Models of pricing and inflation

The Calvo model and the New Keynesian Phillips curve: a review

Let's refresh ourselves on how the New Keynesian Phillips curve is derived from the Calvo assumption.[1] First, we'll assume that the "final good" is produced as a constant-elasticity-of-substitution (CES) aggregate of monopolistically competitive firms' production:

$$Y_t \equiv \left(\int_0^1 Y_t(i)^{1-\frac{1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{1}$$

Suppose, for simplicity, that the production technology for each firm is linear in labor, with a coefficient that we'll normalize to one: $Y_t(i) = N_t(i)$. It follows that the cost of producing $Y_t(i)$ is $W_t Y_t(i)$, where W_t is the nominal wage.

Firm's problem. The Calvo assumption is that a firm is only able to change its price with iid probability $1-\theta$ each period, and fulfills all demand at its current price.

A firm that is able to change at date t chooses its price P_t^* to maximize the current market value of profits generated while the price is in effect, i.e. to maximize

$$\max_{P_t^*} \sum_{k=0}^{\infty} \theta^k \mathbb{E}_t [Q_{t,t+k}(P_t^* - W_{t+k}) Y_{t+k}(P_t^*)] \tag{2}$$

where $Y_{t+k}(P_t^*) = \left(\frac{P_t^*}{P_{t+k}}\right)^{-\varepsilon} Y_{t+k}$ is the demand implied by (1), so that $(P_t^* - W_{t+k})Y_{t+k}(P_t^*)$ is profit in period t, and $Q_{t,t+k} \equiv \beta^k (C_{t+k}/C_t)^{-\sigma}(P_t/P_{t+k})$ is the stochastic discount factor by which nominal payoffs at date $t+k$ are valued at date t, which is just the ratio of a consumer's marginal utilities from nominal spending on consumption at the two dates.

The firm takes everything as given in (2) except P_t^*. Note that the derivative of $Y_{t+k}(P_t^*)$ with respect to P_t^* is

$$Y'_{t+k}(P_t^*) = -\varepsilon (P_t^*)^{-1} \left(\frac{P_t^*}{P_{t+k}} \right)^{-\varepsilon} Y_{t+k} = -\varepsilon \frac{Y_{t+k}(P_t^*)}{P_t^*}$$

Therefore, the derivative of period-t profits $(P_t^* - W_{t+k})Y_{t+k}(P_t^*)$ with respect to P_t^* is

$$Y_{t+k}(P_t^*) - \varepsilon \frac{Y_{t+k}(P_t^*)}{P_t^*}(P_t^* - W_{t+k})$$

which can be rewritten as

$$\frac{Y_{t+k}(P_t^*)}{P_t^*}(P_t^* - \varepsilon P_t^* + \varepsilon W_{t+k}) = -(\varepsilon - 1)\frac{Y_{t+k}(P_t^*)}{P_t^*}(P_t^* - \frac{\varepsilon}{\varepsilon - 1} W_{t+k})$$

and the first-order condition for (2), multiplying both sides by $-P_t^*/(\varepsilon-1)$ to eliminate the constant above, is

$$\sum_{k=0}^{\infty} \theta^k \mathbb{E}_t \left[Q_{t,t+k} Y_{t+k}(P_t^*) \left(P_t^* - \frac{\varepsilon}{\varepsilon - 1} W_{t+k} \right) \right] = 0 \tag{3}$$

[1]This narration is loosely taken from Galí (2015).

If $\theta = 0$, so that we always get to reset and only the $k = 0$ term in (3) is nonzero, then this gives a simple flexible-price markup for prices over wages: $P_t^* = \frac{\varepsilon}{\varepsilon-1} W_t$, where $\frac{\varepsilon}{\varepsilon-1}$ is the usual markup over marginal cost for a monopolistic competitor.

We must similarly have $P^* = \frac{\varepsilon}{\varepsilon-1} W$ in a zero-inflation steady state, since that's the only way that (3) can equal zero with constant P^* and W. Note also that in that steady state, $Q_{t,t+k} = \beta^k (C_{t+k}/C_t)^{-\sigma} (P_t/P_{t+k}) = \beta^k$, because C and P are constant.

Now, linearizing (3) around the zero-inflation steady state, the product rule implies something very nice: since the inner term in parentheses, $\left(P_t^* - \frac{\varepsilon}{\varepsilon-1} W_{t+k}\right)$, is zero in steady state, we can ignore any change in $Q_{t,t+k} Y_{t+k}(P_t^*)$, since it's multiplying something equal to zero.[2] Therefore we get just

$$\sum_{k=0}^{\infty} \theta^k \beta^k \mathbb{E}_t [dP_t^* - \frac{\varepsilon}{\varepsilon - 1} dW_{t+k}] = 0$$

which can be rearranged as

$$\begin{aligned}
\left(\sum_{k=0}^{\infty} (\beta\theta)^k \right) dP_t^* &= \frac{\varepsilon}{\varepsilon - 1} \sum_{k=0}^{\infty} (\beta\theta)^k \mathbb{E}_t [dW_{t+k}] \\
dP_t^* &= (1 - \beta\theta) \frac{\varepsilon}{\varepsilon - 1} \sum_{k=0}^{\infty} (\beta\theta)^k \mathbb{E}_t [dW_{t+k}] \\
\frac{dP_t^*}{P^*} &= (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k \mathbb{E}_t \left[\frac{dW_{t+k}}{W} \right] \qquad (4)
\end{aligned}$$

where in the last step we divide by steady-state $P^* = \frac{\varepsilon}{\varepsilon-1} W$. Letting hats denote log deviations from steady state, so that $\hat{P}_t^* \equiv \frac{dP_t^*}{P^*}$ and $\hat{W}_t \equiv \frac{dW_t}{W}$, then (4) reduces to a simple equation for the relationship between prices and wages, in log deviations from steady state. Firms set prices to match the mean of expected wages k periods in the future, discounted by $(\beta\theta)^k$, where β is the discount rate and θ is the "persistence" of the price, since the Calvo fairy visits with probability $1 - \theta$:

$$\hat{P}_t^* = (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k \mathbb{E}_t [\hat{W}_{t+k}] \qquad (5)$$

Aggregate prices. The derivation summarized in (5) only gives the dynamics of the reset price P_t^* for a firm that chooses prices at date t.

