–125.

Angeletos, George-Marios and Jennifer La'O, "Optimal monetary policy with informational frictions," *Journal of Political Economy*, 2020, *128* (3), 1027–1064.

Aoki, Kosuke, "Optimal monetary policy responses to relative-price changes," *Journal of Monetary Economics*, 2001, *48* (1), 55–80.

Argente, David and Chen Yeh, "Product life cycle, learning, and nominal shocks," *The Review of Economic Studies*, 2022, *89* (6), 2992–3054.

Asriyan, Vladimir and Alexandre N Kohlhas, "The Macroeconomics of Data," Technical Report, Working Paper. NBER Summer Institute 2024.

Baley, Isaac and Laura L Veldkamp, *The Data Economy: Tools and Applications*, Princeton University Press, 2025.

Ball, Laurence, N Gregory Mankiw, and Ricardo Reis, "Monetary policy for inattentive economies," *Journal of Monetary Economics*, 2005, *52* (4), 703–725.

Benigno, Pierpaolo and Michael Woodford, "Inflation stabilization and welfare: The case of a distorted steady state," *Journal of the European Economic Association*, 2005, *3* (6), 1185–1236.

Born, Benjamin, Zeno Enders, Manuel Menkhoff, Gernot J. Müller, and Knut Niemann, "Firm Expectations and News: Micro v Macro," CEPR Discussion Paper 17768, Centre for Economic Policy Research 2025.

Cloyne, James, Clodomiro Ferreira, Maren Froemel, and Paolo Surico, "Monetary Policy, Corporate Finance, and Investment," *Journal of the European Economic Association*, 2023, *21* (6), 2586–2634.

_ , Òscar Jordà, and Alan M. Taylor, "State-Dependent Local Projections: Understanding Impulse Response Heterogeneity," *Journal of Econometrics*, 2024, *244* (2), 105702.

Cong, Lin William, Danxia Xie, and Longtian Zhang, "Knowledge accumulation, privacy, and growth in a data economy," *Management Science*, 2021, *67* (10), 6480–6492.

Corhay, Alexandre, Kejia Hu, Jun E Li, Jincheng Tong, and Chi-Yang Tsou, "Data, markups, and asset prices," *Available at SSRN 5093509*, 2023.

Cúrdia, Vasco and Michael Woodford, "The Central-Bank Balance Sheet as an Instrument of Monetary Policy," *Journal of Monetary Economics*, 2011, *58* (1), 54–79.

David, Joel M, Hugo A Hopenhayn, and Venky Venkateswaran, "Information, misallocation, and aggregate productivity," *The Quarterly Journal of Economics*, 2016, *131* (2), 943–1005.

Eeckhout, Jan and Laura Veldkamp, "Data and Markups: A Macro-Finance Perspective," NBER Working Paper 30022, National Bureau of Economic Research 2022.

Fajgelbaum, Pablo D., Edouard Schaal, and Mathieu Taschereau-Dumouchel, "Uncertainty Traps," *The Quarterly Journal of Economics*, 2017, *132* (4), 1641–1692.

Farboodi, Maryam and Laura Veldkamp, "A Model of the Data Economy," *The Review of Economic Studies*, 2026. Forthcoming.

_ , Roxana Mihet, Thomas Philippon, and Laura Veldkamp, "Big data and firm dynamics," *AEA papers and proceedings*, 2019, *109*, 38–42.

Fernández-Villaverde, Jesús, Yang Yu, and Francesco Zanetti, "Technological Synergies, Heterogeneous Firms, and Uncertainty Shocks," *The Review of Economics and Statistics*, 2026. Forthcoming.

Furlanetto, Francesco, Paolo Gelain, and Marzie Taheri Sanjani, "Output gap, monetary policy trade-offs, and financial frictions," *Review of Economic Dynamics*, 2021, *41*, 52–70.

Galí, Jordi, *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press, 2015.

_ and Tommaso Monacelli, "Monetary Policy and Exchange Rate Volatility in a Small Open Economy," *Review of Economic Studies*, 2005, *72* (3), 707–734.

Groh, Carl-Christian, Oliver Pfäuti, and Farzad Saidi, "Monetary Policy and Business Cycles in the Data Economy," CEPR Discussion Paper 21379 2026.

Hansen, Gary D, "Indivisible labor and the business cycle," *Journal of Monetary Economics*, 1985, *16* (3), 309–327.

Ignaszak, Marek, Daniel Robbins, and Petr Sedláček, "Firm Heterogeneity and the Aggregate Effects of Innovation Policy," 2026. Working paper, University of New South Wales.

