Elasticity of Attention and Optimal Monetary Policy Online Appendix

Shaowen Luo — Virginia Tech
Kwok Ping Tsang — Virginia Tech

1 Linear Approximation of the Firm's Optimization Problem

The production function is subject to an aggregate technology shock A with a mean of 1, $Y_i = AL_i$. The demand function faced by each firm i is $Y^i(P_i) = \left(\frac{P_i}{P}\right)^{-\eta} Y$, where $P = \left(\int (P_i)^{1-\eta} di\right)^{1/(1-\eta)}$. Firm i solves the following profit maximization problem:

$$\max_{P_i} E_i[(1-t_i)P^iY^i(P^i) - WY_i(P_i)/A].$$

Total revenue is reduced by the term $t_i = \frac{1}{1-\eta^{ss}} + u_i$. u_i is an idiosyncratic revenue shock (e.g. liquidity or taxation shock) with mean 0, and η^{ss} is the steady state elasticity of substitution between varieties. The first-order condition for the problem is $E_i\left[(1-t_i)\left(\frac{P_i}{P}\right)^{1-\eta} - \frac{W}{AP}\left(\frac{P_i}{P}\right)^{-\eta}\right] = 0$. Since it contains some random shocks, to find the optimal price P_i we first need to linearize the condition.

Utility maximization by the representative household equates the marginal rate of substitution and the real wage:

$$\frac{W}{P} = \frac{V'(\hat{L})}{U'(C)} = \frac{V'(\hat{L})}{U'(Y)}$$

where $\hat{L} = \int L_i di$.

We consider a symmetric deterministic steady state (denoted by the upper bar) with $P^i = \bar{P}$ and $Y^i = \bar{Y}$, where $\bar{P}$ is some price level chosen by the central bank. In this steady state, we have $V'(\bar{Y})/U'(\bar{Y}) = 1$ and, since the technology shock has a mean of 1, $\bar{L} = \bar{Y}$.

By definition, the sum of labor can be rewritten as $\hat{L} = \int_0^1 \left(\frac{Y_i}{A}\right) di = \left(\frac{Y}{A}\right) \int_0^1 \left(\frac{P_i}{P}\right)^{-\eta} di$, and from it know that $(\hat{L} - \bar{L}) = (Y - \bar{Y}) - (A-1) + \mathcal{O}(1)$, where $\mathcal{O}(1)$ is an approximation error of order smaller than one.

Efficient output level is defined by the condition that $V'(L^*)/U'(Y^*) = A$, or the marginal rate of substitution is equal to the marginal product of labor. In other words, the productivity shock A changes the efficient level of output. Linearizing around the steady state gives us the following condition:

$$\frac{V''(\bar{Y})}{V'(\bar{Y})}(L^* - \bar{Y}) - \frac{U''(\bar{Y})}{U'(\bar{Y})}(Y^* - \bar{Y}) = A - 1.$$

Using the approximation earlier, we have $L^* = Y^* - (A-1) + \mathcal{O}(1)$, and the above condition becomes,

$$\frac{Y^* - \bar{Y}}{\bar{Y}} = \frac{1 + \frac{V''(\bar{Y})}{V'(\bar{Y})}}{\frac{V''(\bar{Y})}{V'(\bar{Y})} - \frac{U''(\bar{Y})}{U'(\bar{Y})}} \frac{1}{\bar{Y}}(A-1) = \frac{V'(\bar{Y}) + V''(\bar{Y})}{V''(\bar{Y}) - U''(\bar{Y})} \frac{1}{\bar{Y}}(A-1).$$

Whenever the productivity shock deviates from its mean, the efficient output deviates from the steady state value.

Now we are ready to linearize the first order condition of firm's problem around the steady state.

$$\frac{P^i-\bar{P}}{\bar{P}} = E_i\left[\frac{P-\bar{P}}{\bar{P}} + \frac{V''(\bar{Y})}{V'(\bar{Y})}(\hat{L}-\bar{Y}) - \frac{U''(\bar{Y})}{U'(\bar{Y})}[(Y-\bar{Y})] - (A-1) + \frac{\eta-\eta^{ss}}{(1-\eta^{ss})\,\eta^{ss}} + u_i + O(1)\right]$$

$$= E_i\left[\frac{P-\bar{P}}{\bar{P}} + \left(\frac{V''(\bar{Y})}{V'(\bar{Y})} - \frac{U''(\bar{Y})}{U'(\bar{Y})}\right)(Y-\bar{Y}) - \left(\frac{V''(\bar{Y})}{V'(\bar{Y})}+1\right)(A-1) + \frac{\tilde{\eta}}{(1-\eta^{ss})\,\eta^{ss}} + u_i + O(1)\right]$$

$$= E_i\left[\frac{P-\bar{P}}{\bar{P}} + \frac{V''(\bar{Y})-U''(\bar{Y})}{V'(\bar{Y})}\bar{Y}\left(\frac{Y-\bar{Y}}{\bar{Y}} - \frac{Y^*-\bar{Y}}{\bar{Y}}\right) + \frac{\tilde{\eta}}{(1-\eta^{ss})\,\eta^{ss}} + u_i + O(1)\right].$$

The lower case letters are used to denote the percentage deviation from their steady state values, i.e. $x \equiv \frac{X-\bar{X}}{\bar{X}}$. Nominal expenditure is denoted by $\pi \equiv y+p$. Also, we define $\alpha \equiv \frac{V''(\bar{Y})U'(\bar{Y})-V'(\bar{Y})U''(\bar{Y})}{(U'(\bar{Y}))^2}\bar{Y}$. Thus, each firm or agent chooses an action p_i to

$$\min \quad E_i\left[p_i-(1-\alpha)p-\alpha(\pi-y^*)-\frac{\tilde{\eta}}{(1-\eta^{ss})\,\eta^{ss}}-u_i\right]^2$$

We rename the markup shock as $\tilde{\theta}=\frac{\tilde{\eta}}{(1-\eta^{ss})\eta^{ss}}$. The optimization problem of firm i is then equivalent to choosing p_i to

$$\min \quad E_i[p_i-(1-\alpha)p-\alpha(\pi-y^*)-\tilde{\theta}-u_i]^2$$

and we have the markup shock $\tilde{\theta}\sim\mathcal{N}(0,\sigma_\theta^2)$ and $\sigma_\theta^2=\frac{\sigma_\eta^2}{[(1-\eta^{ss})\eta^{ss}]^2}$.

2 Derivation of the Welfare Function

To derive the welfare function in log, we approximate the utility function $\Omega\equiv U(Y)-V(L)$ around the steady state at the second order. It is straightforward to show that the deviation from the steady state can be approximated as

$$\Omega-\bar{\Omega}\simeq-\frac{1}{2}[V''(\bar{Y})-U''(\bar{Y})]\bar{Y}^2(y-y^*)^2-\frac{1}{2}V'(\bar{Y})\eta^{ss}\bar{Y}\int(p^i-p)^2di.$$

Using the definition of nominal expenditure $\pi\equiv p+y$, we have the welfare measure $(\pi-p-y^*)^2+\lambda\int(p_i-p)^2di$, where $\lambda\equiv\frac{V'(\bar{Y})\eta^{ss}}{[V''(\bar{Y})-U''(\bar{Y})]\bar{Y}}=\eta^{ss}/\alpha>0$.

Notice that there is a wedge between the central bank and firm's actions due to the existence of aggregate markup shock. Consider the case that there is only aggregate shocks. Central bank would prefer no learning and no reaction to the markup shock, because the markup shock generates inefficient output fluctuations and heterogeneous beliefs of it creates inefficient price dispersion. However, to minimize their own loss, firms would optimally choose to react to the markup shocks. It is the wedge that drives the different results under the setups.