What happens to the aggregate price, i.e. the price P_t of purchasing a single unit of the final good? One can derive from (1) that

$$P_t \equiv \left(\int_0^1 P_t(i)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}} \qquad (6)$$

Now, suppose that a random fraction $1 - \theta$ of firms, which we will label as the subset $i \in [0, 1 - \theta]$ for convenience, is chosen by the Calvo fairy to reset and resets at price P_t^*, while the remaining $(1 - \theta, 1]$ of

[2] This is just one case of the general point that if we're totally differentiating $d(x \cdot y) = dx \cdot y + x \cdot dy$, then the first term is zero if $y = 0$.

the continuum does not and keeps their old prices $P_{t-1}(i)$. Then we have

$$
\begin{aligned}
P_t &= \left(\int_0^{1-\theta} (P_t^*)^{1-\varepsilon} di + \int_{1-\theta}^1 P_{t-1}(i)^{1-\varepsilon} di\right)^{\frac{1}{1-\varepsilon}} \\
&= \left((1-\theta)(P_t^*)^{1-\varepsilon} + \theta \int_0^1 P_{t-1}(i)^{1-\varepsilon} di\right)^{\frac{1}{1-\varepsilon}} \\
&= \left((1-\theta)(P_t^*)^{1-\varepsilon} + \theta(P_{t-1})^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}} \qquad (7)
\end{aligned}
$$

where the crucial step is in the second line, where we wrote $\int_{1-\theta}^1 P_{t-1}(i)^{1-\varepsilon} di = \theta \int_0^1 P_{t-1}(i)^{1-\varepsilon} di$ because of the random Calvo adjustment assumption, so that the measure θ of firms who don't adjust is chosen randomly, and any integral of their prices is just θ times the integral of all prices.[3]

What happens if we log-linearize (7) around the zero-inflation steady state? Then we get simply

$$
\begin{aligned}
\hat{P}_t &= \frac{1}{1-\varepsilon}\left((1-\theta)(1-\varepsilon)\hat{P}_t^* + \theta(1-\varepsilon)\hat{P}_{t-1}\right) \\
&= (1-\theta)\hat{P}_t^* + \theta \hat{P}_{t-1} \qquad (8)
\end{aligned}
$$

which gives a very simple law of motion for aggregate prices. Note that we can rewrite this $AR(1)$ as an $MA(\infty)$ by substituting the same expression for $\hat{P}_{t-1}$, etc, until we get

$$
\hat{P}_t = (1-\theta)\sum_{k=0}^{\infty} \theta^k \hat{P}_{t-k}^* \qquad (9)
$$

which is a nice parallel with (5): whereas (5) states that that pricesetters set their "reset prices" equal to the geometric average of current and future expected nominal marginal costs (in this case, just wages) with weights $(\beta\theta)^k$, (9) states that the overall price index is a geometric average of current and past reset prices, with weights θ^k.

This is all pretty intuitive: someone setting prices today chooses her "reset price" to try and hit an average of future marginal costs while the price is still in effect, which k periods in the future has probability θ^k (adding discounting by β^k to reflect less concern about the future). Then, the overall price today is an average of reset prices from k periods ago based on the fraction of prices today that were set back then, which is proportional to θ^k.

How do we get the New Keynesian Phillips curve? So far, we have two very intuitive expressions, (5) and (9), which we'll repeat below:

$$
\hat{P}_t^* = (1-\beta\theta)\sum_{k=0}^{\infty} (\beta\theta)^k \mathbb{E}_t[\hat{W}_{t+k}] \qquad (5)
$$

$$
\hat{P}_t = (1-\theta)\sum_{k=0}^{\infty} \theta^k \hat{P}_{t-k}^* \qquad (9)
$$

How do we obtain the canonical New Keynesian Phillips curve from these, which is a relationship between inflation $\pi_t \equiv \hat{P}_t - \hat{P}_{t-1}$ and real marginal cost $\widehat{mc}_t \equiv \hat{W}_t - \hat{P}_t$?

[3] Our notation here is a bit dicey, as is always the case when putting random variables on the continuum $[0,1]$. People have complained about this, e.g. Judd (1985), but it's too much of a distraction to try to be more formal.

This requires a few minor miracles. First, let's put both (5) and (9) into recursive, $AR(1)$ form. For the first, we can write

$$\hat{P}_t^* = (1-\beta\theta)\sum_{k=0}^{\infty}(\beta\theta)^k \mathbb{E}_t[\hat{W}_{t+k}]$$
$$= (1-\beta\theta)\hat{W}_t + \beta\theta\mathbb{E}_t\left[(1-\beta\theta)\sum_{k=1}^{\infty}(\beta\theta)^k\mathbb{E}_{t+1}[\hat{W}_{t+k}]\right] = (1-\beta\theta)\hat{W}_t + \beta\theta\mathbb{E}_t[\hat{P}_{t+1}^*] \tag{10}$$

where we use the law of iterated expectations in the second line.

Then, (9) in recursive form is just (8), which we've already written. Let's take (9) and subtract $\hat{P}_{t-1}$ from both sides, obtaining

$$\pi_t \equiv \hat{P}_t - \hat{P}_{t-1} = (1-\theta)(\hat{P}_t^* - \hat{P}_{t-1}) \tag{11}$$

Next, let's try to evaluate (11) by subtracting $\hat{P}_{t-1}$ from the left of (10), and $\hat{P}_{t-1} = \hat{P}_t - (\hat{P}_t - \hat{P}_{t-1}) = \hat{P}_t - \pi_t$ from the right:

$$\hat{P}_t^* - \hat{P}_{t-1} = (1-\beta\theta)(\hat{W}_t - \hat{P}_t) + \beta\theta\mathbb{E}_t[\hat{P}_{t+1}^* - \hat{P}_t] + \pi_t \tag{12}$$

Multiplying both sides by $(1-\theta)$ and applying (11), this becomes

$$\pi_t = (1-\theta)(1-\beta\theta)(\hat{W}_t - \hat{P}_t) + \beta\theta\mathbb{E}_t[\pi_{t+1}] + (1-\theta)\pi_t \tag{13}$$

or, isolating π_t on the left by subtracting $(1-\theta)\pi_t$ and then dividing by θ, just

$$\pi_t = \frac{(1-\theta)(1-\beta\theta)}{\theta}\widehat{mc}_t + \beta\mathbb{E}_t[\pi_{t+1}] \tag{14}$$

where we define *real marginal cost* $\widehat{mc}_t \equiv \hat{W}_t - \hat{P}_t$. This is the canonical New Keynesian Phillips curve relating inflation to real marginal cost and expected future inflation.

Note that the discount factor in (14) is only β, without the factor θ giving the survival probability of the price. This is a big deal: if we calibrate β to be consistent with real interest rates, most likely it's fairly close to 1 on a quarterly basis, while θ will be well below 1 if a decent fraction of firms adjust their prices every quarter. Having β rather than $\beta\theta$ as the discount on future inflation in (14) makes it vastly more forward-looking than we would otherwise have.

This only happened at the very end of our derivation. Prior to that, at (12), we had π_t on the right to reflect the fact that, *conditional* on real marginal cost, nominal marginal cost advances at the rate of inflation. This means that the desired price today, among pricesetters, moves up with inflation (again, conditional on real marginal cost), which creates a "strategic complementarity" among pricesetters. The amplification from this complementarity, which is θ^{-1}, offsets the discounting by θ that otherwise appears in the pricesetter's problem.[4]

Converting this to a Phillips curve in output (or the "output gap"). We often think about "Phillips curves" as being specified in terms of the "output gap" (the deviation of output from the level that would

[4]One way to think about it: as a pricesetter, suppose I first want to increase prices by 1. But then I know that the fraction $1-\theta$ of other pricesetters will do the same, and this will increase nominal marginal cost, so I want to increase my prices by another $1-\theta$, and so on. At the end of the day, I want to increase my prices by $1+(1-\theta)+(1-\theta)^2+\ldots=\theta^{-1}$ times whatever I would have wanted if I ignored the actions of other pricesetters. This offsets the discounting by θ.

prevail in a flexible-price economy), or in terms of unemployment, which tends to be closely related.