Ireland, Peter N., "Technology Shocks in the New Keynesian Model," *The Review of Economics and Statistics*, November 2004, *86* (4), 923–936.

Jarociński, Marek and Peter Karadi, "Deconstructing monetary policy surprises—the role of information shocks," *American Economic Journal: Macroeconomics*, 2020, *12* (2), 1–43.

Jones, Charles I and Christopher Tonetti, "Nonrivalry and the Economics of Data," *American Economic Review*, 2020, *110* (9), 2819–2858.

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

Nieuwerburgh, Stijn Van and Laura Veldkamp, "Learning Asymmetries in Real Business Cycles," *Journal of Monetary Economics*, 2006, *53* (4), 753–772.

Nisticò, Salvatore and Marialaura Seccareccia, "Unconventional policy and idiosyncratic risk," *Journal of Monetary Economics*, 2026, *162*, 103973.

Nuño, Galo and Carlos Thomas, "Optimal Redistributive Inflation," *Annals of Economics and Statistics*, 2022, (146), 3–64.

Ottonello, Pablo and Thomas Winberry, "Financial Heterogeneity and the Investment Channel of Monetary Policy," *Econometrica*, 2020, *88* (6), 2473–2502.

Ou, Shengliang, Donghai Zhang, and Renbin Zhang, "Information frictions, monetary policy, and the paradox of price flexibility," *Journal of Monetary Economics*, 2021, *120*, 70–82.

Paciello, Luigi and Mirko Wiederholt, "Exogenous information, endogenous information, and optimal monetary policy," *Review of Economic Studies*, 2014, *81* (1), 356–388.

Qiu, Zhesheng, Yicheng Wang, Le Xu, and Francesco Zanetti, "Monetary Policy in Open Economies with Production Networks," *Journal of Monetary Economics*, 2026, *159*, [article number].

Rubbo, Elisa, "Networks, Phillips Curves, and Monetary Policy," *Econometrica*, 2023, *91* (4), 1417–1455.

Soffritti, Mirco and Francesco Zanetti, "The Advantage of Tying One's Hands: Revisited," *International Journal of Finance & Economics*, 2008, *13* (2), 135–149.

Stock, James H, "Business Cycle Fluctuations in US Macroeconomic Time Series," *Handbook of Macroeconomics*, 1999, *1*.

Swanson, Eric T, "Measuring the effects of federal reserve forward guidance and asset purchases on financial markets," *Journal of Monetary Economics*, 2021, *118*, 32–53.

Tahbaz-Salehi, Alireza and Jennifer La'O, "Optimal Monetary Policy in Production Networks," *Econometrica*, 2022, *90* (3), 1295–1336.

Veldkamp, Laura and Cindy Chung, "Data and the aggregate economy," *Journal of Economic Literature*, 2024, *62* (2), 458–484.

Woodford, Michael, "The Taylor Rule and Optimal Monetary Policy," *American Economic Review*, May 2001, *91* (2), 232–237.

_ , *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton, NJ: Princeton University Press, 2003.

Appendix

A Derivation of Welfare Loss Function

This appendix provides detailed derivations of the welfare loss function. The first part analyzes the constrained efficiency allocation in a social planner's problem. The second part derives the welfare loss function by approximating the social welfare around the steady state of the constrained efficiency equilibrium.

A.1 Constrained efficient allocation

Social planner solves the following optimization problem:

$$\mathcal{W}_t(D_{t-1}) = \max_{L_t, D_t, \{l_{jt}\}} \mathbb{E}_{\{\epsilon_{jt}\}} \left[U\left(C_t, L_t\right) \mid \{s_{jt}\} \right] + \beta \mathbb{E}_t \mathcal{W}_{t+1}(D_t), \tag{A.1}$$

subject to the consumption basket

$$C_t = \left(\int_0^1 \epsilon_{jt}^{\frac{1}{\eta}} l_{jt}^{\frac{\eta-1}{\eta}} dj \right)^{\frac{\eta}{\eta-1}}, \tag{A.2}$$

the labor resource constraint

$$L_t = \int_0^1 l_{jt} dj, \tag{A.3}$$

the data accumulation equation

$$D_t = (1-\delta) D_{t-1} + \theta L_t, \tag{A.4}$$

and information constraint

$$s_{jt} = \log \epsilon_{jt} + e_{jt}, e_{jt} \sim \mathbb{N}(0, 1/D_{t-1}). \tag{A.5}$$

