Inflation persistence when price stickiness differs between industries*

Kevin D. Sheedy†
London School of Economics

First draft: 22nd February 2005
This draft: 15th May 2007

Abstract

There is much evidence that price-adjustment frequencies vary widely across industries. This paper shows that inflation persistence is lower with heterogeneity in price stickiness than without it, taking as given the degree of persistence in variables affecting inflation. Differences in the frequency of price adjustment mean that the pool of firms which responds to any macroeconomic shock is unrepresentative, containing a disproportionately large number of firms from industries with more flexible prices. Consequently, this group of firms is more likely to reverse any initial price change after a shock has dissipated, making inflation persistence much harder to explain.

JEL Classification: E3

Keywords: Inflation persistence; heterogeneity; price stickiness; New Keynesian Phillips curve

*I am grateful to Petra Geraats for helpful discussions and I would like to thank Kosuke Aoki for his comments. The paper has also benefited from the comments of workshop participants at Cambridge University. I thank the Economic and Social Research Council for financial support.

†Department of Economics, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK. Tel: +44 207 107 5022, Fax: +44 207 955 6592, Email: k.d.sheedy@lse.ac.uk, Web: personal.lse.ac.uk/sheedy/.

1 Introduction

It is widely accepted that there are large differences in the frequency of price adjustment between industries.[1] This paper studies the implications of this phenomenon for the behaviour of inflation, and shows that the assumption of a common degree of price stickiness frequently used in macroeconomic models is not innocuous. Heterogeneity in price stickiness has the effect of reducing inflation persistence relative to what would occur with homogeneity, holding constant the degree of persistence in variables affecting inflation.

This reduction in inflation persistence occurs because when there are differences in the frequency of price adjustment between industries, the group of firms that responds to any macroeconomic shock with a price change is not representative of all firms in the economy. Instead, it contains a disproportionately large number of firms drawn from industries with more flexible prices. But these firms are then more likely to reverse any price changes they have made once the shocks that gave rise to those price changes have dissipated.

In the case where all industries have equally sticky prices, it is just as likely that prices which were left fixed after a shock move at a later time in the direction of those prices that were initially adjusted, than it is the latter subsequently moving in the direction of the former. With heterogeneity on the other hand, there is a greater likelihood that prices which were initially changed gravitating back towards those that remained fixed than vice versa. This increased tendency for prices to change direction once any shocks have passed reduces inflation persistence, thus making it much harder for theoretical models to explain observed inflation persistence once heterogeneity is accounted for. The extent of the inflation persistence puzzle is therefore underestimated in theoretical work that makes the simplifying assumption of equally sticky prices in all industries.

Recent discussions of inflation persistence have drawn a helpful distinction between intrinsic and extrinsic sources of persistence.[2] Intrinsic inflation inertia is the persistence in inflation that is generated directly by whatever frictions or imperfections underlie the short-run Phillips curve, and does not depend on there being any persistence in those variables which are the determinants of inflation. Intrinsic inertia can arise from various sources, such as backwards-looking rules of thumb for price setting (Galí and Gertler, 1999), indexation of prices to past inflation (Christiano, Eichenbaum and Evans, 2005), relative contracting models for wages (Fuhrer and Moore, 1995), or firms preferring to change older rather than newer prices (Sheedy, 2007*a*). On the other hand, extrinsic inflation persistence is whatever persistence is already present in the determinants of inflation (for example, in variables such as unemployment, the output gap, unit labour costs, or the growth rate of the money supply), and which is not itself directly explained by the ideas on which the Phillips curve is founded. This extrinsic persistence is inherited by inflation and it feeds into overall inflation persistence along with any intrinsic inertia.

This paper takes an otherwise standard New Keynesian model of price setting, the Calvo (1983) model, and adds heterogeneity in the frequency of price adjustment across a potentially large number of industries. It is well known that the New Keynesian Phillips curve resulting from Calvo pricing with homogeneity implies no intrinsic inflation inertia, and this aspect of the model has received much criticism.[3] This paper shows analytically that adding heterogeneity (an arbitrary non-degenerate distribution of price-adjustment frequencies) to the model actually makes the problem worse because it always generates the opposite of inflation inertia, that is, a tendency for above-average inflation to be followed by below-average inflation, a feature that can be thought of as *negative* inflation inertia, Thus holding the level of extrinsic persistence

[1] This is attested in the survey evidence of Blinder, Canetti, Lebow and Rudd (1998), and in studies such as Bils and Klenow (2004) and Dhyne, Álvarez, Le Bihan, Veronese, Dias, Hoffmann, Jonker, Lünnemann, Rumler and Vilmunen (2005) using very large databases of prices of individual goods for the U.S. and the Euro area respectively.

[2] A similar taxonomy is employed by Fuhrer (2005), Altissimo, Ehrmann and Smets (2006) and Sheedy (2007*a*) among others.

[3] See the detailed derivation and discussion of the New Keynesian Phillips curve in Woodford (2003).

constant, heterogeneity diminishes overall inflation persistence.

The intuition for this result can best be understood by considering a transitory (serially uncorrelated) shock to one of the determinants of inflation, for example, an increase in unit labour costs. The one-period rise in costs induces some fraction of firms to increase their prices, but others keep theirs fixed. With homogeneous Calvo pricing and the New Keynesian Phillips curve, there is a one-off jump in inflation in response to the shock, which means that the price level rises and then immediately reaches a plateau. There is no inflation persistence.