Here, I haven't specified any shocks that woudl affect the level of output in a flexible-price economy, so I'll treat "output gap" as synonymous with simply "output". Then, I'll note that to first order,[5] the linear production technology we assumed implies $Y_t = C_t = N_t$. If the household has utility in each period of the form

$$\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{N_t^{1+\eta}}{1+\eta} \tag{15}$$

then its first-order condition for consumption vs. labor is

$$\frac{W_t}{P_t} C_t^{-\sigma} = N_t^{\eta} \tag{16}$$

because at a disutility cost of N_t^{η}, the household can earn real wages $\frac{W_t}{P_t}$ that purchase consumption with marginal utility $C_t^{-\sigma}$. Substituting $C_t = N_t = Y_t$, this becomes

$$\frac{W_t}{P_t} = Y_t^{\sigma+\eta} \tag{17}$$

or, in log-linearized terms, just

$$\widehat{mc}_t = (\sigma+\eta)\hat{Y}_t$$

Substituting into (14), we get

$$\pi_t = \frac{(1-\theta)(1-\beta\theta)}{\theta}(\sigma+\eta)\hat{Y}_t + \beta\mathbb{E}_t[\pi_{t+1}] \tag{18}$$

which expresses inflation in terms of deviation of output from the steady steady, which is equivalent for us (in the absence of technology shocks) from the output gap.

The key added factor here is $(\sigma+\eta)$, the sum of the inverse elasticity of substitution and inverse Frisch elasticity of labor supply. When elasticities are lower, households demand high real wages to work in a boom (and vice versa), which puts greater pressure on real marginal costs and therefore inflation for any boom in output.

Empirical and quantitative issues with the New Keynesian Phillips curve

There are three big issues that pop up about the New Keynesian Phillips curve, in form (14) or (18):

- It is extremely forward-looking, discounting the future by only β (which in the representative-agent model equals $1/(1+r)$).
- It is *entirely* forward-looking, with no inertia.
- It has a coefficient of $\frac{(1-\theta)(1-\beta\theta)}{\theta}(\sigma+\eta)$ on the output gap, which can be quite high compared to empirical estimates.

The fact that the New Keynesian Phillips curve is entirely forward-looking means that, in principle, an anticipated decline in inflation should actually coincide with a *boom* in output. (That's what happens if

[5] This is only to first order, since there is a second-order cost of price dispersion that I'll ignore when deriving the first-order laws of the economy.

$\mathbb{E}_t[\pi_{t+1}]$ on the right of (18) is smaller than π_t.) This is in stark contrast to experience in the United States and elsewhere, where disinflation, even when it is anticipated (i.e. no longer fully unexpected), tends to involve an economic downturn.[6]

The lack of inertia in the New Keynesian Phillips curve stands in contrast to the general sense that inflation in the real world, at least once it's established for more than a quarter or two, seems to have some persistence, which is what makes disinflation difficult. Fuhrer and Moore (1995) pointed this out as a major problem, and quantitative models like Christiano, Eichenbaum and Evans (2005) have needed to build in inertia by modifying the model (in Christiano et al. (2005)'s case, by assuming that prices and wages are automatically "indexed" to inflation).

Finally, the coefficient $\frac{(1-\theta)(1-\beta\theta)}{\theta}(\sigma+\eta)$ can be quite high given reasonable-seeming parameters. Nakamura and Steinsson (2008) find that the average rate of price change monthly, excluding sales, is about 9–12%. If we put in $1-\theta = .3$ quarterly and $\beta \approx 1$, we get a quarterly slope $\frac{(1-\theta)(1-\beta\theta)}{\theta} = 0.129$ relating real marginal cost in a quarter and quarterly inflation.

Next, at the very least, $\sigma \geq 1$ and $\eta \geq 1$, so if we plug in $\sigma = 1$ and $\eta = 1$, then we get an additional factor of $\sigma + \eta = 2$ on the slope with respect to the output gap, giving us $\frac{(1-\theta)(1-\beta\theta)}{\theta}(\sigma+\eta) = 0.258$. Further, by "Okun's law", the change in the output gap is usually around twice the change in unemployment, so this would imply an overall coefficient of about 0.5 on unemployment in a quarterly Phillips curve.

What is the actual coefficient identified in the data? The headline Hazell, Herreno, Nakamura and Steinsson (2022) estimate is 0.0062! This is almost *two orders of magnitude* off. Clearly, there's a big quantitative puzzle here.

(Of course, it might seem like a bit less of a puzzle in the last few years, as inflation has exploded and then partially cooled down again. But that's part of the mystery—why did we detect such a weak Phillips curve relationship before, and only now we're seeing something larger? There are many possibilities, some outside the scope of what we'll cover—e.g. a highly convex, nonlinear wage Phillips curve. But in part, we also need to think about changes to the New Keynesian Phillips curve.)

Finally, it's important to be careful about how we adjust these numbers as we change frequency. The coefficient of ~ 0.5 that we derive above is for a *quarterly* Phillips curve, as is fairly standard. It tells us how higher unemployment for a single quarter affects quarterly (not annualized!) inflation. To convert this to the effect on the annualized rate of inflation, we need to multiply by 4. Further, if we're interested in the effect of unemployment being higher for an entire year (rather than just a quarter) on the annualized rate of inflation today, we have to multiply by approximately 4 again to reflect the cumulative effect of 4 quarters of unemployment (assuming roughly no discounting, $\beta \approx 1$, within the year). This would result in an annual Phillips curve coefficient of roughly $4 \times 4 \times 0.5 = 8$.

This is obviously extremely high: it would imply that if monetary policy managed to increase anticipated unemployment by 1 percentage point for the year, the annual rate of inflation would immediately fall by 8 percentage points, from that alone! (Surely the Fed would be delighted if disinflation was so easy.)

[6]This issue, I believe, was originally pointed out by Ball (1994).

Generalizing to time-dependent pricing

The non-recursive form of the Calvo model had two key equations. First, there's what we will call the *policy equation*, which gives the reset price policy for the pricesetters who do adjust, in terms of current and expected future nominal marginal cost (which in this case is just nominal wages):

$$\hat{P}_t^* = \frac{\sum_{k=0}^{\infty}(\beta\theta)^k \mathbb{E}_t[\hat{W}_{t+k}]}{\sum_{k=0}^{\infty}(\beta\theta)^k} = (1-\beta\theta)\sum_{k=0}^{\infty}(\beta\theta)^k \mathbb{E}_t[\hat{W}_{t+k}] \tag{5}$$

Second, there's what we will call the *law of motion*, which gives the actual change in the overall price level in terms of current and past reset prices:

$$\hat{P}_t = \frac{\sum_{k=0}^{\infty}\theta^k \hat{P}_{t-k}^*}{\sum_{k=0}^{\infty}\theta^k} = (1-\theta)\sum_{k=0}^{\infty}\theta^k \hat{P}_{t-k}^* \tag{9}$$

We see that there are geometrically declining weights θ^k in both cases, corresponding to the chance that a price we set today will still be in effect k periods from now for (5), and the change that a price in effect today will have been set k periods in the past for (9). (For (5), there is additional discounting by β^k.)

Now let's consider a generalization, where the chance that a price set today will still be in effect k periods from now is given by an arbitrary *survival function* Φ_k defined for $k = 0, 1, \ldots$. The Calvo case is the special geometric case $\Phi_k = \theta^k$, but in principle we can have any weakly declining Φ_k, as long as $\Phi_0 = 1$ and $\sum_k \Phi_k < \infty$.