Optimal labor allocation is

$$l_{jt} = \left[-\frac{U_N(C_t, L_t) + \theta\beta \mathbb{E}_t \frac{\partial \mathcal{W}_{t+1}}{\partial D_t}}{U_C(C_t, L_t)} \right]^{-\eta} \mathbb{E}\left[\epsilon_{jt}^{\frac{1}{\eta}} \mid s_{jt} \right]^{\eta} C_t, \tag{A.6}$$

which, together with labor resource constraint (A.3) and the law of large numbers, implies

$$l_{jt} = \frac{\mathbb{E}\left[\epsilon_{jt}^{\frac{1}{\eta}} \mid s_{jt}\right]^{\eta}}{\mathbb{E}\left\{\mathbb{E}\left[\epsilon_{jt}^{\frac{1}{\eta}} \mid s_{jt}\right]^{\eta}\right\}} L_t. \tag{A.7}$$

Substituting equation (A.7) into consumption basket (A.2) yields

$$C_t = Z_t L_t, \tag{A.8}$$

where Z_t is the aggregate total factor productivity (TFP), defined as

$$Z_t \equiv \mathbb{E}\left\{\mathbb{E}\left[\epsilon_{jt}^{\frac{1}{\eta}} \mid s_{jt}\right]^{\eta}\right\}^{\frac{1}{\eta-1}}. \tag{A.9}$$

By the signal structure outlined in the information constraint (A.5), we can obtain that

$$Z_t = \exp\left(-\frac{1}{2\eta}\frac{1}{\tau_\epsilon + D_{t-1}}\right). \tag{A.10}$$

Optimal consumption-leisure tradeoff is given by

$$-U_L(C_t, L_t) = U_C(C_t, L_t) Z_t + \theta\beta\mathbb{E}_t \frac{\partial \mathcal{W}_{t+1}}{\partial D_t}, \tag{A.11}$$

where

$$\beta\mathbb{E}_t \frac{\partial \mathcal{W}_{t+1}}{\partial D_t} = \beta\mathbb{E}_t\left[U_C(C_{t+1}, L_{t+1}) Z_{t+1} \frac{1}{2\eta}\frac{L_{t+1}}{(\tau_\epsilon + D_t)^2} + (1-\delta)\beta\mathbb{E}_{t+1}\frac{\partial \mathcal{W}_{t+2}}{\partial D_{t+1}}\right]. \tag{A.12}$$

Define the value of data as $Q_t = -\frac{\theta\beta\mathbb{E}_t \frac{\partial \mathcal{W}_{t+1}}{\partial D_t}}{U_N(C_t, L_t)}$. Then, we have

$$-\frac{U_N(C_t, L_t)}{U_C(C_t, L_t)} = \frac{Z_t}{1 - Q_t}. \tag{A.13}$$

Given the relationships $U(C_t, L_t) = \frac{C_t^{1-\gamma}}{1-\gamma} - \chi L_t$, $U_L(C_t, L_t) = U_L(C_{t+1}, L_{t+1}) = -\chi$, we can further express (A.12) as

$$Q_t = \beta\mathbb{E}_t\left\{\frac{\theta}{2\eta}\frac{(1 - Q_{t+1})\, L_{t+1}}{(\tau_\epsilon + D_t)^2} + (1-\delta) Q_{t+1}\right\}, \tag{A.14}$$

which is essentially the pricing equation of data.

Regarding the total factor productivity, we have

$$Z_t = \chi\left(1 - Q_t\right) C_t^{\gamma}. \tag{A.15}$$

The constrained-effective equilibrium under information frictions can be characterized by the variables $\{C_t, L_t, Q_t, D_t, Z_t\}$, satisfying equations (A.4), (A.8), (A.10), (A.14), and (A.15). In market equilibrium, the corresponding equation for (A.15) is $Z_t = \frac{\sigma}{\sigma-1}\frac{\eta}{\eta-1}\frac{1}{1+\mu_r}\frac{1}{1+\mu_p}\chi\left(1 - Q_t\right) C_t^{\gamma}$ and the remaining equations are the same. Since $\frac{\sigma}{\sigma-1}$ and $\frac{\eta}{\eta-1}$ come from the monopoly power of retailers and intermediate goods producers, to achieve constrained optimality in market equilibrium, the social planner can subsidize retailers and intermediate goods producers by setting $\mu_r = \frac{1}{\sigma-1}$ and $\mu_p = \frac{1}{\eta-1}$ to ensure that (A.15) holds.