3 Information Process

Firms do not have perfect information on the aggregate shock, $\tilde{\theta}\sim\mathcal{N}(0,\sigma_\theta^2)$. In addition, through private learning, each firm i gets a private signal $\theta_i\sim\mathcal{N}(\tilde{\theta},\sigma_i^2)$ where the magnitude of σ_i^2 depends on the learning effort of firm i. From firm i's perspective, Bayes rule implies that the posterior distribution of $\tilde{\theta}$ is,

$$\begin{aligned} P(\tilde{\theta}|\theta_i) &\propto P(\theta_i|\tilde{\theta})P(\tilde{\theta}) \\ &\propto \exp\left[\frac{-(\theta_i-\tilde{\theta})^2}{2\sigma_i^2}\right]\cdot\exp\left[\frac{-(\tilde{\theta}-0)^2}{2\sigma_\theta^2}\right] = \exp\left(-\frac{\sigma_\theta^2\left(\theta_i-\tilde{\theta}\right)^2+\sigma_i^2\tilde{\theta}^2}{2\sigma_i^2\sigma_\theta^2}\right), \end{aligned}$$

with mean $\hat{\theta}_i = \frac{\sigma_\theta^2}{\sigma_i^2+\sigma_\theta^2}\theta_i$ and variance $\frac{\sigma_i^2\sigma_\theta^2}{\sigma_i^2+\sigma_\theta^2}$. The mutual information is defined as $I(\tilde{\theta}|0, \tilde{\theta}|\{0,\theta_i\}) = 0.5x$. It is then calculated as:

$$H(\tilde{\theta}|0) - H(\tilde{\theta}|\{0,\theta_i\}) = \frac{1}{2}\ln\left(\frac{\sigma^2_{\tilde{\theta}|0}}{\sigma^2_{\tilde{\theta}|\{0,\theta_i\}}}\right) = 0.5x \rightarrow \frac{\sigma_i^2}{\sigma_i^2+\sigma_\theta^2} = e^{-x},$$

where x is the (logarithm of the) reduction in variance from devoting attention to learning more about $\tilde{\theta}$. Thus, the expected $\tilde{\theta}$ for firm i is $\hat{\theta}_i = (1-e^{-x})\theta_i$.

Similarly, firm i gets a private signal $e_i \sim \mathcal{N}(u_i, \sigma_e^2)$ about idiosyncratic shock $u_i \sim \mathcal{N}(0, \phi^2)$. The expected u_i becomes $\hat{u}_i = (1-e^{-z})e_i$ and the posterior variance is $\phi^2 e^{-z}$, with $e^{-z} = \frac{\sigma_e^2}{\sigma_e^2+\phi^2}$.

The following equations will be applied for the derivations in later sections:

$$\int (\hat{\theta}_i - \tilde{\theta})^2 di = e^{-x}\sigma_\theta^2; \tag{1}$$

$$\int (\theta_i - \tilde{\theta})^2 di = \sigma_i^2 = \frac{e^{-x}}{1-e^{-x}}\sigma_\theta^2; \tag{2}$$

$$\int (\hat{u}_i - u_i)^2 di = \phi^2 e^{-z}; \tag{3}$$

$$\int \hat{u}_i^2 di = (1-e^{-z})^2 \int e_i^2 di = (1-e^{-z})^2(\phi^2+\sigma_e^2) = \phi^2(1-e^{-z}). \tag{4}$$

4 Solution to the Inelastic Attention Case

Through their choice on $\{x, z\}$, firms can control their information accuracy on the aggregate shock and the idiosyncratic shock. It is assumed that firms face a constraint on learning capacity, $x + z \leq 2k$.

To solve the optimal learning decision $\{x, z\}$, a guess about the aggregate action a is made, following the method of Reis (2011). With complete information, the aggregate action would be perfectly consistent with monetary policy, i.e. $p = \pi - y^* + \frac{1}{\alpha}\tilde{\theta}$. With incomplete information, we make a starting guess $p = \beta_0 + (\beta_1 + \frac{1}{\alpha})\gamma\tilde{\theta} - y^*$, where γ is a variable governed by the learning effort, which value is to be determined later. $\gamma\tilde{\theta}$ identifies general public's posterior of the aggregate shock $\tilde{\theta}$. Thus

$$p_i = (1-\alpha)\left\{\beta_0 + \left(\beta_1 + \frac{1}{\alpha}\right)\gamma\hat{\theta}_i - y^*\right\} + \alpha\left(\beta_0 + \beta_1\hat{\theta}_i\right) + \hat{\theta}_i - \alpha y^* + \hat{u}_i$$

$$= \beta_0 - y^* + (1-\alpha)\left(\beta_1 + \frac{1}{\alpha}\right)\gamma\hat{\theta}_i + \alpha\left(\beta_1 + \frac{1}{\alpha}\right)\hat{\theta}_i + \hat{u}_i$$

$$= \beta_0 - y^* + \left(\beta_1 + \frac{1}{\alpha}\right)[(1-\alpha)\gamma + \alpha]\hat{\theta}_i + \hat{u}_i.$$

$$\int p_i di = \beta_0 - y^* + \left(\beta_1 + \frac{1}{\alpha}\right)[(1-\alpha)\gamma + \alpha](1-e^{-x})\tilde{\theta}.$$

Combining with the results that $p = \beta_0 + \left(\beta_1 + \frac{1}{\alpha}\right)\gamma\tilde{\theta} - y^*$ and $\int p_i di = p$, the above implies

$$\left(\beta_1 + \frac{1}{\alpha}\right)\gamma = \left(\beta_1 + \frac{1}{\alpha}\right)[(1-\alpha)\gamma + \alpha](1-e^{-x}).$$

The above equation can be simplified as

$$\gamma = [(1-\alpha)\gamma + \alpha](1-e^{-x}).$$

Thus,

$$\gamma = \frac{\alpha(1-e^{-x})}{e^{-x} + \alpha(1-e^{-x})} = 1 - \frac{1}{\alpha e^x + (1-\alpha)} = \frac{\alpha e^x - \alpha}{\alpha e^x + (1-\alpha)}$$

and

$$1 - \gamma = \frac{1}{\alpha e^x + (1-\alpha)}. \tag{5}$$

Substituting the representation of p_i, p and π to firm's problem, we obtain

$$\min \mathbb{E}_i\left[p_i - (1-\alpha)p - \alpha(\pi - y^*) - \tilde{\theta} - u_i\right]^2$$

$$= \left\{\left(\beta_1 + \frac{1}{\alpha}\right)[(1-\alpha)\gamma + \alpha]\hat{\theta}_i + \hat{u}_i - (1-\alpha)\left(\beta_1 + \frac{1}{\alpha}\right)\gamma\tilde{\theta} - \alpha\beta_1\tilde{\theta} - \tilde{\theta} - u_i\right\}^2$$

$$= \int\left\{\left[\left(\beta_1 + \frac{1}{\alpha}\right)[(1-\alpha)\gamma + \alpha](\hat{\theta}_i - \tilde{\theta})\right]^2 + (\hat{u}_i - u_i)^2\right\} di.$$

Substituting the posterior variance of shocks,

$$\min \mathbb{E}_i\left[p_i - (1-\alpha)p - \alpha(\pi - y^*) - \tilde{\theta} - u_i\right]^2$$

$$= \left(\beta_1 + \frac{1}{\alpha}\right)^2 [(1-\alpha)\gamma + \alpha]^2 e^{-x}\sigma_\theta^2 + e^{-z}\phi^2$$

$$= \left(\beta_1 + \frac{1}{\alpha}\right)^2 [(1-\alpha)\gamma + \alpha]^2 e^{-x}\sigma_\theta^2 + \phi^2 e^{-2k}e^x,$$

so

$$x = \max\left\{\ln\left(\frac{\left(\beta_1 + \frac{1}{\alpha}\right)[(1-\alpha)\gamma + \alpha]\sigma_\theta e^k}{\phi}\right), 0\right\}. \tag{6}$$

When $x = 0$, $\gamma = 0$. When $x \neq 0$, applying equations (5) and (6). Thus,

$$\gamma = \begin{cases} 1 - \frac{\phi}{\alpha(\beta_1 + \frac{1}{\alpha})\sigma_\theta e^k}, & \text{if } \phi < \alpha(\beta_1 + \frac{1}{\alpha})\sigma_\theta e^k, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Firms may or may not learn the aggregate shock. In particular, by choosing the value β_1, the central bank may be able to influence firms' decision to learn. We will come back to this point when discussing the optimal monetary policy below.