For instance, the Taylor (1980) staggered pricing model features $\Phi_k = 1$ for $k < N$ and $\Phi_k = 0$ for $k \geq N$, for some contract length N, meant to capture some kind of regular contract length. We could imagine mixtures of Taylor and Calvo pricing: for instance, maybe in a quarterly calibration Φ_k drops mildly from Φ_0 through Φ_3, then drops to almost zero at Φ_4, then drops rapidly. This would capture a situation where firms usually wait a year from their last price change to change prices, but sometimes there is some shock that causes them to change prices earlier, and occasionally they wait longer than a year (but are very likely to change if it has been longer than a year).[7]

Given a survival function Φ_k, if a price has survived through date $k-1$, then its chance of being reset in date k is

$$\lambda_k \equiv \frac{\Phi_{k-1} - \Phi_k}{\Phi_{k-1}} = 1 - \frac{\Phi_k}{\Phi_{k-1}} \tag{19}$$

where λ_k is called the *hazard rate* of adjustment. We assume that the chance of being reset is exactly λ_k for every price that has survived through date $k-1$, i.e. that it doesn't depend on how far off the price is from its optimal level. In other words, this is a general *time-dependent* pricing rule rather than a *state-dependent* pricing rule, which we'll cover soon.

How do the policy equation and law of motion change in the general time-dependent world? One can rederive them step-by-step and show that essentially the same derivation goes through, so that the policy equation (5) becomes

$$\hat{P}_t^* = \frac{\sum_{k=0}^{\infty}\beta^k \Phi_k \mathbb{E}_t[\hat{W}_{t+k}]}{\sum_{k=0}^{\infty}\beta^k \Phi_k} \tag{20}$$

i.e. that we simply replace θ^k with the more general Φ_k. This makes sense: if pricesetters weight future

[7] Importantly, though, in this time-dependent model, we assume that this "shock" has to be orthogonal to the economy and the pricesetter's optimal price, which is a bit tenuous as an assumption. It could be some kind of internal planning change, perhaps.

costs by the chances that their prices will still be in effect, then they'll use Φ_k.

The law of motion (9) similarly becomes

$$\hat{P}_t = \frac{\sum_{k=0}^{\infty} \Phi_k \hat{P}^*_{t-k}}{\sum_{k=0}^{\infty} \Phi_k} \tag{21}$$

where the weights on past prices become the chance that prices will have survived that long.[8]

Matrix and vector notation: the pass-through matrix and the generalized Phillips curve

To solve a general time-dependent model, it turns out to be useful to have the same "sequence-space Jacobian" notation that we used for heterogeneous agent models, where sequences are mapped to other sequences.

In particular, if $\mathbf{P}^*$, $\mathbf{W}$, and $\mathbf{P}$ are stacked $\{\hat{P}^*_t\}$, $\{\hat{W}_t\}$, and $\{\hat{P}_t\}$, respectively, then we can write the policy equation (20), dropping the expectations sign, as:

$$\mathbf{P}^* = \frac{1}{\sum_{k=0}^{\infty} \beta^k \Phi_k} \begin{pmatrix} \Phi_0 & \beta\Phi_1 & \beta^2\Phi_2 & \cdots \\ 0 & \Phi_0 & \beta\Phi_1 & \cdots \\ 0 & 0 & \Phi_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \mathbf{W} \tag{22}$$

Similarly, we can write the law of motion (21) as:

$$\mathbf{P} = \frac{1}{\sum_{k=0}^{\infty} \Phi_k} \begin{pmatrix} \Phi_0 & 0 & 0 & \cdots \\ \Phi_1 & \Phi_0 & 0 & \cdots \\ \Phi_2 & \Phi_1 & \Phi_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \mathbf{P}^* \tag{23}$$

Then, if we combine (22) and (23) into a single equation relating nominal marginal cost $\mathbf{W}$ to prices $\mathbf{P}$, we get

$$\mathbf{P} = \underbrace{\frac{1}{\left(\sum_{k=0}^{\infty} \Phi_k\right)\left(\sum_{k=0}^{\infty} \beta^k \Phi_k\right)} \begin{pmatrix} \Phi_0 & 0 & 0 & \cdots \\ \Phi_1 & \Phi_0 & 0 & \cdots \\ \Phi_2 & \Phi_1 & \Phi_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \Phi_0 & \beta\Phi_1 & \beta^2\Phi_2 & \cdots \\ 0 & \Phi_0 & \beta\Phi_1 & \cdots \\ 0 & 0 & \Phi_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}}_{\equiv \Psi} \mathbf{W} \tag{24}$$

where, as in Auclert, Rigato, Rognlie and Straub (2024), we say that the matrix mapping $\mathbf{W}$ to $\mathbf{P}$ is the *pass-through matrix* from costs to prices, which we denote by Ψ.[9] The pass-through matrix entry Ψ_{ts} says how much the aggregate price level at date t responds to nominal marginal cost at date s.

Note that in the benchmark of perfectly flexible prices, with $\Phi_0 = 1$ and $\Phi_k = 0$ for $k > 0$, Ψ is simply the identity matrix $\mathbf{I}$. That's because we are operating in a model where the only thing stopping one-to-one pass-through of costs to prices is the price rigidity. (This is not to say that there aren't interesting other

[8] In extreme cases like Taylor, this requires an additional uniformity assumption to make sure that everyone isn't synchronizing their price change at the same date. But one can dispense with this assumption as long as Φ_1 is strictly less than 1, even by just ϵ, which in a steady state creates the requisite uniform mixing of price resetting across dates.

[9] Sometimes we might write Ψ^{Φ} to clarify that this is the pass-through matrix induced by the survival function Φ.

frictions that limit this pass-through, only that we won't cover them right now.)

Although (24) might seem a little daunting, it isn't so bad when read from right to left: the upper-triangular right matrix says that to get from $\mathbf{W}$ to $\mathbf{P}^*$, we need to take a forward-looking average weighted by $\beta^k\Phi_k$, and then the lower-triangular left matrix says that to get from $\mathbf{P}^*$ to $\mathbf{P}$, we need to take a backward-looking average weighted by Φ_k.

Solving for the generalized Phillips curve. The pass-through matrix Ψ gives the relationship between nominal marginal cost, in this case the deviation $\mathbf{W}$ in nominal wages, and the nominal price level $\mathbf{P}$. How can we get from there to a relationship akin to the New Keynesian Phillips curve (14), relating real marginal cost $\widehat{mc}_t = \hat{W}_t - \hat{P}_t$ to inflation π_t?

Stacking $\widehat{mc}_t$ in the vector $\mathbf{mc}$, we can write

$$\mathbf{P} = \Psi(\mathbf{mc} + \mathbf{P}) \tag{25}$$

where we see that (25) is a fixed-point equation, with the price level $\mathbf{P}$ appearing on both sides.