A.2 Welfare loss function

Denote log-level deviation from steady state as $\hat{x}_t \equiv \log x_t - \log x^{ss}$, where x^{ss} is the steady state of variable x. Second-order Taylor expansion of the utility function U_t with respect to $\hat{C}_t$ and $\hat{L}_t$ around the steady state implies

$$\begin{aligned} U_t - U^{ss} &\simeq U_C C^{ss}\left[\hat{C}_t + \frac{1}{2}\left(1 + \frac{U_{CC}C^{ss}}{U_C}\right)\hat{C}_t^2\right] + U_L L^{ss}\left[\hat{L}_t + \frac{1}{2}\left(1 + \frac{U_{LL}L^{ss}}{U_L}\right)\hat{L}_t^2\right] \\ &= U_C C^{ss}\left(\hat{C}_t + \frac{1-\gamma}{2}\hat{C}_t^2\right) + U_L L^{ss}\left(\hat{L}_t + \frac{1}{2}\hat{L}_t^2\right), \end{aligned}$$

where the last equality holds because $\frac{U_{CC}C^{ss}}{U_C} = -\gamma$ and $\frac{U_{LL}L^{ss}}{U_L} = 0$, given $U(C_t, L_t) = \frac{C_t^{1-\gamma}}{1-\gamma} - \chi L_t$. According to equations (A.8), (A.10) and (A.15), in the constrained efficiency equilibrium, we have $U_L L^{ss} = -\frac{1}{1-Q^{ss}} U_C C^{ss}$. The economic implication is that, due to the benefits of increased labor in data accumulation, the social planner increases household labor supply compared to the standard New Keynesian model, as reflected in $\frac{1}{1-Q^{ss}} > 1$. The second-order approximation of the utility function U_t can be further expressed as

$$\begin{aligned} U_t - U^{ss} &\approx U_C C^{ss}\left(\hat{C}_t + \frac{(1-\gamma)}{2}\hat{C}_t^2\right) - \frac{1}{1-Q^{ss}} U_C C^{ss}\left(\hat{L}_t + \frac{1}{2}\hat{L}_t^2\right) \\ &= U_C C^{ss}\left[\hat{C}_t + \frac{(1-\gamma)}{2}\hat{C}_t^2 - \frac{1}{1-Q^{ss}}\left(\hat{L}_t + \frac{1}{2}\hat{L}_t^2\right)\right]. \end{aligned} \tag{A.16}$$

Aggregate production function $C_t = Z_t L_t / \Xi_t$ implies

$$\hat{L}_t = \hat{C}_t + \hat{\Xi}_t - \hat{Z}_t. \tag{A.17}$$

Second-order approximation of $Z_t = \exp\left(-\frac{1}{2\eta}\frac{1}{\tau_\epsilon + D_{t-1}}\right)$ yields

$$\hat{Z}_t \approx \Theta \hat{D}_{t-1} + \frac{1}{2}\Theta \frac{\tau_\epsilon - D^{ss}}{\tau_\epsilon + D^{ss}} \hat{D}_{t-1}^2, \tag{A.18}$$

where $\Theta = \frac{1}{2\eta}\frac{D^{ss}}{(\tau_\epsilon + D^{ss})^2}$.

According to the standard New Keynesian literature, we have

$$\hat{\Xi}_t \simeq \frac{\sigma}{2}\mathbf{Var}(\tilde{P}_{it}), \tag{A.19}$$

where **Var** is the unconditional variance of individual prices $\tilde{P}_{it}$.

Combining equations (A.16), (A.17), (A.18), and (A.19) yields

$$\begin{aligned}
-\frac{U_t - U^{ss}}{U_C C^{ss}} &\approx -\hat{C}_t - \frac{(1-\gamma)}{2}\hat{C}_t^2 + \frac{1}{1-Q^{ss}}\left[\hat{C}_t + \hat{\Xi}_t - \hat{Z}_t + \frac{1}{2}(\hat{C}_t + \hat{\Xi}_t - \hat{Z}_t)^2\right] \\
&\approx -\hat{C}_t - \frac{1-\gamma}{2}\hat{C}_t^2 + \frac{1}{1-Q^{ss}}\left[\hat{C}_t + \frac{1}{2}(\hat{C}_t - \Theta\hat{D}_{t-1})^2 + \hat{\Xi}_t - \Theta\hat{D}_{t-1} - \frac{1}{2}\Theta\frac{\tau_\epsilon - D^{ss}}{\tau_\epsilon + D^{ss}}\hat{D}_{t-1}^2\right] \\
&\approx \frac{Q^{ss}}{1-Q^{ss}}\hat{C}_t - \frac{\Theta}{1-Q^{ss}}\hat{D}_{t-1} - \frac{(1-\gamma)}{2}\hat{C}_t^2 + \frac{1}{2}\frac{1}{1-Q^{ss}}\left[(\hat{C}_t - \Theta\hat{D}_{t-1})^2 + \sigma\mathbf{Var}(\tilde{P}_{i,t}) - \Theta\frac{\tau_\epsilon - D^{ss}}{\tau_\epsilon + D^{ss}}\hat{D}_{t-1}^2\right],
\end{aligned}$$

where the second and third equalities are derived by dropping the higher-order terms.