Welfare loss calculation: First, consider price dispersion. Applying equations (5) and (7)

$$
\begin{aligned}
\int (p_i - p)^2 di &= \int \left\{ \left(\beta_1 + \frac{1}{\alpha}\right) \left[(1-\alpha)\gamma + \alpha\right] (1 - e^{-x})(\theta_i - \tilde{\theta}) + \hat{u}_i \right\} di \\
&= \left(\beta_1 + \frac{1}{\alpha}\right)^2 \gamma^2 \sigma_i^2 + \int \hat{u}_i^2 di \\
&= \left(\beta_1 + \frac{1}{\alpha}\right)^2 \gamma^2 \frac{e^{-x}}{1 - e^{-x}} \sigma_\theta^2 + \int \hat{u}_i^2 di \\
&= \frac{\phi^2}{e^{2k}\alpha^2} \frac{\gamma^2}{(1-\gamma)^2} \frac{e^{-x}}{1-e^{-x}} + \phi^2 (1 - e^{-2k+x}) \\
&= \frac{\phi^2}{e^{2k}} (e^x - 1) + \phi^2 (1 - e^{-2k+x}) \\
&= \phi^2 \left(1 - e^{-2k}\right).
\end{aligned}
$$

Output gap follows

$$
\begin{aligned}
(\pi - p - y^*)^2 &= \left[\beta_0 + \beta_1 \tilde{\theta} - \beta_0 + y^* - \left(\beta_1 + \frac{1}{\alpha}\right) \gamma \tilde{\theta} - y^*)\right]^2 \\
&= \left\{\beta_1 \tilde{\theta} - \left(\beta_1 + \frac{1}{\alpha}\right) \gamma \tilde{\theta} \right\}^2 = \left[\beta_1 - \left(\beta_1 + \frac{1}{\alpha}\right)\gamma\right]^2 \sigma_\theta^2.
\end{aligned}
$$

When there is learning on aggregate shock: Equation (7) implies that $\beta_1 - (\beta_1 + \frac{1}{\alpha})\gamma = \frac{\phi}{\alpha\sigma_\theta e^k} - \frac{1}{\alpha}$. Output gap is independent of monetary policy:

$$
(\pi - p - y^*)^2 = \left[-\frac{1}{\alpha} + \frac{\phi}{\alpha \sigma_\theta e^k}\right]^2 \sigma_\theta^2.
$$

Thus, total welfare loss in this scenario follows,

$$
(\pi - p - y^*)^2 + \lambda \int (p_i - p)^2 di = \left[-\frac{1}{\alpha} + \frac{\phi}{\alpha \sigma_\theta e^k}\right]^2 \sigma_\theta^2 + \lambda \phi^2 \left(1 - e^{-2k}\right).
$$

Welfare loss is independent of monetary policy, and the choice of β_1 does not matter, as long as under that β_1 firms learn. That is, we need the choice to make sure that $\phi < \alpha(\beta_1 + \frac{1}{\alpha})\sigma_\theta e^k$, or any β_1 such that $\beta_1 > \frac{\phi}{\alpha\sigma_\theta e^k} - \frac{1}{\alpha}$.

Output gap is created by the wedge between aggregate nominal expenditure π and the aggregate action of firms p, and as a result it is related to the reaction to $\tilde{\theta}$. When π is more responsive to $\tilde{\theta}$, p also gets more responsive to $\tilde{\theta}$. Due to the existence of total learning capacity constraint ($x + z \leq 2k$), the wedge is fixed and it is independent of policy.

Price dispersion comes from two sources: dispersed reaction to the aggregate condition and to the idiosyncratic condition. Price dispersion increases with both x and z. First, information accuracy increases with x, so heterogeneous belief weakens with x. However, firms are more responsive to the aggregate condition when information is more accurate, which further increases price dispersion due to the reaction to the aggregate condition. Ultimately, the second effect dominates, and price dispersion increases with x. Price dispersion also increases with z, which governs the information accuracy on the idiosyncratic condition and thus the sensitivity of the action to it. Given the learning capacity constraint, we have x and z negatively correlated. Thus, in the end, we have price dispersion being independent of the learning effort and monetary policy.

When there is no learning on aggregate shock: Under this case we have

$$\int (p_i - p)^2 di = \phi^2(1 - e^{-2k}), \qquad (\pi - p - y^*)^2 = [\beta_1 \tilde{\theta}]^2 = \beta_1^2 \sigma_\theta^2.$$

Price dispersion is independent of monetary policy, so monetary policy can focus on minimizing the other component. The optimal choice is $\beta_1 = 0$, reducing that component to zero. Referring back to equation (7), this case satisfies the condition that $\phi \geq \sigma_\theta e^k$.

Now we can present the results by combining the two cases above. The total welfare loss in the inelastic attention setup under the optimal monetary policy follows

$$(\pi - p - y^*)^2 + \lambda \int (p_i - p)^2 di = \beta_1^2 \sigma_\theta^2 + \lambda \phi^2 (1 - e^{-2k})$$

$$= \begin{cases} \lambda \phi^2 (1 - e^{-2k}) & \text{if } \phi \geq \sigma_\theta e^k, \\ \left(\frac{\phi}{\alpha \sigma_\theta e^k} - \frac{1}{\alpha} \right)^2 \sigma_\theta^2 + \lambda \phi^2 (1 - e^{-2k}), & \text{otherwise.} \end{cases}$$

Therefore, the optimal monetary policy in this case is,

$$\beta_1 = \begin{cases} 0, & \text{if } \phi \geq \sigma_\theta e^k; \\ > \frac{\phi}{\alpha \sigma_\theta e^k} - \frac{1}{\alpha}, & \text{otherwise.} \end{cases} \tag{8}$$

The two cases in the above results corresponding to some learning and no learning on the aggregate shock respectively. The second inequality means that when it is optimal to have learning on the aggregate condition, any β_1 that leads to learning is optimal.

If we pick the optimal policy that stabilizes the price level the most, we have the inequality in Equation 8 being equality, i.e. $\beta_1 = \frac{\phi}{\alpha \sigma_\theta e^k} - \frac{1}{\alpha}$ if $\phi < \sigma_\theta e^k$.

Therefore, the choice depends on whether $\phi \geq \sigma_\theta e^k$. In words, the condition says that the idiosyncratic shock is relatively imprecise, meaning firms find it more attractive to learn. If the condition is met, the optimal policy choice is not to respond to the aggregate shock at all ($\beta_1 = 0$) and hence there is no learning on the aggregate shock. Firms focus on the idiosyncratic shock and social loss is $\lambda \phi^2 (1 - e^{-2k})$. If we have $\phi < \sigma_\theta e^k$ instead, there is learning on the aggregate shock even when $\beta_1 = 0$. The central bank should then respond to the aggregate shock in any way such that the learning is positive, i.e., by choosing $\beta_1 \geq \frac{\phi}{\alpha \sigma_\theta e^k} - \frac{1}{\alpha}$. If we pick the optimal policy that stabilizes the aggregate price level the most, then we have $\beta_1 = \frac{\phi}{\alpha \sigma_\theta e^k} - \frac{1}{\alpha}$.

5 Solution to the Elastic Case

Firm i's problem is

$$\min[p_i - (1-\alpha)\,p - \alpha(\pi - \tilde{\theta}) - u_i]^2 + \mu k.$$

We can follow the same steps as in the previous section to obtain

$$\min\left(\beta_1 + \frac{1}{\alpha}\right)^2 [(1-\alpha)\gamma + \alpha]^2 \sigma_\theta^2 e^{-x} + \phi^2 e^{-z} + \mu 0.5(x+z),$$

where $k \equiv I(\{0, \tilde{\theta}\}, \theta_i) + I(u_i, e_i)$ and $x + z = 2k$. From now on the elastic case differs from the inelastic case. Differentiating the objective function with respect to x and z, we get:

$$x = \ln\left(\frac{\left(\beta_1 + \frac{1}{\alpha}\right)^2 [(1-\alpha)\gamma + \alpha]^2 \sigma_\theta^2}{0.5\mu}\right), z = \ln\left(\frac{\phi^2}{0.5\mu}\right).$$

Notice that z only depends on the variance of the idiosyncratic shock and the cost of learning, and it is not affected by monetary policy. The expression of x is similar to that in the inelastic case.