One approach to solving (25) might be to iterate: first assume on the right that $\mathbf{P} = 0$, getting a guess $\Psi\mathbf{mc}$ for $\mathbf{P}$, then feed that back into the right to obtain $\Psi(\mathbf{mc} + \Psi\mathbf{mc}) = (\Psi + \Psi^2)\mathbf{mc}$, and so on. Ultimately, we'll get $(\Psi + \Psi^2 + \ldots)\mathbf{mc}$, reflecting infinitely many rounds of feedback from real marginal cost to prices. It is nontrivial to handle the mathematical details necessary to show that this converges and that it actually equals the equilibrium price level, but we do so in an appendix of Auclert et al. (2024). We can then write

$$\mathbf{P} = (\mathbf{I} - \Psi)^{-1}\Psi\mathbf{mc} = \left(\sum_{k=1}^{\infty} \Psi^k\right)\mathbf{mc} \tag{26}$$

The inflation rate is then the first difference $\pi_t = \hat{P}_t - \hat{P}_{t-1}$. Stacking π_t in $\boldsymbol{\pi}$ and letting $\mathbf{L}$ be the "lag operator", we can write $\boldsymbol{\pi} = (\mathbf{I} - \mathbf{L})\mathbf{P}$, can combine with (26) to obtain

$$\boldsymbol{\pi} = \underbrace{(\mathbf{I} - \mathbf{L})(\mathbf{I} - \Psi)^{-1}\Psi}_{\equiv \mathbf{K}}\mathbf{mc} \tag{27}$$

where we call the matrix $\mathbf{K}$ that maps the vector of real marginal costs $\mathbf{mc}$ to the vector of inflation $\boldsymbol{\pi}$ the *generalized Phillips curve*, again as in Auclert et al. (2024).

Note that in the Calvo case, the standard New Keynesian Phillips curve (14) gives a very special structure for $\mathbf{K}$, namely

$$\mathbf{K}^{calvo} \equiv \frac{(1-\theta)(1-\beta\theta)}{\theta}\begin{pmatrix} 1 & \beta & \beta^2 & \cdots \\ 0 & 1 & \beta & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{28}$$

i.e. that the generalized Phillips curve $\mathbf{K}^{calvo}$ is upper triangular with entries that decline geometrically as we move above and to the right of the main diagonal. But more generally, we can evaluate the formula $\mathbf{K} = (\mathbf{I} - \mathbf{L})(\mathbf{I} - \Psi)^{-1}\Psi$ in (27) for any pass-through matrix and see what we get.

An important intuition: intrinsic inflation persistence. One longstanding idea is that with the appropriate time-dependent rules, we should get *inflation persistence*, unlike in the purely forward-looking New

Keynesian Phillips curve. In particular, the idea is that this is true when time-dependent rules have *increasing hazard rates*: where you are unlikely to adjust a price that you just adjusted, and then the probability of adjusting rises as time goes on. (An extreme example of an increasing hazard rate is the Taylor case, where the hazard rate starts at zero and then goes to one.) Sheedy (2010) is probably the most relevant paper doing this in a modern context, with the same generalized time-dependent rules we have above.

The intuition is that when today's pricesetters are the ones who haven't set prices in a long time, then if there has been inflation in the recent past, they will need to increase prices in order to *catch up* with the inflation that has already occurred, even if there are no additional shocks to real marginal cost. Intuitively, this can lead to a cycle of inflation: today's pricesetters increase their prices to catch up with yesterday's inflation, and then tomorrow's pricesetters increase their prices further to catch up with today's inflation, and so on.

It turns out that this intuition is true in the model, but perhaps not quantitatively as strong as one might think. An increasing hazard rate does lead to persistence, and inversely a decreasing hazard rate leads to anti-persistence (a partial reversal of past inflation), but this often dies out within just a few periods after the shock—a sharp contrast to the very forward-looking anticipatory behavior we get with the New Keynesian Phillips curve.

Can alternative time-dependent models make the Phillips curve less forward-looking? One of the leading problems with the New Keynesian Phillips curve is its lack of inertia, and as we discuss above, these more general models can partly fix this (but not necessarily by enough).

Another issue was its extreme forward-lookingness: the fact that the future is discounted only with β, which might be very close to 1, and is much higher than the survival probability of prices.

It turns out that in the general time-dependent case, although there isn't the exact forward-looking structure with a discount rate of β each period as in (28), *asymptotically* the effect of a real marginal cost shock in the future still decays at a rate of β. We'll see this in our computations, and it is a consequence of Proposition 3 in Auclert et al. (2024).

State-dependent ("menu cost") models

One conceptual weakness of the models we've seen so far is that the probability of adjusting prices is exogenous, and doesn't depend on whether a price is low or high—it's just a constant rate $1-\theta$ in the Calvo case, and the time-dependent hazard rate λ_k in the general time-dependent case.

This seems unrealistic, at least at the extremes: surely if your price is far, far higher than you want it to be, or far, far, lower, and this is losing you money, then you are probably going to change your price, regardless of what some arbitrary time-dependent schedule says.

An alternative class of models, *state-dependent models*, addresses this concern. In these models, the decision of whether or not to adjust prices is conditional on the firm's "state", i.e. the price itself vs. the fundamentals of cost, etc. In the simplest case, *menu cost models*, which we'll look at, adjusting prices incurs a fixed "menu cost", and each period firms decide whether or not to adjust.

There is a long history and vast variety of menu cost models. The modern literature, with a seminal paper being Golosov and Lucas Jr (2007), generally assumes that firms face idiosyncratic shocks to either productivity or demand (or both), and are constantly deciding whether or not to pay the menu cost in response to these idiosyncratic shocks. We then see what happens when we introduce an aggregate shock that changes costs (in Golosov and Lucas Jr (2007), a "money" shock that exogenously changes total nominal demand for goods).

One can try to tame these models by introducing a variety of simplifications, including a second-order approximation to the firm's profit function, and an assumption that firms' productivity follows a random walk, so that all that is relevant is the *gap* between the current price and productivity.[10] With all these simplifying assumptions, many of them made in Alvarez, Le Bihan and Lippi (2016) and also in the appendix of Auclert et al. (2024), we obtain a simplified problem of the form

$$\min_{\{p_{it}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{2}(p_{it} - p_{it}^* - \log W_t)^2 + \xi 1_{\{p_{it} \neq p_{it-1}\}}\right] \tag{29}$$

where the firm faces a quadratic cost of having its actual log price p_{it} deviate from the sum of its idiosyncratically optimal log price p_{it}^* and log nominal wages $\log W_t$, which we normalize in steady state to $\log W = 0$. We assume that p_{it}^* follows a random walk $p_{it}^* = p_{it-1}^* + \epsilon_{it}$, with some iid ϵ_{it} symmetric around zero. The firm balances this cost of having a suboptimal price against the "menu cost" ξ of adjusting its price.

If we define the *price gap* $x_{it} \equiv p_{it} - p_{it}^*$ to be the difference between a firm's actual log price p_{it} and its idiosyncratically optimal log price p_{it}^*, then we can rewrite (29) in terms of the price gap as

$$\min_{\{x_{it}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{2}(x_{it} - \log W_t)^2 + \xi 1_{\{x_{it} \neq x_{it-1} - \epsilon_{it}\}}\right] \tag{30}$$

Once the problem is written in this form, we can write a Bellman equation using the single state variable x rather than the two state variables p and p^*:

$$V_t(x) = \frac{1}{2}(x - \log W_t)^2 + \beta \mathbb{E}\left[\min\left(V_{t+1}(x+\epsilon), \xi + \min_{x^*} V_{t+1}(x^*)\right)\right] \tag{31}$$

[10] Technically, for this to work properly, one also needs to introduce demand shocks that are simultaneous with and in the opposite direction as productivity shocks. In response to a simultaneous positive productivity shock and negative demand shock, firms' optimal price p^* will fall, but their total demand at that price will not change. This is a technical hack needed to prevent some firms from asymptotically gaining all the demand in the economy. See Midrigan (2011). We won't worry too much about the details here, and will instead just take the specification (29) of the problem as given.

where here $V_t(x)$ is the value function given a *post-adjustment* price gap of x (not including any costs paid to adjust), which happens to be a convenient way to formulate the probelm.