We now define the welfare loss function as

$$\mathcal{L} = -\sum_{t=0}^{\infty}\beta^t \frac{U_t - U^{ss}}{U_C C^{ss}}. \tag{A.20}$$

Based on Woodford (2003), we can express $\mathbf{Var}(\tilde{P}_{it})$ as

$$\sum_{t=0}^{\infty}\beta^t \mathbf{Var}(\tilde{P}_{it}) = \frac{\omega}{(1-\beta\omega)(1-\omega)}\sum_{t=0}^{\infty}\beta^t \hat{\Pi}_t^2. \tag{A.21}$$

The welfare loss function can be further expressed as:

$$\mathcal{L} = \mathbb{E}_0 \sum_{t=0}^{\infty}\beta^t \left\{ \begin{array}{c} \frac{Q^{ss}}{1-Q^{ss}}\hat{C}_t - \frac{\Theta}{1-Q^{ss}}\hat{D}_{t-1} - \frac{1-\gamma}{2}\hat{C}_t^2 \\ +\frac{1}{2}\frac{1}{1-Q^{ss}}\left[(\hat{C}_t - \Theta\hat{D}_{t-1})^2 + \frac{\sigma}{\phi}\hat{\Pi}_t^2 - \Theta\frac{\tau_\epsilon - D^{ss}}{\tau_\epsilon + D^{ss}}\hat{D}_{t-1}^2\right] \end{array} \right\} \tag{A.22}$$

where $\phi = \frac{(1-\omega)(1-\beta\omega)}{\omega}$ and $\Theta = \frac{1}{2\eta}\frac{D^{ss}}{(\tau_\epsilon + D^{ss})^2}$.

We follow the approach Benigno and Woodford (2005) to eliminate the linear terms in $\mathcal{L}$. A second-order approximation on the data accumulation equation $D_t = (1-\delta)D_{t-1} + \theta L_t$ gives

$$\left[(1-\delta)\hat{D}_{t-1} + \delta\hat{L}_t - \hat{D}_t\right] + \left[\tfrac{1}{2}(1-\delta)\hat{D}_{t-1}^2 + \tfrac{1}{2}\delta\hat{L}_t^2 - \tfrac{1}{2}\hat{D}_t^2\right] = 0, \tag{A.23}$$

where

$$\begin{aligned}\hat{L}_t &= \hat{C}_t - \hat{Z}_t + \hat{\Xi}_t \\ &= \hat{C}_t - \Theta\hat{D}_{t-1} - \frac{1}{2}\Theta\frac{\tau_\epsilon - D^{ss}}{\tau_\epsilon + D^{ss}}\hat{D}_{t-1}^2 + \hat{\Xi}_t\end{aligned}$$

Collecting terms, we can obtain that

$$\begin{aligned}0 &= -\delta\hat{C}_t - (1-\delta-\delta\Theta)\hat{D}_{t-1} + \hat{D}_t - \delta\hat{\Xi}_t - \tfrac{1-\delta}{2}\hat{D}_{t-1}^2 \\ &- \tfrac{\delta}{2}\left(\hat{C}_t - \Theta\hat{D}_{t-1}\right)^2 + \frac{1}{2}\delta\Theta\frac{\tau_\epsilon - D^{ss}}{\tau_\epsilon + D^{ss}}\hat{D}_{t-1}^2 + \tfrac{1}{2}\hat{D}_t^2 \qquad \text{(A.24)}\end{aligned}$$

Multiplying (A.24) by $\frac{1}{\delta}\frac{Q^{ss}}{1-Q^{ss}}$ and substituting it into welfare loss function (A.22), we can obtain that

$$\mathcal{L} = \mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\left\{\begin{array}{c}\frac{1}{\delta}\frac{1}{1-Q^{ss}}\left[Q^{ss}\hat{D}_t - (\delta\Theta + Q^{ss}(1-\delta-\Theta\delta))\hat{D}_{t-1}\right] \\ -\frac{1-\gamma}{2}\hat{C}_t^2 + \frac{1}{2}\frac{\sigma}{\phi}\hat{\Pi}_t^2 + \frac{1}{2}\left(\hat{C}_t - \Theta\hat{D}_{t-1}\right)^2 \\ -\frac{1-\delta}{2}\frac{1}{\delta}\frac{Q^{ss}}{1-Q^{ss}}\hat{D}_{t-1}^2 - \frac{1}{2}\Theta\frac{\tau_\epsilon - D^{ss}}{\tau_\epsilon + D^{ss}}\hat{D}_{t-1}^2 + \frac{1}{2}\frac{1}{\delta}\frac{Q^{ss}}{1-Q^{ss}}\hat{D}_t^2\end{array}\right\} \qquad \text{(A.25)}$$