Welfare loss calculation: We can write down each component as

$$\int (p_i - p)^2\, di = \left(\beta_1 + \frac{1}{\alpha}\right)^2 \gamma^2 \sigma_\theta^2 \frac{e^{-x}}{1 - e^{-x}} + \phi^2(1 - e^{-z}),$$

$$(\pi - p - y^*)^2 = \left\{\beta_1 \tilde{\theta} - (\beta_1 + \frac{1}{\alpha})\gamma\tilde{\theta}\right\}^2 = \left[\beta_1 - (\beta_1 + \frac{1}{\alpha})\gamma\right]^2 \sigma_\theta^2.$$

If there is no learning on the aggregate condition, then $\gamma = 0$ and the optimal monetary policy has $\beta_1 = 0$. And $\beta_1 = 0$ indeed leads to no learning, if $\mu > 2\sigma_\theta^2$. Otherwise, we proceed to the following derivation,

We can solve from the above result for x that:

$$(\beta_1 + 1/\alpha)\gamma = \frac{\sqrt{0.5\mu e^x}}{\sigma_\theta}(1 - e^{-x}),$$

$$\beta_1 = \frac{\sqrt{0.5\mu e^x}}{\sigma_\theta}\frac{(\alpha + e^{-x} - \alpha e^{-x})}{\alpha} - \frac{1}{\alpha} = \frac{\sqrt{0.5\mu e^x}}{\sigma_\theta}(1 - e^{-x} + e^{-x}/\alpha) - \frac{1}{\alpha}.$$

Output gap can then be expressed as

$$(\pi - p - y^*)^2 = \left[\sqrt{0.5\mu e^x}\frac{e^{-x}}{\alpha\sigma_\theta} - \frac{1}{\alpha}\right]^2 \sigma_\theta^2 = \left[\frac{\sqrt{0.5\mu e^{-x}}}{\alpha\sigma_\theta} - \frac{1}{\alpha}\right]^2 \sigma_\theta^2, = \left[\frac{\sqrt{0.5\mu}}{\sigma_\theta\alpha}T - \frac{1}{\alpha}\right]^2 \sigma_\theta^2,$$

where we define $T \equiv e^{-0.5x} \in [0, 1]$. Similarly, price dispersion can be written in terms of T:

$$\int (p_i - p)^2\, di = \left(\beta_1 + \frac{1}{\alpha}\right)^2 \gamma^2 q \sigma_\epsilon^2 \frac{e^{-x}}{1 - e^{-x}} + \phi^2(1 - e^{-z})$$

$$= \lambda 0.5\mu(1 - e^{-x}) + \lambda\phi^2(1 - e^{-z}) = \lambda 0.5\mu(1 - T^2) + \lambda\phi^2(1 - e^{-z}).$$

Combining them both gives us the welfare loss:

$$(\pi - p - y^*)^2 + \lambda \int (p_i - p)^2 \, di$$

$$= \left[\frac{\sqrt{0.5\mu}}{\alpha\sigma_\theta} T - \frac{1}{\alpha} \right]^2 \sigma_\theta^2 + \lambda 0.5\mu(1 - T^2) + \lambda\phi^2(1 - e^{-z}). \tag{9}$$

Notice only the terms in the welfare loss that contain T are affected by monetary policy. Optimal monetary policy finds the T, which implies some level of learning x, that satisfies the first order condition:

$$2\frac{\sqrt{0.5\mu}}{\alpha\sigma_\theta} \left[\frac{\sqrt{0.5\mu}}{\alpha\sigma_\theta} T - \frac{1}{\alpha} \right] \sigma_\theta^2 - \lambda\mu T = 0$$

$$T = \frac{2\sqrt{0.5\mu}\sigma_\theta}{\mu(1 - \lambda\alpha^2)}$$

Remember that T is bounded between 0 and 1, so the solution above only works for some specific parameter values. More generally, we have

$$T = \begin{cases} 1, & \text{if } \lambda\alpha^2 > 1 - \frac{2}{\sqrt{0.5\mu}}\sigma_\theta \text{ (Corner Condition)}, \\ \frac{2\sqrt{0.5\mu}\sigma_\theta}{\mu(1-\lambda\alpha^2)}, & \text{if } \lambda\alpha^2 \le 1 - \frac{2}{\sqrt{0.5\mu}}\sigma_\theta, \end{cases} \tag{10}$$

However, when $\mu \le 2\sigma_\theta^2$, we always have the corner solution. Thus, we have $T = 1$.

The optimal β_1 can be backed out from the solution for T

$$\beta_1 = \begin{cases} \le \frac{\sqrt{0.5\mu}}{\sigma_\theta\alpha} - \frac{1}{\alpha}, & \text{if } \mu \le 2\sigma_\theta^2, \\ 0, & \text{if } \mu > 2\sigma_\theta^2. \end{cases} \tag{11}$$

When it is optimal to have no learning (i.e., corner conditions 1 and 2 in equation (10)), any level of β_1 that leads to no learning on the aggregate condition is optimal.

If we pick the optimal policy that stabilizes the price level the most, we have the inequalities in Equation 11 being equality, i.e. $\beta_1 = \frac{\sqrt{0.5\mu}}{\sigma_\theta\alpha} - \frac{1}{\alpha}$ in the two corner conditions.

6 Without Idiosyncratic Shock

If there is no idiosyncratic shock or firms have perfect knowledge on the shock, the model is similar to Adam (2007).

6.1 Inelastic Attention

We have $x = 2k$ and $\gamma = 1 - \frac{1}{\alpha e^{2k} + 1 - \alpha}$, and welfare loss becomes

$$(\pi - p - y^*)^2 + \lambda \int (p_i - p)^2 di$$

$$= \left[\beta_1 - \left(\beta_1 + \frac{1}{\alpha} \right) \gamma \right]^2 \sigma_\theta^2 + \lambda \left(\beta_1 + \frac{1}{\alpha} \right)^2 \gamma^2 \frac{e^{-x}}{1 - e^{-x}} \sigma_\theta^2 + t.i.p$$

$$= \left(\beta_1 + \frac{1}{\alpha} \right)^2 \gamma^2 \sigma_\theta^2 - 2\beta_1 \left(\beta_1 + \frac{1}{\alpha} \right) \gamma\sigma_\theta^2 + \beta_1^2\sigma_\theta^2 + \lambda \left(\beta_1 + \frac{1}{\alpha} \right)^2 \gamma^2 \frac{e^{-x}}{1 - e^{-x}} \sigma_\theta^2 + t.i.p.$$

The first order condition gives us

$$\beta_1 = \frac{1}{\alpha}\frac{1-\gamma}{(1-\gamma)^2 + \lambda\gamma^2 e^{-2k}/(1-e^{-2k})} - \frac{1}{\alpha} = \frac{(1-\lambda\alpha)(e^{2k}-1)}{1+\lambda\alpha^2(e^{2k}-1)}.$$

It is straightforward to show that the optimal monetary policy is identical to the one in Adam (2007).

6.2 Elastic Attention

Firm i's problem is the same as before and we have

$$x = \ln\left(\frac{\left(\beta_1 + \frac{1}{\alpha}\right)^2 [(1-\alpha)\gamma + \alpha]^2 \sigma_\theta^2}{0.5\mu}\right).$$

The welfare loss and the implied optimal monetary policy are the same as the case with idiosyncratic shock, and it is also the same as the case with inelastic attention. Plugging in $x = 2k$ to $\beta_1 = \frac{\sqrt{0.5\mu e^x}}{\sigma_\theta}\left(1 - e^{-x} + \frac{e^{-x}}{\alpha}\right) - \frac{1}{\alpha}$, we have

$$\beta_1 = \frac{\alpha e^{2k} + 1 - \alpha}{\alpha(1-\lambda\alpha^2)},$$

7 Other Results

Result 1. Learning on the idiosyncratic condition z depends on the monetary policy in the inelastic case, but not in the elastic case. In both cases, we have $x = 0$ under the optimal monetary policy. A non-zero response to the markup shock causes inefficient consumption fluctuations and inefficient price dispersion. Thus, the Central Bank try to minimize the learning incentive on the aggregate shock. Correspondingly, we have $z = 2k$ in the inelastic attention case, given the attention capacity constraint. However, in the elastic case, learning on the idiosyncratic shock is completely indepdent of the monetary policy, since information processing on the two conditions are completely separate.

Result 2. In both cases, optimal monetary policy incentivize no learning on the markup shock. Let's go back to the definition of the welfare function. First, output gap depends on a wedge between firms' decision and the central bank's decision:

$$\text{Output gap: } (\pi - p - y^*)^2 = \left[\beta_1\tilde{\theta} - \left(\beta_1 + \frac{1}{\alpha}\right)\gamma\tilde{\theta}\right]^2.$$

Comparing the two terms in the square bracket, we find monetary policy and firms' action having different loadings on $\tilde{\theta}$. Output gap is minimized when $\gamma = 0$, i.e. no learning and no reaction to the markup shock.