It turns out that the optimal policy for price gaps x_{it} that solves (30), or equivalently (31), will generally take the so-called *Ss* form, with

$$x_{it} = \begin{cases} x_{it-1} - \epsilon_{it} & x_{it-1} - \epsilon_{it} \in [\underline{x}_t, \overline{x}_t] \\ x_t^* & \text{otherwise} \end{cases} \tag{32}$$

where the price is adjusted whenever yesterday's price gap x_{it-1}, adjusted for today's shock ϵ_{it}, lies outside the *adjustment bands* $[\underline{x}_t, \overline{x}_t]$, and is not adjusted otherwise. The optimal reset price gap x_t^* is the same no matter what the incoming price is, since the menu cost is fixed and does not depend on how much we adjust.

To a first-order approximation, one can write the aggregate price level $\log P_t$ as the cross-sectional average across all firms (which we assume has measure one) of price gaps:

$$\log P_t = \int p_{it} di = \int x_{it} di \tag{33}$$

where the integrals are equal because we normalize the average of p_{it}^* to zero (which is possible since shocks ϵ_{it} have mean zero).

We now have a complete model mapping log nominal wages $\log W_t$, which appear in (30), to log nominal prices $\log P_t$, which appear in (33). How might this model differ from the Calvo and generalized time-dependent models we've seen so far?

To evaluate this, we'll look at first-order shocks around the steady state. Given the assumption that the distribution of shocks ϵ_{it} is symmetric around 0, and that steady-state $\log W_t$ is zero in (30), so that the quadratic objective in (30) is symmetric around zero, the steady-state *Ss* rule will have the simple form

$$\overline{x} = -\underline{x}$$
$$x^* = 0$$

i.e. the adjustment bands for the price gap will be symmetric around zero, and the optimal reset point for the price gap will be zero. This will lead to a steady-state initial density of price gaps $g(x)$, which obeys the equation

$$g(x) = \text{freq} \cdot f(x) + \int_{\underline{x}}^{\overline{x}} f(x - x')g(x')dx'$$

where $f(\cdot)$ is the distribution of shocks ϵ_{it}, and $\text{freq} = 1 - \int_{\underline{x}}^{\overline{x}} g(x)dx$ is the fraction of prices that are reset each period.

We'll then look at how shocks $\hat{W}_t$ to $\log W_t$ lead to changes $\hat{P}_t$ in the path of aggegate prices, obtaining a pass-through matrix and generalized Phillips curve just like for the time-dependent case.

Reducing this to a mixture of two time-dependent models

Until recently, it was thought that these models, even in this simplified form, are quite hard to solve, and that some mix of brute-force numerical calculation and additional simplifying assumptions was needed. Proposition 1 in Auclert et al. (2024), however, provides an analytical result that obtains the pass-through matrix for the problem above as a convex combination of the pass-through matrices for two time-dependent models.[11]

Expectation function. Recursively define

$$E^t(x) = \int_{-\bar{x}}^{\bar{x}} f(x'-x) E^{t-1}(x') dx' \tag{34}$$

to be the *expectation function* for the price gap, with the initial condition $E^0(x) = x$. This function answers the question: in the aggregate steady state, if the (post-adjustment) price gap is x today, what is the expected price gap t periods in the future, assuming that x? The recursive formula (34) effectively applies the law of iterated expectations: the expectation t periods from now, given that the price gap today is x, is the same as the expectation today of tomorrow's expectation $t-1$ periods in the future, taking the expectation today over all possible x' tomorrow. Note that by symmetry, it is easy to show that $E^t(0) = 0$ for all t, so that only one term appears on the right of (34), taking the expectation of E^{t-1} conditional on there *not* being an adjustment tomorrow.

Immediately from the definition, we have that $E^0(x) = x$. In general, $E^t(x)$ converges to zero as $t \to \infty$ for two reasons. First, there is always a probability that we will adjust the price gap to zero, and then the expectation from then on is zero. Second, and more subtly, conditional on *not adjusting*, price gaps will tend to move closer to zero. That's because if prices ever went outside the adjustment bands, they would have been reset to zero; hence, the very fact that they have survived and not been reset means that they more likely moved away from the adjustment bands and toward zero. This is the so-called "selection effect" for prices.

A technical detour (for those interested). A bit more subtly, $E^t(x)$ will converge to multiples of an *eigenfunction*, and then decay at the corresponding eigenvalue. The reason is that (34) is a linear operator on functions, linearly mapping the function E^{t-1} to E^t. Repeatedly applying a matrix generally gives us a multiple of the eigenvector corresponding to the largest eigenvalue (since ones with other eigenvalues decay more quickly). Analogously, repeatedly applying this operator gives us the eigenfunction corresponding to a leading eigenvalue—but in this case the largest eigenvalue on *odd* functions, because all the $E^t(x)$ are odd, which is not the largest eigenvalue overall.[12]

Meanwhile, if we defined $\Phi^{actual,t}(x)$ to be the probability of survival for at least t periods if the price gap is currently x, then we would have $\Phi^{actual,0}(x) = 1$ and then the same recursion as (34) would apply:

$$\Phi^{actual,t}(x) = \int_{-\bar{x}}^{\bar{x}} f(x'-x) \Phi^{actual,t-1}(x') dx' \tag{35}$$

Since this is the same recursion, it's the same linear operator as in (34). but now since the initial condition is even $\Phi^{actual,0}(x) = 1$, $\Phi^{actual,t}(x)$ is even as well. It follows that it decays to the eigenfunction corresponding to the largest *even* eigenvalue, which turns out to be larger.[13] This is a manifestation of the "selection effect", which we'll see strongly in our computations: the E^t functions decay far, far faster than the $\Phi^{actual,t}$

[11] The baseline model for which this is proven in Auclert et al. (2024) is slightly more general than the model above, allowing for an additional iid Calvo-like probability λ of "free" price adjustments, so that with a λ probability, firms adjust prices no matter what.

[12] Formally, we need a bit more to ensure that there are discrete eigenvalues and eigenfunctions, e.g. that $E^t(x)$ is also a "compact" operator—but that gets into even more of a technical detour.

[13] See Alvarez and Lippi (2022) and Auclert et al. (2024) for more on this.

functions.

Auclert et al. (2024) result: getting the pass-through matrix. Our result (a bit too complicated to prove formally here) is that pass-through matrix of the menu cost model is given by

$$\Psi = \alpha \Psi^{\Phi^e} + (1-\alpha)\Psi^{\Phi^i} \tag{36}$$

where Ψ^{Φ^e} and Ψ^{Φ^i} are the pass-through matrices generated by the "virtual" survival functions Φ^e and Φ^i, which represent the extensive and intensive margins of the model,[14] and are given by

$$\Phi^e_t \equiv \frac{E^t(\overline{x})}{\overline{x}} \tag{37}$$

$$\Phi^i_t \equiv E^{t\prime}(0) = \lim_{x \to 0} \frac{E^t(x)}{x} \tag{38}$$

and the weight α on the extension margin is given by:

$$\alpha \equiv 2g(\overline{x})\overline{x} \cdot \sum_{t=0}^{\infty} \frac{E^t(\overline{x})}{\overline{x}} \tag{39}$$

The virtual survival functions Φ^e_t and Φ^i_t are fairly easy to interpret: they are the local persistences of the price gap at the intensive and extensive margin.