Since $\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\hat{D}_{t-1} = \hat{D}_{-1} + \mathbb{E}_0\sum_{t=0}^{\infty}\beta^{t+1}\hat{D}_t$ and $\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\hat{D}_{t-1}^2 = \hat{D}_{-1}^2 + \mathbb{E}_0\sum_{t=0}^{\infty}\beta^{t+1}\hat{D}_t^2$, where $\hat{D}_{-1}$ is predetermined, by rearranging terms, we can obtain

$$\mathcal{L} = \mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\left\{\begin{array}{c}\frac{1}{\delta}\frac{1}{1-Q^{ss}}\left[Q^{ss} - \beta(\delta\Theta + Q^{ss}(1-\delta-\Theta\delta))\right]\hat{D}_t \\ +\frac{1}{2}\gamma\left(\hat{C}_t - \frac{1}{\gamma}\Theta\hat{D}_{t-1}\right)^2 + \frac{1}{2}\frac{\sigma}{\phi}\hat{\Pi}_t^2 + \frac{1}{2}\beta\left(1-\frac{1}{\gamma}\right)\Theta^2\hat{D}_t^2 \\ +\frac{1}{2}\left(\frac{1}{\delta}\frac{Q^{ss}}{1-Q^{ss}}(1-\beta(1-\delta)) - \beta\Theta\frac{\tau_\epsilon - D^{ss}}{\tau_\epsilon + D^{ss}}\right)\hat{D}_t^2\end{array}\right\} + t.i.p. \qquad \text{(A.26)}$$

where t.i.p. denotes terms that are independent of policy.

In the constrained-efficient steady state, by data value equation (A.14) and steady state data stock $D^{ss} = \theta L_{ss}/\delta$, the steady state data value satisfies

$$Q^{ss} = \frac{\beta\delta\Theta}{1-\beta(1-\delta-\Theta\delta)}, \qquad \text{(A.27)}$$

which implies $\frac{1}{\delta}\frac{Q^{ss}}{1-Q^{ss}} = \frac{\beta\Theta}{1-\beta(1-\delta)}$ and $Q^{ss} = \beta(\delta\Theta + Q^{ss}(1-\delta-\Theta\delta))$.

Therefore, we have

$$\mathcal{L} = \frac{1}{2}\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\left[\gamma\left(\hat{C}_t - \frac{1}{\gamma}\Theta\hat{D}_{t-1}\right)^2 + \frac{\sigma}{\phi}\hat{\Pi}_t^2 + \xi_D\hat{D}_{t-1}^2\right] + t.i.p. \qquad \text{(A.28)}$$

where $\phi = \frac{(1-\omega)(1-\beta\omega)}{\omega}$, $\Theta = \frac{1}{2\eta}\frac{D^{ss}}{(\tau_\epsilon + D^{ss})^2}$, $\xi_D = \Theta^2\frac{\gamma-1}{\gamma} + \Theta\frac{2D^{ss}}{\tau_\epsilon + D^{ss}}$, and D^{ss} is the steady-state level of data stock, respectively.

B Derivation of optimal monetary policy

We now derive the optimal monetary policy with commitment. The government solves a Ramsey problem, which minimizes the welfare loss by choosing the paths of $\{\hat{C}_t, \hat{\Pi}_t, \hat{R}_t, \hat{D}_t, \hat{Q}_t\}$, taking as given the market equilibrium system that consists of the dynamic IS curve (57), the NKPC (58), the aggregate data accumulation (59), and the data pricing equation (60). The optimization problem can be written as follows.