Second, price dispersion is affected by heterogeneous beliefs on the aggregate shock and the dispersed reaction to the idiosyncratic shock. Under inelastic attention, price dispersion is independent of monetary policy and equal to

$$\text{Price dispersion (inelastic): } \lambda \int (p_i - p)^2 di = \lambda\phi^2\left(1 - e^{-2k}\right).$$

It increases with k, the learning capacity. Under elastic attention, price dispersion depends on x and z, thus the policy:

$$\text{Price dispersion (elastic): } \lambda \int (p_i - p)^2 di = \lambda \left[0.5\mu(1 - e^{-x}) + \phi^2(1 - e^{-z}) \right].$$

The first term in the square bracket comes from heterogeneous beliefs on the aggregate condition. Information accuracy increases with x, and that means dispersed reactions weaken with it too. But the first term also increases with the sensitivity of firms' actions to the aggregate condition. The two effects counteract each other. The latter effect dominates, and hence the first term increases with x. The second term in the square bracket comes from dispersed reactions due to the idiosyncratic condition. The more accurate the information on the idiosyncratic condition, the more responsive firms' action is to it. Thus, the second term increases with z. However, z is independent of the policy.

Therefore, when $\beta_1 = 0$ generates $x = 0$, then the optimal policy sets $\beta_1 = 0$. Otherwise, a level of β_1 with the smallest absolute value that creates $x = 0$ is picked. It is easy to show that that level of β_1 is the same in the two cases when the optimal $\beta_1 \neq 0$.[1]

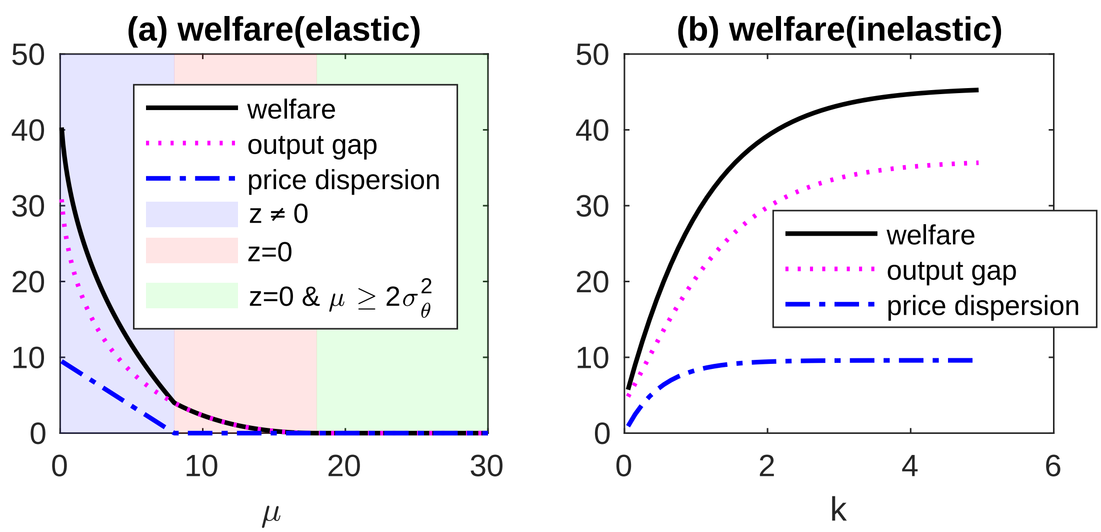
Result 3. Under certain conditions, β_1 is negative in the elastic attention case, while being zero or non-exist in the inelastic attention case. Figure 1 in the paper shows how β_1 changes with the cost of learning μ. When the cost is low, the two cases are equivalent (and as we shall see, the choices of x and z are also the same), and the coefficient for optimal monetary policy is negative. The intuition is that, in order to make firms not to learn about the aggregate shock ($x = 0$), the central bank acts in a way that firms have no incentive to learn. It can be seen from the optimization problem solved by the firms (equation 8): as β gets more and more negative, the weight put on the markup shock decreases. When learning is cheap, the central bank needs to make the shock itself less relevant. Algebraically, the central bank has to set β_1 so that firms hit the corner solution and choose x. As μ goes up, the optimal coefficient increases. When μ is large enough that creates zero z, the total learning is zero. In the inelastic case, we set $k = x + z = 0$ and thus $\beta_1 = 0$. However, in the elastic case, we sill have negative β_1 unless $\mu > 2\sigma_\theta^2$. In fact, the difference of β_1 comes from the condition differences in Equations 7 and 8 in the paper.

Result 4. Figure 1 (c) in the paper plots the welfare loss as a function of μ. Clearly, perfect information creates much higher levels of inefficient fluctuation than the cases with imperfect inforamtion, a result consistent with the existing literature (Adam and Paciello). If there is no information friction on aggregate condition, welfare loss is $\sigma_\theta^2/\alpha^2 + \lambda\phi^2$ as the dotted blue line.

Further, we decompose the welfare loss to output gap and price dispersion in Figure 1. Panel (a) presents the elastic attention case. Both price dispersion and output gap decrease with μ, since z decreases with μ. Panel (b) presents the welfare decomposition for the inelastic case as a function of k. Both price dispersion and output gap increase with μ.

[1] When $\mu \geq 2\phi^2$, we have $z = ln(\phi^2/0.5\mu)$. Taking a value of $k = 0.5z$, we have $\phi/\alpha\sigma_\theta e^k - 1/\alpha = \sqrt{0.5\mu}/\alpha\sigma_\theta - 1/\alpha$.

Figure 1: Welfare Decomposition



8 With Public Signal

Firms still do not have perfect information on the aggregate shock, $\tilde{\theta} \sim \mathcal{N}(0, \sigma_\theta^2)$, but now they also receive a noisy public signal $\theta^P \sim \mathcal{N}(\tilde{\theta}, \sigma_\epsilon^2)$. Again, through private learning, firm i gets a private signal θ_i, $\theta_i \sim \mathcal{N}(\tilde{\theta}, \sigma_i^2)$, where the magnitude of σ_i^2 depends on the learning effort of the firm. From firm i's perspective, Bayes rule implies that the posterior distribution of $\tilde{\theta}$ follows

$$\begin{aligned} P(\tilde{\theta}|\theta_i, \theta^P) &\propto P(\theta_i|\tilde{\theta})P(\theta^P|\tilde{\theta})P(\tilde{\theta}) \\ &\propto \exp\left[\frac{-(\theta_i - \tilde{\theta})^2}{2\sigma_i^2}\right] \cdot \exp\left[\frac{-(\theta^P - \tilde{\theta})^2}{2\sigma_\epsilon^2}\right] \cdot \exp\left[\frac{-(\tilde{\theta} - 0)^2}{2\sigma_\theta^2}\right] \\ &= \exp\left(-\frac{\sigma_\epsilon^2\sigma_\theta^2(\theta_i - \tilde{\theta})^2 + \sigma_i^2\sigma_\theta^2(\theta^P - \tilde{\theta})^2 + \sigma_\epsilon^2\sigma_i^2\tilde{\theta}^2}{2\sigma_i^2\sigma_\epsilon^2\sigma_\theta^2}\right), \end{aligned}$$

with mean,

$$\mu_{\tilde{\theta}|\theta_i,\theta} = \frac{\sigma_\epsilon^2\sigma_\theta^2}{\sigma_\epsilon^2\sigma_i^2 + \sigma_\epsilon^2\sigma_\theta^2 + \sigma_\theta^2\sigma_i^2}\theta_i + \frac{\sigma_i^2\sigma_\theta^2}{\sigma_\epsilon^2\sigma_i^2 + \sigma_\epsilon^2\sigma_\theta^2 + \sigma_\theta^2\sigma_i^2}\theta^P,$$

and variance,

$$var_{\tilde{\theta}|\theta_i,\theta} = \frac{\sigma_\epsilon^2\sigma_i^2\sigma_\theta^2}{\sigma_\epsilon^2\sigma_i^2 + \sigma_\epsilon^2\sigma_\theta^2 + \sigma_\theta^2\sigma_i^2}.$$