The equilibrium rate of inflation that occurs once the shock has dissipated can be understood in terms of two countervailing effects.[4] First there is the "catch-up" effect of firms that did not change their money prices initially, but subsequently want price increases to bring them back into line with the now-higher general price level. The second is the "roll-back" effect of firms that did initially raise their prices, but now find they are too high relative to the general price level, and consequently want price cuts to bring themselves into line with others.[5] When the catch-up effect is larger than the roll-back effect, there continues to be above-average inflation; and when the roll-back effect is dominant, inflation now falls below average. With homogeneous Calvo pricing, the two effects always exactly cancel out for transitory shocks.

The addition of heterogeneity into the story upsets the precarious balance between the catch-up and roll-back effects. Now the group of firms that want to catch up is disproportionately drawn from industries with stickier prices; and the group that wants to roll back features a preponderance of firms from industries with more flexible prices. This clearly strengthens the roll-back effect at the expense of the catch-up effect. Because the roll-back effect now dominates, inflation actually falls below average after the shock. A spell of higher-than-average inflation is thus followed by a spell of lower-than-average inflation. Since positive inflation persistence is nothing other than higher-than-average inflation following higher-than-average, and lower-than-average following lower-than-average, the swing from above to below average means that inflation persistence is actually negative with heterogeneous price stickiness and transitory shocks.

Earlier work on incorporating heterogeneous price stickiness into New Keynesian models has focused on the implications for optimal monetary policy (Aoki, 2001; Benigno, 2004). More recent studies by Cavalho (2005) and Dixon and Kara (2005) have addressed the effects of such heterogeneity on short-run macroeconomic dynamics. In addition, de Walque, Smets and Wouters (2006) have argued that heterogeneity in price stickiness combined with changes in industry-specific technology could be an additional source of cost-push shocks at the aggregate level.

The current paper differs from these studies in a number of respects. First, the paper presents analytical results on short-run dynamics, rather than relying on simulations of calibrated models.[6] Second, unlike the work by Cavalho and by Dixon and Kara, the focus here is on inflation persistence, instead of persistence in real variables such as output and unemployment. Moreover, this paper looks specifically at intrinsic inflation inertia as well as overall persistence. The advantage of this is that the effect of heterogeneity on intrinsic inflation inertia may be a structural feature of the economy if there are inherent reasons for the different price-adjustment frequencies prevailing across industries. On the other hand, the amount of overall inflation persistence is sensitive to assumptions made about aggregate demand, the conduct of monetary policy, the persistence of the shocks hitting the economy, among many other things. In addition, when addressing certain issues such as the cost of disinflation, it is crucial to focus only on intrinsic inertia.

The analysis in the current paper has some connection to the work on heterogeneity by Álvarez, Burriel and Hernando (2005). They show that estimates of the hazard function for price changes (the probability

[4]See (Sheedy, 2007*a*) for another application of this analysis.

[5]Note that whenever a fraction of firms changes price, the average percentage change in their prices alone must necessarily exceed the overall inflation rate.

[6]Calibrations are used in this paper to assess the quantitative significance of heterogeneity, but the direction of its effect is established analytically.

of a price change as a function of the age of the current price) using microeconomic data are biased towards detecting a negative slope when there is heterogeneity in the stickiness of the prices that make up the sample. In a macroeconomic context, Sheedy (2007*a*) shows that upwards-sloping hazard functions generate positive intrinsic inflation inertia and downwards-sloping ones negative intrinsic inertia. While there is no formal equivalence between a model with heterogeneity and one with a downwards-sloping hazard function, there is a close connection between the two which helps to explain the intuition for the results presented here.

It is important to contrast the analysis presented in this paper with that of Altissimo, Mojon and Zaffaroni (2007). They argue that heterogeneity can increase inflation persistence because the overall inflation rate is an average of many industry-specific inflation rates, each with a different degree of persistence.[7] However, such a claim depends on the shocks to industry-level inflation rates being independent across sectors. This clearly does not apply when any macroeconomic shocks are present, such as monetary policy shocks. Furthermore, even if this explanation does contribute to understanding the observed persistence of economy-wide inflation, it could not automatically be used to draw conclusions about the dynamic effects of macroeconomic shocks, which is one of the principal motivations for the study of inflation persistence. The results presented below in this paper are directly applicable to analysing the effects of macroeconomic shocks.

The plan of the paper is as follows. Section 2 sets out the model and studies firms' profit-maximizing price-setting decisions when prices are not changed continually. Section 3 then aggregates firms' behaviour across industries with different degrees of price stickiness to obtain a Phillips curve, and derives analytical results on intrinsic inflation inertia and inflation dynamics. Section 4 then presents a calibration of the model to assess the quantitative significance of the results, and also discusses how the analysis is connected with other work on heterogeneity and hazard functions for price changes. Finally, section 5 draws some conclusions.

2 The model

2.1 Assumptions

The economy contains a continuum of firms producing differentiated goods. Firms producing goods with similar costs of production, similar degrees of substitutability to customers, and similar frequencies of price adjustment are grouped together into industries. There are $n \geq 2$ industries and each firm belongs to one and only one industry. Industry i has size $0 < \omega_i < 1$, as measured by the proportion of the economy's firms that are based within it. The industry sizes ω_i must therefore sum to one. Firms in the economy are distributed along the unit interval, which is partitioned into separate industries as follows,

$$\bigcup_{i=1}^{n} \Omega_i = [0,1) \quad , \quad \Omega_i \equiv \left[\sum_{j=1}^{i-1} \omega_j, \sum_{j=1}^{i} \omega_j\right) \tag{2.1.1}$$

with Ω_i denoting the set of firms in industry i.