When we're choosing what price to set at the intensive margin in response to shocks, for instance, it turns out that what matters for us is the *effective persistence* of that price: if we decide to choose a price gap x^*_t that is a bit higher than zero today (because, say, costs have risen above the steady state today), how long will x^*_t stay above zero? At the margin, this is given by Φ^i_t above.

Similarly, when we decide whether to adjust to zero today around the adjustment threshold $\overline{x}$ (or, symmetrically, $-\overline{x}$), what matters is the effective persistence of that decision: if we decide to lower the price gap from $\overline{x}$ to zero today, how will that lower the price gap in the future?

These effective persistences turn out to be the analogs, in this model, of the generalized survival functions in the time-dependent case—such close analogs, in fact, that we can literally treat the extensive and intensive margins of adjustment individually as being equivalent to generalized time-dependent problems.

It turns out that for both the intensive and extensive margins, the "effective persistences" will be lower than the actual survival rate of a given price (i.e. how long is it until a price actually adjusts). This is because of the *selection effect* mentioned above: conditional on a price not adjusting, it is likely that the price gap is much closer to zero.

What this means is that the effective ("virtual") survival rate in the time-dependent models that comprise the menu cost model will decline much more quickly than in time-dependent models calibrated to the same rate of adjustment. This implies that a menu cost model will behave as if aggregate prices are much more flexible than we might expect from a calibrated Calvo model (where we already had a puzzle of too high a slope of the New Keynesian Phillips curve). This is a consistent finding from the quantitative menu cost literature, perhaps most prominently in Golosov and Lucas Jr (2007).

A slightly different way to put it is the following: since price-setters assume that they will adjust prices whenever their price gaps get very far out of line, their effective horizons are quite short. Whatever they

[14] The extensive margin is the decision of whether or not to adjust prices, and the intensive margin is the decision of what price to set conditional on adjustment.

choose today in terms of prices (both on the intensive and extensive margins) will not have a very long-lasting impact. Therefore, they should choose it with only the near future in mind—and they get closer to the case of perfect flexibility.

Ends up close to Calvo. As we'll see when we implement (36), the pass-through matrix Ψ and the resulting generalized Phillips curve $\mathbf{K}$ end up looking remarkably close to the New Keynesian Phillips curve produced by the Calvo case, just with a higher slope due to the selection effect. Hence, this doesn't give us much—it doesn't lessen forward-lookingness or add inertia, and it aggravates the existing puzzle of the Calvo model producing too high a slope for the Phillips curve.

Why is this model so close to Calvo? The main reason is that the virtual survival curves Φ_t^e and Φ_t^i behave similarly to the Calvo case, decreasing at a roughly constant hazard rate. Why is this? Technically, it's because there is rapid convergence to the leading odd eigenfunction discussed above, and then the hazard rate for both is the eigenvalue corresponding to this eigenfunction. But why is there this rapid convergence in practice? Loosely, it's because we the adjustment bands $[-\bar{x}, \bar{x}]$ to be relatively narrow in order to calibrate to a high enough rate of price changes (say, at least once annually, or 0.25 quarterly). This narrowness means that conditional on staying within the bands, there is a lot of mixing between different price gaps, and the initial price gap quickly becomes irrelevant. (One way to say this is that the distribution of your price gap several periods ahead, conditional on not having reset, quickly becomes independent of your position today.) So Φ_t^e and Φ_t^i quickly decay at the same, constant rate.

Initially, the hazards are different: Φ_t^i declines more slowly, and Φ_t^e declines more quickly, because the chance of resetting (and the associated selection effect) is much lower when we start from 0 than when we start from $\bar{x}$. But these differences go away quickly enough that there is not much quantitative impact.

The "Calvo-plus" model: adding free resets. The basic menu cost model implies that there are only large price changes, which does not seem true in the data: sometimes there are small price changes too, although prices do seem more likely to be reset when price gaps are especially large.

One easy way to fix this, achieving a blend of the menu cost and Calvo models, is to suppose that the menu cost is actually a stochastic ξ_{it}, equaling the positive constant $\xi > 0$ with iid probability $1 - \lambda$, and equaling zero with probablity λ. Whenever the menu cost is zero (a "free reset"), you'll reset no matter what; this Calvo-like adjustment leads to some small price changes. This model was dubbed the "Calvo-plus" model by Nakamura and Steinsson (2010).

As Auclert et al. (2024) show, essentially all the same results above go through with this modification. In the code, we'll implement this, finding that the resulting generalized Phillips curve is still very similar to the usual Calvo New Keynesian Phillips curve, with a slope that is still higher than the pure Calvo curve calibrated to the same rate of adjustment, but lower than the pure menu cost model.

Interestingly, the shape of the Phillips curve differs a *bit* more from Calvo than in the menu cost model—although it's still very similar. This unintuitive result (why would making the model more Calvo-like make the shape of the curve a bit less Calvo-like?) comes from the fact that there are wider adjustment bands, so that Φ_t^e and Φ_t^i do not converge quite as quickly to the same constant hazard rate.

Three simple modifications to the Calvo model

So far, we've seen that a menu cost model, and to a lesser degree time-dependent models in general, seem to behave fairly similarly to Calvo, with all its weaknesses. Now we'll consider the consequences of three simple modifications.

Strategic complementarity

In our models so far, a firm ideally wants to change its price in each period by the same log amount that nominal marginal cost changes. In some models, however, there is an additional motivation for firms to try and stick closer to the average price level in the economy—so that, if the firm's price was flexible, it would target $\chi MC_t + (1-\chi)P_t$ rather than just MC_t, for some $\chi \in [0,1]$. This is called *strategic complementarity*: firms want to be close to other firms' pricing decisions.

Recall that we derived (25) by writing $\mathbf{P} = \Psi\mathbf{MC}$,[15] and then replacing $\mathbf{MC} = \mathbf{mc} + \mathbf{P}$. Now, we can instead write $\mathbf{P} = \Psi(\chi\mathbf{MC} + (1-\chi)\mathbf{P})$, which becomes

$$\mathbf{P} = \Psi(\chi\mathbf{mc} + \mathbf{P}) \tag{40}$$

This is the same equation as (25), just with the coefficient χ multiplying $\mathbf{mc}$. Unsurprisingly, if we solve it to obtain inflation, we get

$$\boldsymbol{\pi} = \chi(\mathbf{I}-\mathbf{L})(\mathbf{I}-\Psi)^{-1}\Psi\mathbf{mc} \tag{41}$$

which is the same as (27), just with an additional χ factor multiplying the old $\mathbf{K} = (\mathbf{I}-\mathbf{L})(\mathbf{I}-\Psi)^{-1}\Psi$. So we obtain a simple result: strategy complementarity placing a weight of $1-\chi$ on the price level leads the generalized Phillips curve to be shrunk by a factor of χ.