Without private learning, posterior mean is simply $\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\epsilon^2}\theta^P \equiv q\theta^P$ and variance is $\frac{\sigma_\theta^2\sigma_\epsilon^2}{\sigma_\theta^2+\sigma_\epsilon^2}$. Limitted attention as a bound on information flow gives $I(\tilde{\theta}|\theta^P, \tilde{\theta}|\{\theta^P, \theta_i\}) = 0.5x$, with which we can calculate the mutual entropy:

$$H(\tilde{\theta}|\theta^P) - H(\tilde{\theta}|\{\theta^P, \theta_i\}) = \frac{1}{2}\ln\left(\frac{\sigma^2_{\tilde{\theta}|\theta^P}}{\sigma^2_{\tilde{\theta}|\{\theta^P,\theta_i\}}}\right) = 0.5x,$$

$$\frac{\sigma_i^2(\sigma_\theta^2 + \sigma_\epsilon^2)}{\sigma_\epsilon^2\sigma_i^2 + \sigma_\epsilon^2\sigma_\theta^2 + \sigma_\theta^2\sigma_i^2} = e^{-x}.$$

x is the (logarithm of the) reduction in variance from devoting attention to learning more about θ^P. Thus, the posterior of firm i's beliefs on $\tilde{\theta}$ is a linear interpolation of the private and the public signals,

$$\mu_{\tilde{\theta}} = (1-e^{-x})\theta_i + e^{-x}\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\epsilon^2}\theta^P = (1-e^{-x})\theta_i + e^{-x}q\theta^P \tag{12}$$

where $q \equiv \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\epsilon^2}$. Alternatively, we can represent in terms of accuracy

$$e^{-x} = \frac{\sigma_\epsilon^{-2}+\sigma_\theta^{-2}}{\sigma_\epsilon^{-2}+\sigma_\theta^{-2}+\sigma_i^{-2}},$$

and the posterior variance is

$$var_{\tilde{\theta}|\theta_i,\theta} = \left(\sigma_\epsilon^{-2}+\sigma_\theta^{-2}+\sigma_i^{-2}\right)^{-1}.$$

The sum of attention satisfies $x+z \le 2k$.

Monetary policy now is a linear function of state variables $\tilde{\theta}$ and θ^P, $\pi = \beta_0 + \beta_1\tilde{\theta} + \beta_2\theta^P$, where β_0, β_1, and β_2 are the policy choices. To solve for the optimal learning decision $\{x, z\}$, a guess about the aggregate action (p) is made. With complete information, the aggregate action would be perfectly consistent with monetary policy, i.e. $p = \pi - y^* + \frac{1}{\alpha}\tilde{\theta}$. With incomplete information, the following starting guess is made:

$$p = \beta_0 + \left(\beta_1 + \frac{1}{\alpha}\right)[\gamma\tilde{\theta} + (1-\gamma)q\theta] + \beta_2\theta - y^*,$$

where γ is a variable governed by the learning effort, the value of which is to be determined later. Also, $\gamma\tilde{\theta} + (1-\gamma)q\theta$ identifies firms' posterior of the aggregate shock $\tilde{\theta}$. Thus

$$\begin{aligned}
p_i &= (1-\alpha)\left\{\beta_0 + \left(\beta_1+\frac{1}{\alpha}\right)[\gamma\hat{\theta}_i + (1-\gamma)q\theta] + \beta_2\theta - y^*\right\} + \alpha\left(\beta_0 + \beta_1\hat{\theta}_i + \beta_2\theta\right) + \hat{\theta}_i - \alpha y^* + \hat{u}_i \\
&= \beta_0 + \beta_2\theta - y^* + (1-\alpha)\left(\beta_1+\frac{1}{\alpha}\right)[\gamma\hat{\theta}_i + (1-\gamma)q\theta] + \alpha\left(\beta_1+\frac{1}{\alpha}\right)\hat{\theta}_i + \hat{u}_i \\
&= \beta_0 + \beta_2\theta - y^* + \left(\beta_1+\frac{1}{\alpha}\right)[(1-\alpha)\gamma+\alpha]\hat{\theta}_i + \left(\beta_1+\frac{1}{\alpha}\right)(1-\alpha)(1-\gamma)q\theta + \hat{u}_i,
\end{aligned}$$

and

$$\begin{aligned}
\int p_i di &= \beta_0 + \beta_2\theta - y^* + \left(\beta_1+\frac{1}{\alpha}\right)[(1-\alpha)\gamma+\alpha](1-e^{-x})\tilde{\theta} \\
&+ \left(\beta_1+\frac{1}{\alpha}\right)[(1-\alpha)\gamma+\alpha]e^{-x}q\theta + (1-\alpha)\left(\beta_1+\frac{1}{\alpha}\right)(1-\gamma)q\theta.
\end{aligned}$$

Combining with $p = \beta_0 + \left(\beta_1+\frac{1}{\alpha}\right)[\gamma\tilde{\theta} + (1-\gamma)q\theta] + \beta_2\theta - y^*$ and $\int p_i di = p$, we obtain

$$(\beta_1 + \frac{1}{\alpha})\gamma = \left(\beta_1+\frac{1}{\alpha}\right)[(1-\alpha)\gamma+\alpha](1-e^{-x}), \tag{13}$$

$$\left(\beta_1+\frac{1}{\alpha}\right)(1-\gamma)q = \left(\beta_1+\frac{1}{\alpha}\right)[(1-\alpha)\gamma+\alpha]e^{-x}q + \left(\beta_1+\frac{1}{\alpha}\right)(1-\alpha)(1-\gamma)q. \tag{14}$$

It is straightforward to prove that if one of the above two equations holds, the other also holds. The above two equations can be simplified as

$$\gamma = [(1-\alpha)\gamma + \alpha]\,(1-e^{-x}),$$

$$(1-\gamma) = [(1-\alpha)\gamma + \alpha]\,e^{-x} + (1-\alpha)(1-\gamma).$$

As a result, we have

$$\gamma = \frac{\alpha(1-e^{-x})}{e^{-x}+\alpha(1-e^{-x})} = 1 - \frac{e^{-x})}{e^{-x}+\alpha(1-e^{-x})} = 1 - \frac{1}{\alpha e^{x} + (1-\alpha)} = \frac{\alpha e^{x} - \alpha}{\alpha e^{x} + (1-\alpha)}$$

and

$$1-\gamma = \frac{1}{\alpha e^{x} + (1-\alpha)}.$$

8.1 Inelastic attention

Substituting p_i and equation (12) to firm i's problem we have

$$\min \mathbb{E}_i \left[p_i - (1-\alpha)p - \alpha(\pi - y^*) - \tilde{\theta} - u_i\right]^2$$

$$= \left\{ \left(\beta_1 + \frac{1}{\alpha}\right)[(1-\alpha)\gamma + \alpha]\hat{\theta}_i + \left(\beta_1 + \frac{1}{\alpha}\right)(1-\alpha)(1-\gamma)q\theta + w\hat{u}_i \right.$$

$$\left. -(1-\alpha)\left(\beta_1 + \frac{1}{\alpha}\right)\gamma\tilde{\theta} - (1-\alpha)\left(\beta_1 + \frac{1}{\alpha}\right)(1-\gamma)q\theta - \alpha\beta_1\tilde{\theta} - \tilde{\theta} \right\}^2$$

$$= \int \left\{ \left[\left(\beta_1 + \frac{1}{\alpha}\right)[(1-\alpha)\gamma + \alpha]\,(\hat{\theta}_i - \tilde{\theta})\right]^2 + (\hat{u}_i - u_i)^2 \right\} di.$$

Applying equation (13), we obtain

$$\min \mathbb{E}_i \left[p_i - (1-\alpha)p - \alpha(\pi - y^*) + \tilde{\theta} - u_i\right]^2$$

$$= \left(\beta_1 + \frac{1}{\alpha}\right)^2 [(1-\alpha)\gamma + \alpha]^2\, e^{-x} q\sigma_\epsilon^2 +^2 \phi^2 e^{-y}$$

$$= \left(\beta_1 + \frac{1}{\alpha}\right)^2 [(1-\alpha)\gamma + \alpha]^2\, e^{-x} q\sigma_\epsilon^2 +^2 \phi^2 e^{-2k} e^{x}.$$

Thus we have

$$x_i^* = \max\left\{ \ln\left(\frac{(\beta_1 + \frac{1}{\alpha})\,[(1-\alpha)\gamma + \alpha]\, q^{0.5}\sigma_\epsilon e^{k}}{w\phi} \right), 0 \right\}. \tag{15}$$

When $x = 0$, it is obvious that $\gamma = 0$. When $x \neq 0$, applying equations (13) and (15) and we have

$$\gamma = \begin{cases} 1 - \frac{\phi}{\alpha(\beta_1 + \frac{1}{\alpha})q^{0.5}\sigma_\epsilon e^{k}}, & \text{if } \phi < \alpha(\beta_1 + \frac{1}{\alpha})q^{0.5}\sigma_\epsilon e^{k}, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

Welfare loss calculation: First, consider price dispersion. Combining equations (13), (14) and (16), we have

$$\int (p_i - p)^2 di = \int \left\{ \left(\beta_1 + \frac{1}{\alpha}\right) [(1-\alpha)\gamma + \alpha] (1 - e^{-x})(\theta_i - \tilde{\theta}) + E(u_i) \right\} di.$$