$$\min_{\{\hat{C}_t,\hat{\Pi}_t,\hat{R}_t,\hat{D}_t,\hat{Q}_t\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left\{ \begin{array}{l} \frac{1}{2}\left[\gamma(\hat{C}_t - \frac{1}{\gamma}\Theta\hat{D}_{t-1})^2 + \frac{\sigma}{\phi}\hat{\Pi}_t^2 + \xi_D\hat{D}_{t-1}^2\right] \\ +\lambda_t^R\left[\hat{C}_{t+1} - \frac{1}{\gamma}\left(\hat{R}_t - \hat{\Pi}_{t+1}\right) - \hat{C}_t\right] \\ +\lambda_t^\Pi\left[\hat{\Pi}_t - \beta\hat{\Pi}_{t+1} - \phi\left(\gamma\hat{C}_t - \frac{Q^{ss}}{1-Q^{ss}}\hat{Q}_t - \Theta\hat{D}_{t-1}\right) - e_t\right] \\ +\lambda_t^D\left[\hat{D}_t - (1-\delta(1+\Theta))\hat{D}_{t-1} - \delta\hat{C}_t\right] \\ +\lambda_t^Q\left[\hat{Q}_t - (1-\beta(1-\delta))\hat{C}_{t+1} - \zeta_Q\hat{Q}_{t+1} + \zeta_D\hat{D}_t\right] \end{array} \right\},$$

where $\phi = \frac{(1-\omega)(1-\beta\omega)}{\omega}$ and $\Theta = \frac{1}{2\eta}\frac{D^{ss}}{(\tau_\epsilon + D^{ss})^2}$; λ_t^R, λ_t^Π, λ_t^D and λ_t^Q are Lagrange multipliers of constraints (57), (58), (59), and (60), respectively.

First order condition of nominal interest rate $\hat{R}_t$ yields $\lambda_t^R = 0$, implying that the dynamic IS curve is not binding in the Ramsey problem.

First order condition of inflation $\hat{\Pi}_t$ implies

$$\lambda_t^\Pi = -\frac{\sigma}{\phi}\hat{\Pi}_t + \lambda_{t-1}^\Pi. \tag{B.1}$$

Backward iteration yields

$$\lambda_t^\Pi = -\frac{\sigma}{\phi}\hat{P}_t, \tag{B.2}$$

where $\hat{P}_t = \log P_t - \log P_{-1}$.

First order condition of inflation path of the final output (consumption) $\hat{C}_t$ yields

$$\gamma(\hat{C}_t - \frac{1}{\gamma}\Theta\hat{D}_{t-1}) = \gamma\phi\lambda_t^\Pi + \delta\lambda_t^D + (\frac{1}{\beta} - (1-\delta))\lambda_{t-1}^Q. \tag{B.3}$$

The optimal condition for the data stock $\hat{D}_t$ is given by

$$\begin{aligned} \lambda_t^D + \zeta_D\lambda_t^Q &= -\beta\xi_D\hat{D}_t + \beta\Theta\mathbb{E}_t(\hat{C}_{t+1} - \frac{1}{\gamma}\Theta\hat{D}_t) + \beta\left[1-\delta(1+\Theta)\right]\mathbb{E}_t\lambda_{t+1}^D - \beta\Theta\phi\mathbb{E}_t\lambda_{t+1}^\Pi \\ &= -\beta\xi_D\hat{D}_t + \beta\Theta\frac{1}{\gamma}\delta\mathbb{E}_t\lambda_{t+1}^D + \beta\Theta\frac{1}{\gamma}(\frac{1}{\beta} - (1-\delta))\lambda_t^Q + \beta\left[1-\delta(1+\Theta)\right]\mathbb{E}_t\lambda_{t+1}^D \end{aligned}$$

where the second equality is obtained by substituting into eq. (B.3). Using $\zeta_D = (1-\beta(1-\delta))(\Theta + \frac{2D^{ss}}{\tau_\epsilon + D^{ss}})$ and $\xi_D = \Theta^2 \frac{\gamma-1}{\gamma} + \Theta \frac{2D^{ss}}{\tau_\epsilon + D^{ss}}$, rearranging terms yields

$$\lambda_t^D = -\beta \xi_D \hat{D}_t - (1-\beta(1-\delta)) \frac{\xi_D}{\Theta_D} \lambda_t^Q + \beta[1-\delta(1+\Theta(1-\frac{1}{\gamma}))] \mathbb{E}_t \lambda_{t+1}^D. \tag{B.4}$$

The optimal condition for the data value $\hat{Q}_t$ is given by

$$\lambda_t^Q - \zeta_Q \beta^{-1} \lambda_{t-1}^Q = -\frac{Q^{ss}}{1-Q^{ss}} \phi \lambda_t^\Pi. \tag{B.5}$$

Using $\zeta_Q = \beta[1-\delta(1+\Theta)]$ and $\frac{Q^{ss}}{1-Q^{ss}} = \frac{\beta\delta\Theta}{1-\beta(1-\delta)}$, we can further obtain that

$$\lambda_t^Q = -\frac{\beta\delta\Theta}{1-\beta(1-\delta)} \phi \lambda_t^\Pi + (1-\delta(1+\Theta)) \lambda_{t-1}^Q. \tag{B.6}$$