With enough strategic complementarity, in principle we can shrink the slope of the Phillips curve as much as we want. This is often the strategy used in practice by quantitative papers to reconcile the Calvo model (with a reasonable adjustment frequency) with macro estimates of the Phillips curve slope.

What is the justification for strategic complementarity? There are three main ways of getting it, which we get by modifying *firm-specific marginal costs*, *markups*, or the *structure of production.*

- Suppose that firms have an upward-sloping marginal cost curve: it's more costly for them to produce additional units when they're producing a lot. This creates a motive for them to set their price closer to the aggregate price level: if they have an unusually high price, then they'll have low demand and therefore low costs (causing them to want to lower their price), and if they have an unusually low price, then they'll have high demand and therefore high costs (causing them to want to raise their price).
- Alternatively, suppose that the elasticity of demand is not constant, and instead is lower when you have a relatively low price and are selling more. Then if you have a low price, you'll want to raise it a bit, because the optimal markup is now higher with the lower elasticity of demand (and vice versa). An extreme case is a so-called "kinked" demand curve where the elasticity is lower when you have a low price (perhaps raising prices above the norm will send your customers away, but lowering prices

[15] We actually wrote $\mathbf{P} = \Psi\mathbf{W}$, since nominal marginal cost simply equalled the wage; but it's more general to put nominal marginal cost there.

won't get too many customers). See Klenow and Willis (2016). This is perhaps the most common way to obtain strategic complementarity.

- Yet another approach is to suppose that we have "roundabout" production, where some other firms' goods are needed in the production of our own good. If we define MC_t to exclude this part of cost, and suppose (for simplicity) that producing one unit of the intermediate good requires a mass $1-\chi$ of the final good, then we get exactly $\chi MC_t + (1-\chi)P_t$.[16] This is perhaps the simplest way to get strategic complementarity: other firms' prices directly affect our own costs, creating a motivation to keep our prices close to theirs.

Mixture models with heterogeneous pricing

So far we've considered models where firms are *homogeneous*. In the Calvo case, they all have the same Calvo rate of price change.

An alternative is to suppose that firms are heterogeneous. For instance, we could suppose that some firms have one Calvo parameter and other firms have another Calvo parameter (and that otherwise they are completely identical).

This is easy to handle: in the case where otherwise identical firms have different pricing frictions leading to different pass-through matrices, we simply combine the pass-through matrices to obtain the aggregate, and then proceed as before. For instance, if there are relatively "sticky" and "flexible" firms, we can write:

$$\Psi = a^{sticky}\Psi^{sticky} + (1-a^{sticky})\Psi^{flexible}$$

This leads to two main consequences. First, this sort of mixture between firms tends to lead to *anti-persistence*: following the inflation from a positive real marginal cost shock, inflation becomes negative, as the flexible firms reverse their earlier increase and get closer to the sticky firms' prices, which anchor the economy. This is the opposite of what we're looking for, although something resembling this does happen in practice when there are energy or food price shocks, which tend to create short bursts of inflation that then recede.

Second, the overall slope of the resulting generalized Phillips curve, especially if one excludes the extreme boom-and-bust inflation dynamic around the time of the shock, tends to be closer to what one would get from assuming that only the sticky pricers existed in the economy. This is a general point that has been observed by many people, i.e. Carvalho (2006), and it is one way that we can match an aggegate Phillips curve with a lower slope—by realizing that observed price change frequencies actually reflect a mix of high-frequency and low-frequency firms, and that aggregate behavior will more closely resemble what we'd expect from the low-frequency firms.

Why do the sticky firms dominate? Because the flexible firms can adjust, they tend to adjust their prices to be close to the sticky firms at any moment, so that the sticky firms' dynamics are the most important.

This approach is generalized to an explicit model with many sectors and input-output relationships in Rubbo (2022) and Afrouzi and Bhattarai (2023). Rubbo (2022) shows that with the right weighting on sectoral inflation—proportional to sectoral sales, and inversely proportional to flexibility—we obtain a price index that still obeys the standard New Keynesian Phillips curve.

[16] Basu (1995) is usually cited for this model.

Sticky expectations (or other behavioral frictions)

We can also modify the pass-through matrix using the tools from last lecture to reflect *sticky expectations*: the fact that following a date-0 shock, some firms have not updated their expectations on either real marginal costs or prices.

To do this, assuming that firms form expectations about real marginal costs and prices directly, we simply modify the pass-through matrix as we did other Jacobians in the previous lecture.

What we get is not surprising: a low initial response to future real marginal cost changes, as many firms have not updated their expectations yet (which feeds into prices, which feeds into the policies of firms that have updated their expectations). In response to a reasonable AR(1) shock, this can lead to a hump.

We still get anti-persistence, however: once a real marginal cost shock has passed, if there are some firms that have still not updated their expectations, then those firms won't have changed their prices at all. The firms that have updated their expectations realize this, and move their prices closer to the non-updating firms, creating deflation that partly offsets the earlier inflation.

References

Afrouzi, H. and Bhattarai, S. (2023). Inflation and GDP Dynamics in Production Networks: A Sufficient Statistics Approach .

Alvarez, F., Le Bihan, H. and Lippi, F. (2016). The real effects of monetary shocks in sticky price models: a sufficient statistic approach. *American Economic Review* 106(10):2817–2851.

Alvarez, F. and Lippi, F. (2022). The analytic theory of a monetary shock. *Econometrica* 90(4):1655–1680.

Auclert, A., Rigato, R., Rognlie, M. and Straub, L. (2024). New Pricing Models, Same Old Phillips Curves? *The Quarterly Journal of Economics* 139(1):121–186.

Ball, L. (1994). Credible disinflation with staggered price-setting. *The American Economic Review* 84(1):282–289.

Basu, S. (1995). Intermediate goods and business cycles: Implications for productivity and welfare. *The American Economic Review* 85(3):512.

Carvalho, C. (2006). Heterogeneity in price stickiness and the real effects of monetary shocks. *Frontiers in Macroeconomics* 2(1).

Christiano, L.J., Eichenbaum, M. and Evans, C.L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy* 113(1):1–45.

Fuhrer, J. and Moore, G. (1995). Inflation persistence. *The Quarterly Journal of Economics* 110(1):127–159.

Galí, J. (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and its Applications*. Princeton University Press.

Golosov, M. and Lucas Jr, R.E. (2007). Menu costs and phillips curves. *Journal of Political Economy* 115(2):171–199.

Hazell, J., Herreno, J., Nakamura, E. and Steinsson, J. (2022). The slope of the phillips curve: evidence from us states. *The Quarterly Journal of Economics* 137(3):1299–1344.

Judd, K.L. (1985). The Law of Large Numbers with a Continuum of iid Random Variables. *Journal of Economic theory* 35(1):19–25.

Klenow, P.J. and Willis, J.L. (2016). Real rigidities and nominal price changes. *Economica* 83(331):443–472.

Midrigan, V. (2011). Menu costs, multiproduct firms, and aggregate fluctuations. *Econometrica* 79(4):1139–1180.

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

Nakamura, E. and Steinsson, J. (2010). Monetary non-neutrality in a multisector menu cost model. *The Quarterly journal of economics* 125(3):961–1013.

Rubbo, E. (2022). Networks, Phillips Curves, and Monetary Policy .

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

Taylor, J.B. (1980). Aggregate dynamics and staggered contracts. *Journal of political economy* 88(1):1–23.