Firms' customers (households, government, other firms) allocate their spending between different products to minimize the cost of buying a given amount of a basket of goods. Baskets of goods at the industry and economy level are defined using Dixit-Stiglitz aggregators,

$$Y_{it} \equiv \left(\frac{1}{\omega_i}\int_{\Omega_i} Y_t(\imath)^{\frac{\varepsilon_{\mathsf{f}}-1}{\varepsilon_{\mathsf{f}}}} d\imath\right)^{\frac{\varepsilon_{\mathsf{f}}}{\varepsilon_{\mathsf{f}}-1}} \quad , \quad Y_t \equiv \left(\sum_{i=1}^{n} \omega_i Y_{it}^{\frac{\varepsilon_{\mathsf{s}}-1}{\varepsilon_{\mathsf{s}}}}\right)^{\frac{\varepsilon_{\mathsf{s}}}{\varepsilon_{\mathsf{s}}-1}} \tag{2.1.2}$$

[7] This conclusion is based on Granger's (1980) aggregation theorem.

where $Y_t(\imath)$ is the output of firm $\imath \in [0, 1)$ at time t, Y_{it} is industry i output, and Y_t is aggregate output. The parameters $\varepsilon_{\mathsf{f}} > 1$ and $\varepsilon_{\mathsf{s}} \geq 0$ are respectively the elasticities of substitution between the products of firms within one industry, and between the products of different sectors of the economy.[8] Customers' expenditure minimization for the basket of goods at the industry level implies that firms face the following demand functions,

$$Y_t(\imath) = \left(\frac{P_t(\imath)}{P_{it}}\right)^{-\varepsilon_{\mathsf{f}}} Y_{it} \quad , \quad P_{it} \equiv \left(\frac{1}{\omega_i}\int_{\Omega_i} P_t(\imath)^{1-\varepsilon_{\mathsf{f}}} d\imath\right)^{\frac{1}{1-\varepsilon_{\mathsf{f}}}} \tag{2.1.3}$$

where $P_t(\imath)$ is the money price charged by firm $\imath$ and P_{it} is the industry i price index. Similarly, expenditure minimization for the economy-wide basket of goods implies the following industry-level demand functions:

$$Y_{it} = \left(\frac{P_{it}}{P_t}\right)^{-\varepsilon_{\mathsf{s}}} Y_t \quad , \quad P_t \equiv \left(\sum_{i=1}^{n} \omega_i P_{it}^{1-\varepsilon_{\mathsf{s}}}\right)^{\frac{1}{1-\varepsilon_{\mathsf{s}}}} \tag{2.1.4}$$

In the above, P_{it} is the industry i price level from (2.1.3) and P_t denotes the economy-wide price level. By putting together the demand functions in (2.1.3) and (2.1.4), the following consolidated demand function for firm $\imath$ in industry i ($\imath \in \Omega_i$) is obtained:

$$Y_t(\imath) = \left(\frac{P_t(\imath)}{P_t}\right)^{-\varepsilon_{\mathsf{f}}} \left(\frac{P_{it}}{P_t}\right)^{\varepsilon_{\mathsf{f}}-\varepsilon_{\mathsf{s}}} Y_t \tag{2.1.5}$$

Firm $\imath$ in industry i can produce output $Y_t(\imath)$ at total real cost $\mathcal{C}\left(Y_t(\imath); Y_t^*, \mathcal{Z}_{it}\right)$,

$$\mathcal{C}\left(Y_t(\imath); Y_t^*, \mathcal{Z}_{it}\right) \equiv \frac{\mathcal{Z}_{it}}{1+\eta_{cy}} \frac{Y_t(\imath)^{1+\eta_{cy}}}{Y_t^{*\eta_{cy}}} \tag{2.1.6}$$

where Y_t^* is the economy's potential output, $\mathcal{Z}_{it}$ represents any other exogenous factors influencing costs in industry i, and $\eta_{cy} \geq 0$ is the elasticity of real marginal cost $\mathcal{C}_Y\left(Y_t(\imath); Y_t^*, \mathcal{Z}_{it}\right)$ with respect to an individual firm's output $Y_t(\imath)$. Potential output is defined as the level of output where real marginal cost is equal to one in the absence of any exogenous shocks to costs, that is, $\mathcal{C}_Y\left(Y_t^*; Y_t^*, 1\right) \equiv 1$. Potential output Y_t^* is taken to be exogenous in this paper.

Since each good is produced by only one firm and is an imperfect substitute for rival products, all firms have some market power and are price setters in the market for their own good. Prices are set in money terms, with $P_t(\imath)$ being the money price at time t of the good produced by firm $\imath$. Let $\varrho_t(\imath) \equiv P_t(\imath)/P_t$ and $\varrho_{it} \equiv P_{it}/P_t$ be the implied relative prices of the products of firm $\imath$ and industry i respectively. Total real profits at time t for firm $\imath$ in industry i are given by the profit function $\digamma\left(\varrho_t(\imath); \varrho_{it}, Y_t, Y_t^*, \mathcal{Z}_{it}\right)$, which is obtained by subtracting total real cost $\mathcal{C}\left(Y_t(\imath); Y_t^*, \mathcal{Z}_{it}\right)$ from the level of total real revenue implied by demand function (2.1.5):

$$\digamma\left(\varrho_t(\imath); \varrho_{it}, Y_t, Y_t^*, \mathcal{Z}_{it}\right) \equiv \varrho_t(\imath)^{1-\varepsilon_{\mathsf{f}}} \varrho_{it}^{\varepsilon_{\mathsf{f}}-\varepsilon_{\mathsf{s}}} Y_t - \mathcal{C}\left(\varrho_t(\imath)^{-\varepsilon_{\mathsf{f}}} \varrho_{it}^{\varepsilon_{\mathsf{f}}-\varepsilon_{\mathsf{s}}} Y_t; Y_t^*, \mathcal{Z}_{it}\right) \tag{2.1.7}$$