Next, we show that $\lambda_t^D = 0$ and $\lambda_t^Q = -\frac{\beta\Theta}{1-\beta(1-\delta)} \hat{D}_t$. Define $h_t = \lambda_t^Q + \frac{\beta\Theta}{1-\beta(1-\delta)} \hat{D}_t$. Then by (B.4),

$$\lambda_t^D = -(1-\beta(1-\delta)) \frac{\xi_D}{\Theta_D} h_t + \beta[1-\delta(1+\Theta(1-\frac{1}{\gamma}))] \mathbb{E}_t \lambda_{t+1}^D. \tag{B.7}$$

Substituting (B.3) into data accumulation equation (59) yields

$$\hat{D}_t = [1-\delta(1+\Theta(1-\frac{1}{\gamma}))] \hat{D}_{t-1} + \delta\phi\lambda_t^\Pi + \frac{\delta}{\gamma}(\frac{1}{\beta} - (1-\delta)) \lambda_{t-1}^Q + \frac{\delta^2}{\gamma} \lambda_t^D. \tag{B.8}$$

Multiplying the above equation by $\frac{\beta\Theta}{1-\beta(1-\delta)}$ and combining it with (B.6),

$$\lambda_t^Q + \frac{\beta\Theta}{1-\beta(1-\delta)} \hat{D}_t = [1-\delta(1+\Theta(1-\frac{1}{\gamma}))] \left(\lambda_{t-1}^Q + \frac{\beta\Theta}{1-\beta(1-\delta)} \hat{D}_{t-1} \right) + \frac{\beta\Theta}{1-\beta(1-\delta)} \frac{\delta^2}{\gamma} \lambda_t^D,$$

which implies

$$h_t = [1-\delta(1+\Theta(1-\frac{1}{\gamma}))] h_{t-1} + \frac{\beta\Theta}{1-\beta(1-\delta)} \frac{\delta^2}{\gamma} \lambda_t^D. \tag{B.9}$$

Since $h_{-1} = 0$, equations (B.9) and (B.7) imply that

$$\lambda_t^D = h_t = 0 \quad \Rightarrow \quad \lambda_t^Q = -\frac{\beta\Theta}{1-\beta(1-\delta)} \hat{D}_t, \quad t \geq 0. \tag{B.10}$$

Substituting (B.10) into (B.3) implies

$$\hat{C}_t = \phi\lambda_t^{\Pi} = -\sigma\hat{P}_t, \tag{B.11}$$

which is the same as the standard New Keynesian model.

Substituting (B.11) into the linearized NKPC (58), data accumulation equation (59) and data value equation (60), the optimal ramsey policy path can be summarized by the linear system of equations of $\hat{P}_t$, $\hat{D}_t$, and $\hat{Q}_t$.

$$\begin{aligned}
(1+\beta+\phi\gamma\sigma)\hat{P}_t &= \hat{P}_{t-1} + \beta\mathbb{E}_t\hat{P}_{t+1} - \phi\left(\frac{Q^{ss}}{1-Q^{ss}}\hat{Q}_t + \Theta\hat{D}_{t-1}\right) + e_t, && \text{(B.12)}\\
\hat{D}_t &= \left[1-\delta\left(1+\Theta\right)\right]\hat{D}_{t-1} - \delta\sigma\hat{P}_t, && \text{(B.13)}\\
\hat{Q}_t &= -\sigma\left[1-\beta\left(1-\delta\right)\right]\mathbb{E}_t\hat{P}_{t+1} + \beta[1-\delta(1+\Theta)]\mathbb{E}_t\hat{Q}_{t+1} - \zeta_D\hat{D}_t, && \text{(B.14)}
\end{aligned}$$

where $\zeta_D = (1-\beta(1-\delta))(\Theta + \frac{2D^{ss}}{\tau_\epsilon + D^{ss}})$.

Substituting (B.11) into the dynamic IS curve (57) yield the path of interest rate $\hat{R}_t$:

$$\hat{R}_t = (1-\gamma\sigma)\mathbb{E}_t\hat{\Pi}_{t+1}, \tag{B.15}$$

which, by (B.12) gives

$$\hat{R}_t = \frac{1-\gamma\sigma}{\beta}\left[(1+\phi\gamma\sigma)\hat{P}_t - \hat{P}_{t-1} + \phi\left(\frac{Q^{ss}}{1-Q^{ss}}\hat{Q}_t + \Theta\hat{D}_{t-1}\right) - e_t\right]. \tag{B.16}$$

In the data economy, the interest rate path under the optimal Ramsey policy must be adjusted according to $\hat{Q}_t$ and $\hat{D}_{t-1}$.