Using the result that $\int E(u_i)^2 di = (1 - e^{-z})\phi^2$, we obtain

$$\int (p_i - p)^2 di = \left(\beta_1 + \frac{1}{\alpha}\right)^2 \gamma^2 q \sigma_\epsilon^2 \frac{e^{-x}}{1 - e^{-x}} + \phi^2 (1 - e^{-2k+x})$$

$$= \left(\beta_1 + \frac{1}{\alpha}\right)^2 \gamma^2 q \sigma_\epsilon^2 \frac{e^{-x}}{1 - e^{-x}} + \phi^2 (1 - e^{-2k+x}) = \left(\beta_1 + \frac{1}{\alpha}\right)^2 \gamma^2 q \sigma_\epsilon^2 \frac{e^{-x}}{1 - e^{-x}} + \phi^2 (1 - e^{-2k+x}).$$

Next, we make use of $e^x = 1 + \frac{\gamma}{\alpha(1-\gamma)}$ and $1 - \gamma = \frac{\phi}{(\beta_1 + \frac{1}{\alpha}) q^{0.5} \sigma_\epsilon e^k \alpha}$ to get

$$\int (p_i - p)^2 di = \frac{\phi^2 \gamma^2}{e^{2k} \alpha^2 (1-\gamma)^2} \frac{e^{-x}}{1 - e^{-x}} - \phi^2 e^{-2k} e^x + \phi^2$$

$$= \phi^2 e^{-2k} (e^x - 1)^2 \frac{e^{-x}}{1 - e^{-x}} - \phi^2 e^{-2k} e^x + \phi^2 = \phi^2 e^{-2k} (e^x - 1) - \phi^2 e^{-2k} e^x + \phi^2.$$

$$= \phi^2 (1 - e^{-2k}).$$

Now consider the output gap:

$$(\pi - p - y^*)^2$$

$$= \left[p_t + \beta_1 \tilde{\theta} + \beta_2 \theta - \beta_0 - \beta_2 \theta + y^* - \left(\beta_1 + \frac{1}{\alpha}\right) [\gamma \tilde{\theta} + (1-\gamma) q \theta] - y^*) \right]^2$$

$$= \left\{ \beta_1 \tilde{\theta} - \left(\beta_1 + \frac{1}{\alpha}\right) [\gamma \tilde{\theta} + (1-\gamma) q \theta] \right\}^2$$

$$= \left\{ \left(\beta_1 + \frac{1}{\alpha}\right) (1-\gamma) q (\tilde{\theta} - \theta) + \left[\beta_1 - \left(\beta_1 + \frac{1}{\alpha}\right) \gamma - \left(\beta_1 + \frac{1}{\alpha}\right) (1-\gamma) q \right] \tilde{\theta} \right\}^2$$

$$= \left(\beta_1 + \frac{1}{\alpha}\right)^2 (1-\gamma)^2 q^2 \sigma_\epsilon^2 + \left[\beta_1 - \left(\beta_1 + \frac{1}{\alpha}\right) [\gamma + (1-\gamma) q] \right]^2 \sigma_\theta^2.$$

When there is learning on aggregate shock: We obtain the following expression of the output gap after applying $1 - \gamma = \frac{\phi}{\alpha(\beta_1 + \frac{1}{\alpha}) q^{0.5} \sigma_\epsilon e^k}$ and $\gamma = \frac{\alpha(\beta_1 + \frac{1}{\alpha}) q^{0.5} \sigma_\epsilon e^k - \phi}{\alpha(\beta_1 + \frac{1}{\alpha}) q^{0.5} \sigma_\epsilon e^k}$, so

$$\left(\beta_1 + \frac{1}{\alpha}\right) [\gamma + (1-\gamma) q] = \beta_1 + \frac{1}{\alpha} - \frac{(1-q)\phi}{\alpha q^{0.5} \sigma_\epsilon e^k}$$

and also

$$(\pi - p - y^*)^2 = \frac{\phi^2 q}{\alpha^2 e^{2k}} + \left[-\frac{1}{\alpha} + \frac{(1-q)\phi}{\alpha q^{0.5} \sigma_\epsilon e^k} \right]^2 \sigma_\theta^2 = \frac{\phi^2 q}{\alpha^2 e^{2k}} + \frac{1}{\alpha^2} \left[-1 + \frac{1-q}{q^{0.5} \sigma_\epsilon} \frac{\phi}{e^k} \right]^2 \sigma_\theta^2.$$

We have just shown that output gap is independent of monetary policy. Total welfare loss is calculated as

$$(\pi - p - y^*)^2 + \lambda \int (p_i - p)^2 di = \frac{\phi^2 q}{\alpha^2 e^{2k}} + \frac{1}{\alpha^2} \left[-1 + \frac{1-q}{q^{0.5} \sigma_\epsilon} \frac{\phi}{e^k} \right]^2 \sigma_\theta^2 + \lambda \phi^2 (1 - e^{-2k}).$$

When there is no learning on aggregate shock: Under this case we have the two components of the loss function as

$$\int (p_i - p)^2 di = \phi^2 \left(1 - e^{-2k}\right),$$

$$(\pi - p - y^*)^2 = \left[\beta_1 \tilde{\theta} - \left(\beta_1 + \frac{1}{\alpha}\right) q\theta\right]^2 = \left[\left(\beta_1 + \frac{1}{\alpha}\right) q(\theta - \tilde{\theta}) + \left[\beta_1 - \left(\beta_1 + \frac{1}{\alpha}\right) q\right] \tilde{\theta}\right]^2$$

$$= \left(\beta_1 + \frac{1}{\alpha}\right)^2 q^2 \sigma_\epsilon^2 + \left[\beta_1 - \left(\beta_1 + \frac{1}{\alpha}\right) q\right]^2 \sigma_\theta^2 = \left(\beta_1 + \frac{1}{\alpha}\right)^2 q^2 \sigma_\epsilon^2 + \left[\left(\beta_1 + \frac{1}{\alpha}\right)(1 - q) - \frac{1}{\alpha}\right]^2 \sigma_\theta^2.$$

$$(\pi - p - y^*)^2 + \lambda \int (p_i - p)^2 di = \left(\beta_1 + \frac{1}{\alpha}\right)^2 q^2 \sigma_\epsilon^2 + \left[\left(\beta_1 + \frac{1}{\alpha}\right)(1 - q) - \frac{1}{\alpha}\right]^2 \sigma_\theta^2 + \lambda \phi^2 (1 - e^{-2k}).$$

Differentiating the above with respect to β_1 and setting it to zero, we have

$$\beta_1 = 0.$$

At this level of β_1, welfare loss is equal to

$$\frac{q\sigma_\theta^2}{\alpha^2} + \lambda \phi^2 (1 - e^{-2k}).$$

Combining the two cases, we the following result for total welfare loss:

$$(\pi - p - y^*)^2 + \lambda \int (p_i - p)^2 di$$

$$= \begin{cases} \frac{q\sigma_\theta^2}{\alpha^2} + \lambda \phi^2 (1 - e^{-2k}), & \text{if } \phi > (1-q)^{0.5} \sigma_\theta e^k; \\ \frac{\phi^2 q}{\alpha^2 e^{2k}} + \frac{1}{\alpha^2} \left[-1 + \frac{(1-q)^{0.5}}{\sigma_\theta} \frac{\phi}{e^k}\right]^2 \sigma_\theta^2 + \lambda (1 - e^{-2k}), & \text{otherwise,} \end{cases}$$

Optimal monetary policy follows immediately

$$\beta_1 = \begin{cases} 0, & \text{if } \phi > (1-q)^{0.5} \sigma_\theta e^k; \\ \frac{\phi}{\alpha (1-q)^{0.5} \sigma_\theta e^k} - \frac{1}{\alpha}, & \text{otherwise.} \end{cases}$$

It is straightforward to show that welfare loss increases with the size of q: given that $\frac{q - q^2 + 1}{(2-q)^2}$ increases with $q \in (0, 1)$ and $\frac{\phi}{\sigma_\theta e^k} < 1$ in the set of $\phi > \frac{q^{0.5} \sigma_\epsilon e^k}{2-q}$.