By substituting in the functional form for the cost function (2.1.6), and defining the output gap $\mathcal{Y}_t \equiv Y_t/Y_t^*$, the profit function (2.1.7) can be written as:

$$\digamma\left(\varrho_t(\imath); \varrho_{it}, Y_t, Y_t^*, \mathcal{Z}_{it}\right) = \left\{\varrho_t(\imath)^{1-\varepsilon_{\mathsf{f}}} \varrho_{it}^{\varepsilon_{\mathsf{f}}-\varepsilon_{\mathsf{s}}} - \frac{1}{1+\eta_{cy}} \varrho_t(\imath)^{-\varepsilon_{\mathsf{f}}(1+\eta_{cy})} \varrho_{it}^{(\varepsilon_{\mathsf{f}}-\varepsilon_{\mathsf{s}})(1+\eta_{cy})} \mathcal{Y}_t^{\eta_{cy}} \mathcal{Z}_{it}\right\} Y_t \tag{2.1.8}$$

Not all firms change their money prices in every time period. The frequency of price adjustment

[8] The most plausible case is where $\varepsilon_{\mathsf{f}} > \varepsilon_{\mathsf{s}}$ so products from the same industry are more substitutable than the products of different industries, though this assumption is not actually necessary for any of the results.

is modelled using the assumption of Calvo (1983) price-setting, but allowing for heterogeneity between industries. Every firm in industry i has a constant probability α_i of changing price in any given time period. Some industries have stickier prices than others so there is a dispersion of price-adjustment probabilities. The precise nature of this distribution of probabilities over industries does not need to be specified, but without loss of generality, it is convenient to number the industries in increasing order of price flexibility, so industry 1 has the stickiest prices and industry n the most flexible prices. In addition to this, it is assumed for simplicity that no industry has completely sticky or completely flexible prices, and no two industries have exactly the same probabilities of price adjustment. These assumptions are not very restrictive since all the results apply to economies arbitrarily close to any of these cases. The above statements are summarized by the following chain of inequalities for the industry-specific probabilities of price adjustment $\{\alpha_i\}_{i=1}^{n}$:

$$0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{n-1} < \alpha_n < 1 \tag{2.1.9}$$

Finally, when firms do change price, their prices are set to maximize the discounted value of the stream of profits they generate. Future profits are discounted using the nominal interest rate.[9]

2.2 Profit-maximizing price setting

Firms anticipate that the prices they choose are likely to remain sticky for at least some period of time. This means that they must take into account the effect on expected future profits when choosing a new price. At time t, consider a firm in industry i that is deciding what price to set. Its choice of price in money terms is denoted by R_{it} and is referred to as a reset price. The reset price is selected to maximize the discounted value of the stream of future profits generated by the price. In addition to the discounting of future profits by financial markets, it is necessary to contemplate the possibility that another new reset price will have been chosen before some of these future profits are realized. Using the Calvo pricing assumption, the probability that a firm in industry i which changes price at time t will still be using the same price in period $\tau \geq t$ is given by $(1-\alpha_i)^{\tau-t}$. The objective function for firms that incorporates both sources of discounting is

$$\max_{R_{it}} \sum_{\tau=t}^{\infty} (1-\alpha_i)^{\tau-t} \mathbb{E}_t \left[\left(\prod_{s=t+1}^{\tau} \frac{\mathit{\Pi}_s}{\mathcal{I}_s} \right) \digamma \left(\frac{R_{it}}{P_\tau}; \varrho_{i\tau}, Y_\tau, Y_\tau^*, \mathcal{Z}_{i\tau} \right) \right] \tag{2.2.1}$$

where $\mathbb{E}_t[\cdot]$ denotes the mathematical expectation conditional on all available information in period t, $\mathit{\Pi}_t \equiv P_t/P_{t-1}$ is the gross inflation rate between $t-1$ and t, $\mathcal{I}_t$ is the gross nominal interest rate also between periods $t-1$ and t, and $\digamma(\varrho_t(\imath); \varrho_{it}, Y_t, Y_t^*, \mathcal{Z}_{it})$ is the single-period real profit function defined in (2.1.7). The first-order condition characterizing the profit-maximizing reset price R_{it} in (2.2.1) is:

$$\sum_{\tau=t}^{\infty} (1-\alpha_i)^{\tau-t} \mathbb{E}_t \left[\left(\prod_{s=t+1}^{\tau} \frac{1}{\mathcal{I}_s} \right) \digamma_\varrho \left(\frac{R_{it}}{P_\tau}; \varrho_{i\tau}, Y_\tau, Y_\tau^*, \mathcal{Z}_{i\tau} \right) \right] = 0 \tag{2.2.2}$$

The derivative of the single-period profit function $\digamma(\varrho_t(\imath); \varrho_{it}, Y_t, Y_t^*, \mathcal{Z}_{it})$ in (2.1.8) with respect to a firm's own relative price $\varrho_t(\imath)$ is given by

$$\begin{aligned} \digamma_\varrho(\varrho_t(\imath); \varrho_{it}, Y_t, Y_t^*, \mathcal{Z}_{it}) = (1-\varepsilon_\mathsf{f}) \Bigg\{ & \varrho_t(\imath)^{-\varepsilon_\mathsf{f}} \varrho_{it}^{\varepsilon_\mathsf{f}-\varepsilon_\mathsf{s}} \\ & - \left(\frac{\varepsilon_\mathsf{f}}{\varepsilon_\mathsf{f}-1} \right) \varrho_t(\imath)^{\varepsilon_\mathsf{f}(1+\eta_{cy})-1} \varrho_{it}^{(\varepsilon_\mathsf{f}-\varepsilon_\mathsf{s})(1+\eta_{cy})} \mathcal{Y}_t^{\eta_{cy}} \mathcal{Z}_{it} Y_t \Bigg\} \end{aligned} \tag{2.2.3}$$

[9] The conclusions of this paper are not affected by making other assumptions about asset markets.

where $\varepsilon_{\mathsf{f}}/(\varepsilon_{\mathsf{f}}-1)$ is firms' desired (gross) markup of price on marginal cost if prices were fully flexible. Since $\varepsilon_{\mathsf{f}} > 1$ this markup is always well defined and greater than one. By defining $G_t \equiv Y_t/Y_{t-1}$ to be the gross growth rate of aggregate output Y_t, and $r_{it} \equiv R_{it}/P_t$ to be the reset price of industry i relative to all other prices in the economy, an expression for industry i's profit-maximizing reset price is obtained from equations (2.2.2) and (2.2.3):

$$r_{it} = \left(\frac{\frac{\varepsilon_{\mathsf{f}}}{\varepsilon_{\mathsf{f}}-1} \sum_{\tau=t}^{\infty}(1-\alpha_i)^{\tau-t}\mathbb{E}_t\left[\left(\prod_{s=t+1}^{\tau} \frac{G_s \mathit{\Pi}_s^{\varepsilon_{\mathsf{f}}+(1+\varepsilon_{\mathsf{f}}\eta_{cy})}}{\mathcal{I}_s}\right) \varrho_{i\tau}^{(\varepsilon_{\mathsf{f}}-\varepsilon_{\mathsf{s}})(1+\eta_{cy})} \mathcal{Y}_\tau^{\eta_{cy}} \mathcal{Z}_{i\tau}\right]}{\sum_{\tau=t}^{\infty}(1-\alpha_i)^{\tau-t}\mathbb{E}_t\left[\left(\prod_{s=t+1}^{\tau} \frac{G_s \mathit{\Pi}_s^{\varepsilon_{\mathsf{f}}}}{\mathcal{I}_s}\right) \varrho_{i\tau}^{\varepsilon_{\mathsf{f}}-\varepsilon_{\mathsf{s}}}\right]} \right)^{\frac{1}{1+\varepsilon_{\mathsf{f}}\eta_{cy}}} \tag{2.2.4}$$

Since the cost and demand conditions faced by two firms in the same industry are identical, equation (2.2.4) shows that all firms changing price at the same time in the same industry choose a common profit-maximizing reset price.

Because of the Calvo pricing assumption that the probability of price adjustment in industry i is always α_i, the proportion of firms in that industry which are using a price set j periods ago will eventually converge to $\alpha_i(1-\alpha_i)^j$. It is assumed that the economy has already reached this unique stationary distribution of the duration of price stickiness. Since all firms in the same industry that change price at the same time choose identical reset prices, the industry price index P_{it} from (2.1.3) can be written in terms of current and past reset prices from that industry:

$$P_{it} = \left(\sum_{j=0}^{\infty} \alpha_i(1-\alpha_i)^j R_{i,t-j}^{1-\varepsilon_{\mathsf{f}}} \right)^{\frac{1}{1-\varepsilon_{\mathsf{f}}}} \tag{2.2.5}$$

The equations for the economy-wide price level P_t from (2.1.4) and the industry price levels from (2.2.5) can be recast in terms of relative prices $\varrho_{it} \equiv P_{it}/P_t$, relative reset prices $r_{it} \equiv R_{it}/P_t$ and gross inflation rates $\mathit{\Pi}_t \equiv P_t/P_{t-1}$:

$$1 = \sum_{i=1}^{n} \omega_i \varrho_{it}^{1-\varepsilon_{\mathsf{s}}} \quad , \quad \varrho_{it}^{1-\varepsilon_{\mathsf{f}}} = \sum_{j=0}^{\infty} \alpha_i(1-\alpha_i)^j r_{i,t-j}^{1-\varepsilon_{\mathsf{f}}} \left(\prod_{k=0}^{j-1} \mathit{\Pi}_{t-k}^{\varepsilon_{\mathsf{f}}-1} \right) \tag{2.2.6}$$

For given stochastic processes for the output gap $\{\mathcal{Y}_t\}$, real output growth $\{G_t\}$, nominal interest rates $\{\mathcal{I}_t\}$, and exogenous cost-push shocks $\{\mathcal{Z}_{it}\}$, equations (2.2.4) and (2.2.6) determine relative prices ϱ_{it}, relative reset prices r_{it}, and the overall gross rate of inflation $\mathit{\Pi}_t \equiv P_t/P_{t-1}$, with the level of any money price being indeterminate unless a nominal anchor is specified. However, it is impossible to find an exact solution of these equations in most cases, so instead the equations are log-linearized around a steady state in order to obtain a first-order accurate approximation to the solution.[10] A steady state where inflation and real output growth are zero is chosen for simplicity.[11] Full details of the steady-state values of all variables are given in appendix A.1.