8.2 Elastic attention

Firm i's problem follows

$$\min [p_i - (1 - \alpha) p - \alpha(\pi - \tilde{\theta}) - u_i]^2 + \mu k.$$

Substituting the expression for p_i, p, π, and the posterior variance of shocks, we have

$$\min \left(\beta_1 + \frac{1}{\alpha}\right)^2 [(1 - \alpha)\gamma + \alpha]^2 q \sigma_\epsilon^2 e^{-x} + \omega^2 \phi^2 e^{-z} + \mu 0.5(x + z),$$

where $I(\{\theta, \tilde{\theta}\}, \theta_i) + I(u_i, e_i) = k$, $x + z = 2k$. First order conditions imply

$$x=\ln\left(\frac{\left(\beta_1+\frac{1}{\alpha}\right)^2[(1-\alpha)\gamma+\alpha]^2q\sigma_\epsilon^2}{0.5\mu}\right), \qquad z=\ln\left(\frac{\phi^2}{0.5\mu}\right).$$

Welfare loss calculation: First, consider price dispersion

$$\int(p_i-p)^2\,di=\left(\beta_1+\frac{1}{\alpha}\right)^2\gamma^2q\sigma_\epsilon^2\frac{e^{-x}}{1-e^{-x}}+\phi^2(1-e^{-z}).$$

Applying $\gamma=\frac{\alpha e^x-\alpha}{\alpha e^x+1-\alpha}=\frac{\alpha(1-e^{-x})}{e^{-x}+\alpha(1-e^{-x})}$ and $1-\gamma=\frac{1}{\alpha e^x+1-\alpha}$, $\gamma+(1-\gamma)q=\frac{\alpha e^x-\alpha+q}{\alpha e^x+1-\alpha}$, we obtain

$$\int(p_i-p)^2\,di=\left(\beta_1+\frac{1}{\alpha}\right)^2q\sigma_\epsilon^2\frac{e^{-x}}{1-e^{-x}}\left(\frac{\alpha e^x-\alpha}{\alpha e^x+1-\alpha}\right)^2+\phi^2(1-e^{-z})$$

$$=\left(\beta_1+\frac{1}{\alpha}\right)^2\gamma^2q\sigma_\epsilon^2\frac{e^{-x}}{1-e^{-x}}+\phi^2(1-e^{-z})=0.5\mu(1-e^{-x})+\phi^2(1-e^{-z}).$$

Next, consider the output gap,

$$(\pi-p-y^*)^2=\left\{\beta_1\tilde{\theta}-\left(\beta_1+\frac{1}{\alpha}\right)[\gamma\tilde{\theta}+(1-\gamma)q\theta]\right\}^2$$

$$=\left\{\left(\beta_1+\frac{1}{\alpha}\right)(1-\gamma)q(\theta-\tilde{\theta})+\left[\beta_1-\left(\beta_1+\frac{1}{\alpha}\right)\gamma-\left(\beta_1+\frac{1}{\alpha}\right)(1-\gamma)q\right]\tilde{\theta}\right\}^2$$

$$=\left(\beta_1+\frac{1}{\alpha}\right)^2(1-\gamma)^2q^2\sigma_\epsilon^2+\left[\beta_1-\left(\beta_1+\frac{1}{\alpha}\right)[\gamma+(1-\gamma)q]\right]^2\sigma_\theta^2.$$

Given that $0.5\mu e^x=\frac{(\beta_1+1/\alpha)^2\gamma^2q\sigma_\epsilon^2}{(1-e^{-x})^2}$, $\left(\beta_1+\frac{1}{\alpha}\right)^2=\frac{0.5\mu e^x(1-e^{-x})^2}{q\sigma_\epsilon^2}\frac{(\alpha e^x+1-\alpha)^2}{(\alpha e^x-\alpha)^2}=\frac{0.5\mu}{q\sigma_\epsilon^2}\frac{(\alpha e^x+1-\alpha)^2}{e^x\alpha^2}$, we have

$$\left(\beta_1+\frac{1}{\alpha}\right)^2(1-\gamma)^2q^2\sigma_\epsilon^2=0.5\mu e^x\frac{(1-\gamma)^2}{\gamma^2}\frac{q}{(1-e^{-x})^2}=\frac{0.5\mu q}{\alpha^2e^x},$$

$$\left(\beta_1+\frac{1}{\alpha}\right)[\gamma+(1-\gamma)q]=(\beta_1+1/\alpha)\frac{\alpha e^x-\alpha+q}{\alpha e^x+1-\alpha}=\frac{\sqrt{0.5\mu}}{\sqrt{q}\sigma_\epsilon\alpha e^{0.5x}}[\alpha e^x-\alpha+q],$$

$$\beta_1-\left(\beta_1+\frac{1}{\alpha}\right)[\gamma+(1-\gamma)q]=\frac{\sqrt{0.5\mu}[\alpha e^x+1-\alpha]}{\sqrt{q}\sigma_\epsilon\alpha e^{0.5x}}-\frac{1}{\alpha}-\frac{\sqrt{0.5\mu}[\alpha e^x-\alpha+q]}{\sqrt{q}\sigma_\epsilon\alpha e^{0.5x}}=\frac{\sqrt{0.5\mu}(1-q)}{\sqrt{q}\sigma_\epsilon\alpha e^{0.5x}}-\frac{1}{\alpha}.$$

As a result, output gap follows

$$(\pi-p-\tilde{\theta})^2=\frac{0.5\mu q}{\alpha^2e^x}+\left[\frac{\sqrt{0.5\mu}(1-q)}{\sqrt{q}\sigma_\epsilon\alpha e^{0.5x}}-\frac{1}{\alpha}\right]^2\sigma_\theta^2=\frac{0.5\mu q}{\alpha^2}T^2+\left[\frac{\sqrt{0.5\mu}(1-q)}{\sqrt{q}\sigma_\epsilon\alpha}T-\frac{1}{\alpha}\right]^2\sigma_\theta^2,$$

where we define $T\equiv e^{-0.5x}$. Therefore, total welfare loss is written as

$$(\pi-p-\tilde{\theta})^2+\lambda\int(p_i-p)^2\,di=\frac{0.5\mu q}{\alpha^2}T^2+\left[\frac{\sqrt{0.5\mu}(1-q)}{\sqrt{q}\sigma_\epsilon\alpha}T-\frac{1}{\alpha}\right]^2\sigma_\theta^2+\lambda0.5\mu(1-T^2)+\lambda\phi^2(1-e^{-z})$$

$$= \frac{0.5\mu q}{\alpha^2}T^2 + \left[\frac{\sqrt{0.5\mu(1-q)}}{\sigma_\theta \alpha}T - \frac{1}{\alpha}\right]^2 \sigma_\theta^2 + \lambda 0.5\mu\left(1-T^2\right) + \lambda\phi^2\left(1-e^{-z}\right)$$
$$= \frac{0.5\mu}{\alpha^2}T^2 - \frac{2\sigma_\theta\sqrt{0.5\mu(1-q)}}{\alpha^2}T + \frac{\sigma_\theta^2}{\alpha^2} + \lambda 0.5\mu\left(1-T^2\right) + \lambda\phi^2\left(1-e^{-z}\right).$$

We can see that it is a decreasing function of q. Once again, the public signal reduces welfare. Optimal monetary policy generates an equilibrium T that satisfies the following first order condition:

$$\left[\frac{\mu q}{\alpha^2} - \lambda\mu + \frac{1}{\alpha^2}\mu(1-q)\right]T = 2\frac{\sigma_\theta^2}{\alpha^2}\frac{\sqrt{0.5\mu(1-q)}}{\sigma_\theta}.$$

The above can be rewritten as

$$T = \begin{cases} 1, & \text{if } \lambda\alpha^2 > 1 - \frac{2\sqrt{0.5\mu(1-q)}}{\mu}\sigma_\theta, \\ \frac{2\sqrt{0.5\mu(1-q)}\sigma_\theta}{\mu(1-\lambda\alpha^2)}, & \text{if } \lambda\alpha^2 \leq 1 - \frac{2\sqrt{0.5\mu(1-q)}}{\mu}\sigma_\theta \end{cases} \tag{17}$$

However, when the learning condition is satisfied, we always have the corner solution. Thus, the optimal monetary policy follows

$$\beta_1 = \begin{cases} \frac{\sqrt{0.5\mu}}{\alpha(1-q)^{0.5}\sigma_\theta} - \frac{1}{\alpha}, & \text{if } \sqrt{0.5\mu} \leq (1-q)^{0.5}\sigma_\theta, \\ 0, & \text{if } \sqrt{0.5\mu} > (1-q)^{0.5}\sigma_\theta \end{cases} \tag{18}$$