In what follows, a bar over a variable denotes its steady-state value, and a sans serif letter denotes the log deviation of the corresponding roman letter from its steady-state value. When a variable is indeterminate in the steady state (such as any money price) the sans serif letter denotes just the logarithm of the variable. Hence, $\bar{G}$ denotes the steady-state gross growth rate of aggregate output, and $\mathsf{G}_t \equiv \log G_t - \log \bar{G}$ is the log deviation of the growth rate from its steady-state value. On the other hand, $\mathsf{P}_t \equiv \log P_t$ is just the log of the general price level P_t. In addition to this convention, $\pi_t \equiv \log \mathit{\Pi}_t - \log \bar{\mathit{\Pi}}$ denotes the log deviation of the economy-wide inflation rate, $\rho_{it} \equiv \log \varrho_{it} - \log \bar{\varrho}$ the log deviation of industry i's relative price,

[10] This is standard practice in many New Keynesian models. See Woodford (2003) for further details.
[11] These assumptions can be relaxed without substantially altering the results.

$\mathsf{y}_t \equiv \log \mathcal{Y}_t - \log \bar{\mathcal{Y}}$ the log deviation of the output gap, and $\mathsf{z}_{it} \equiv \log \mathcal{Z}_{it} - \log \bar{\mathcal{Z}}_i$ the log deviation of the industry-i cost-push shock.

Appendix A.1 shows how log-linearized versions of the economy-wide and industry-level price indices in (2.1.4) and (2.2.5) can be derived from (2.2.6):

$$\mathsf{P}_t = \sum_{i=1}^{n} \omega_i \mathsf{P}_{it} \quad , \quad \mathsf{P}_{it} = \sum_{j=0}^{\infty} \alpha_i (1-\alpha_i)^j \mathsf{R}_{i,t-j} \tag{2.2.7}$$

The aggregate price level is a weighted average of industry-specific price levels, which are in turn weighted averages of current and past reset prices. When prices are sticky, the current price indices obviously depend on past pricing decisions, with the weights on the past decaying more rapidly in industries with more flexible prices (higher α_i).

It is also demonstrated in appendix A.1 that the log-linearized version of the profit-maximizing reset price equation (2.2.4) is

$$\mathsf{R}_{it} = (1-\beta(1-\alpha_i)) \sum_{j=0}^{\infty} \beta^j (1-\alpha)^j \mathbb{E}_t[\mathsf{P}_{i\mathcal{T}} - \eta_\rho \rho_{i,t+j} + \eta_y \mathsf{y}_{t+j} + \eta_z \mathsf{z}_{i,t+j}] \tag{2.2.8}$$

where η_ρ, η_y and η_z are positive constants defined in equation (A.1.6) of appendix A.1, and the parameter $0 < \beta < 1$ is the steady-state real interest rate expressed as a discount factor. The industry-specific profit-maximizing reset prices R_{it} depend on weighted averages of current and expected future price levels $\mathsf{P}_{i\mathcal{T}}$, relative prices $\rho_{i\mathcal{T}}$, output gaps $\mathsf{y}_\mathcal{T}$, and exogenous cost-push shocks $\mathsf{z}_{i\mathcal{T}}$. They are increasing in the industry-specific price levels, the output gap and cost-push shocks, and decreasing in the industry-specific relative prices. When prices are sticky, profit-maximization clearly requires firms to take account of both current and expected future conditions, with the weights attached to the future decaying more slowly in industries with stickier prices (lower α_i).

3 The Phillips curve and inflation persistence

3.1 Aggregation

The next step is to aggregate the profit-maximizing behaviour of firms derived in section 2.2 into a Phillips curve determining economy-wide inflation. To set up a system of equations that determines the overall inflation rate, it is convenient to use vector and matrix notation to represent as a block the pricing equations for all industries. In what follows, boldface letters are used to denote the $n \times 1$ vectors of the corresponding industry-specific variables. For example, $\mathbf{P}_t$ is the vector of industry-specific (log) price levels P_{it}, and $\mathbf{R}_t$ is the vector of (log) reset prices R_{it}. The vector of relative prices ρ_{it} is given by $\boldsymbol{\rho}_t$, and the vector of inflation rates π_{it} by $\boldsymbol{\pi}_t$.

If $\boldsymbol{\omega}$ is the vector of industry sizes ω_i, the aggregate price level equation in (2.2.7) can be written as $\mathsf{P}_t = \boldsymbol{\omega}'\mathbf{P}_t$ in vector notation. Relative prices can be expressed as $\boldsymbol{\rho}_t = \mathbf{P}_t - \boldsymbol{\iota}\mathsf{P}_t$, where $\boldsymbol{\iota}$ is a $n \times 1$ vector of 1s, or equivalently, $\boldsymbol{\rho}_t$ can be obtained by premultiplying the price-level vector $\mathbf{P}_t$ by a $n \times n$ matrix $\boldsymbol{\mathcal{R}}$. This matrix is defined by $\boldsymbol{\mathcal{R}} \equiv \mathbf{I} - \boldsymbol{\iota}\boldsymbol{\omega}'$, with $\mathbf{I}$ denoting the $n \times n$ identity matrix.

The following result shows how the set of equations for the profit-maximizing reset prices (2.2.8) and for the price indices (2.2.7) can be combined to obtain a relationship between the vector of industry-specific price levels $\mathbf{P}_t$, the output gap y_t, and cost-push shocks $\mathbf{z}_t$. This aggregate supply relationship can also be stated in terms of a series of industry Phillips curves, relating the vectors of industry-specific inflation rates $\boldsymbol{\pi}_t$ and relative prices $\boldsymbol{\rho}_t$ to the output gap and cost-push shocks.

Proposition 1 *By combining equations (2.2.7) and (2.2.8) there exists a $n \times n$ positive definite and diagonal matrix $\mathbf{K}$, and a $n \times n$ positive definite matrix $\boldsymbol{\mathcal{M}}$, such that the aggregate supply relationship between the vector of prices $\mathbf{P}_t$ and the output gap y_t is given by:*

$$\mathbf{P}_t = \boldsymbol{\mathcal{M}}^{-1}\left(\mathbf{P}_{t-1} + \beta \mathbb{E}_t \mathbf{P}_{t+1} + \mathbf{K}(\eta_y \boldsymbol{\iota} \mathsf{y}_t + \eta_z \mathbf{z}_t)\right) \tag{3.1.1}$$

An equivalent system of industry Phillips curves involving industry-specific inflation rates $\boldsymbol{\pi}_t$ and relative prices $\boldsymbol{\rho}_t$ is:

$$\boldsymbol{\pi}_t = \beta \mathbb{E}_t \boldsymbol{\pi}_{t+1} - \eta_\rho \mathbf{K} \boldsymbol{\rho}_t + \mathbf{K}(\eta_y \boldsymbol{\iota} \mathsf{y}_t + \eta_z \mathbf{z}_t) \tag{3.1.2}$$

The diagonal matrix $\mathbf{K} \equiv \operatorname{diag}\{\kappa_i\}_{i=1}^n$ contains the industry-specific component of the short-run Phillips curve slopes, denoted by κ_i for industry i. These satisfy the following chain of inequalities:

$$0 < \kappa_1 < \kappa_2 < \cdots < \kappa_{n-1} < \kappa_n < \infty \tag{3.1.3}$$

Industries with more flexible prices (larger α_i) have steeper short-run Phillips curves.

Proof See appendix A.6. ■

Equations (3.1.1) and (3.1.2) can be averaged over industries using the weights in the vector $\boldsymbol{\omega}$ to obtain the overall price level $\mathsf{P}_t = \boldsymbol{\omega}'\mathbf{P}_t$ and economy-wide inflation rate $\pi_t = \boldsymbol{\omega}'\boldsymbol{\pi}_t$.

According to equation (3.1.1), the current price level vector $\mathbf{P}_t$ depends positively on its past and expected future values, and positively on the output gap y_t and cost-push shocks $\mathbf{z}_t$. The lagged price vector $\mathbf{P}_{t-1}$ appears because some firms will continue to use past prices in the current time period. This directly affects the period t price index, as well as the decisions of firms changing price at time t. Current pricing decisions are also influenced by expectations of future prices $\mathbb{E}_t\mathbf{P}_{t+1}$ because firms anticipate that their own prices may remain sticky for some time and thus overlap with prices that will be newly set in the future.

When the relationship between nominal and real variables is recast as a set of industry Phillips curves in (3.1.2), these take on a form with some similarities to that of the standard New Keynesian Phillips curve (NKPC), which is itself given by:

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa(\eta_y \mathsf{y}_t + \eta_z \mathsf{z}_t) \tag{3.1.4}$$

This is the economy-wide Phillips curve that would be obtained were all the price-adjustment probabilities equal. In both (3.1.2) and (3.1.4), current inflation depends positively on expected future inflation because when firms anticipate that their own prices are likely to remain sticky during a period in which others' prices are rising, they want larger price increases today to keep pace. The key difference between (3.1.2) and (3.1.4) is the presence of the vector of relative prices $\boldsymbol{\rho}_t$ when there is heterogeneity in the speed of price adjustment between industries.

Like the NKPC in (3.1.4), the system of equations (3.1.2) for an economy with heterogeneity has no explicit dependence on past inflation because of the assumption of Calvo price setting in each industry, and at first glance, it might appear that there are no state variables at all. But this is not true because the current vector of inflation rates $\boldsymbol{\pi}_t$ and the current vector of relative prices $\boldsymbol{\rho}_t$ cannot move independently of one another. Taking as given the past vector of relative prices $\boldsymbol{\rho}_{t-1}$, current relative prices are necessarily given by $\boldsymbol{\rho}_t = \boldsymbol{\rho}_{t-1} + \boldsymbol{\mathcal{R}}\boldsymbol{\pi}_t$ since $\boldsymbol{\rho}_t = \boldsymbol{\mathcal{R}}\mathbf{P}_t$ and $\boldsymbol{\pi}_t = \mathbf{P}_t - \mathbf{P}_{t-1}$. Thus past relative prices should be counted as a state variable in (3.1.2).

While past economy-wide inflation is not in itself a state variable, it should not be thought that it exerts no influence on relative prices, with these being affected only by idiosyncratic factors. When there are differences in the speed of price adjustment, shocks affecting economy-wide inflation will call forth different price responses across industries, which disturb the relative price vector. When this vector then becomes a state variable for next period's inflation, past economy-wide inflation can influence current inflation through this channel.

A more precise examination of this mechanism requires that the determinants of prices and inflation be decomposed into forwards- and backwards-looking components.

Proposition 2 *For each non-degenerate distribution of price-adjustment frequencies $\{\alpha_i\}_{i=1}^{n}$ there exists a unique $n \times n$ matrix $\boldsymbol{\Lambda}$ with n distinct, real, and positive eigenvalues ($n-1$ inside the unit circle, one equal to unity) such that the equation for the vector of price levels $\boldsymbol{\mathsf{P}}_t$ in (3.1.1) can be expressed as:*

$$\boldsymbol{\mathsf{P}}_t = \boldsymbol{\Lambda}\boldsymbol{\mathsf{P}}_{t-1} + \boldsymbol{\Lambda}\sum_{j=0}^{\infty}(\beta\boldsymbol{\Lambda})^j\mathbf{K}\mathbb{E}_t[\eta_y\boldsymbol{\iota}\mathsf{y}_{t+j} + \eta_z\boldsymbol{\mathsf{z}}_{t+j}] \tag{3.1.5}$$

The equation for the vector of inflation rates $\boldsymbol{\pi}_t$ in (3.1.2) is equivalent to:

$$\boldsymbol{\pi}_t = -(\mathbf{I}-\boldsymbol{\Lambda})\boldsymbol{\rho}_{t-1} + \boldsymbol{\Lambda}\sum_{j=0}^{\infty}(\beta\boldsymbol{\Lambda})^j\mathbf{K}\mathbb{E}_t[\eta_y\boldsymbol{\iota}\mathsf{y}_{t+j} + \eta_z\boldsymbol{\mathsf{z}}_{t+j}] \tag{3.1.6}$$

The forwards-looking components in (3.1.5) and (3.1.6) are the same for both prices and inflation, depending on current and expected future values of the output gap y_t and cost-push shocks $\boldsymbol{\mathsf{z}}_t$. The backwards-looking component for prices depends on the past vector of industry price levels $\boldsymbol{\mathsf{P}}_{t-1}$, whereas for inflation it is the past vector of relative prices $\boldsymbol{\rho}_{t-1}$ that matters.

Proof See appendix A.7. ■

The forwards-looking components of equations (3.1.5) and (3.1.6) resemble the "solved forwards" version of the New Keynesian Phillips curve given in (3.1.4):

$$\pi_t = \sum_{j=0}^{\infty}\beta^j\kappa\mathbb{E}_t[\eta_y\mathsf{y}_{t+j} + \eta_z\mathsf{z}_{t+j}] \tag{3.1.7}$$

But unlike the standard New Keynesian Phillips curve, the presence of heterogeneity across industries implies that there is now a backwards-looking component of inflation in (3.1.6) as well. The nature of this component is analysed in the following section.

3.2 Intrinsic inflation inertia

The intrinsic inertial component of the economy-wide inflation rate at time t is defined to be the backwards-looking component of equation (3.1.6) averaged across all n industries. The current level of intrinsic inflation inertia is denoted by ш_t:

$$\text{ш}_t \equiv \boldsymbol{\omega}'\boldsymbol{\Lambda}\boldsymbol{\rho}_{t-1} \tag{3.2.1}$$

At time t the inertial component ш_t is predetermined and depends on the vector of past relative prices $\boldsymbol{\rho}_{t-1}$, which are the state variables for current inflation. While it may seem surprising to define inflation inertia using past relative prices rather than past inflation rates, the relative price vector $\boldsymbol{\rho}_{t-1}$ is systematically related to the history of past inflation rates $\{\pi_{t-1}, \pi_{t-2}, \ldots\}$ among other things. This dependence occurs with heterogeneous price stickiness because shocks that affect the economy-wide inflation rate will call forth a range of price responses across industries and thus perturb relative prices.

In order to understand the relationship between actual inflation and current inflation inertia, the forwards-looking component of equation (3.1.6) is split into two parts, one depending on the output gap

y_t, the other on the exogenous cost-push shocks $\boldsymbol{\mathsf{z}}_t$, which are then averaged across all n industries. The values of these two components at time t are denoted by y_t and $\mathfrak{z}_t$ respectively, and expressions for them are obtained from (3.1.6):

$$\mathrm{y}_t \equiv \eta_y \boldsymbol{\omega}' \boldsymbol{\Lambda} \sum_{j=0}^{\infty} (\beta \boldsymbol{\Lambda})^j \boldsymbol{\kappa} \mathbb{E}_t \mathsf{y}_{t+j} \quad , \quad \mathfrak{z}_t \equiv \eta_z \boldsymbol{\omega}' \boldsymbol{\Lambda} \sum_{j=0}^{\infty} (\beta \boldsymbol{\Lambda})^j \mathbf{K} \mathbb{E}_t \boldsymbol{\mathsf{z}}_{t+j} \tag{3.2.2}$$

The cost-push component $\mathfrak{z}_t$ depends on current and expected future vectors of cost-push disturbances $\boldsymbol{\mathsf{z}}_t$, comprising both economy-wide and idiosyncratic shocks. Similarly, the aggregate demand component y_t in (3.2.2) is a sum involving current and expected future output gaps y_t. The expression for y_t can be written as a linear combination of current and future output gaps, all of which have positive coefficients:

Proposition 3 *The aggregate demand component y_t defined in (3.1.6) can be expressed as follows*

$$\mathrm{y}_t = \sum_{j=0}^{\infty} \mu_j \mathbb{E}_t \mathsf{y}_{t+j} \tag{3.2.3}$$

using a sequence of weights $\{\mu_j\}_{j=0}^{\infty}$. For any non-degenerate distribution of price-adjustment frequencies $\{\alpha_i\}_{i=1}^{n}$, each of the weights μ_j is strictly positive and the sequence decays at a faster rate than β, that is, $0 < \mu_{j+1} < \beta\mu_j$ for all $j \geq 0$.

Proof See appendix A.8. ∎

Therefore y_t is increasing in all current and expected future output gaps, and the weights on future output gaps decay more rapidly with heterogeneous price stickiness than in the New Keynesian Phillips curve (3.1.7) implied by homogeneous Calvo pricing. This means that with heterogeneity, expected future cost movements in the near term become more important than longer-term developments.

The decomposition given in equation (3.1.6), together with the definitions in (3.2.1) and (3.2.2), implies that the determinants of economy-wide inflation π_t can be stated succinctly as:

$$\pi_t = \text{ш}_t + \mathrm{y}_t + \mathfrak{z}_t \tag{3.2.4}$$

Actual inflation π_t is the sum of current intrinsic inflation inertia ш_t, and the aggregate demand y_t and cost-push $\mathfrak{z}_t$ components. Thus one interpretation of intrinsic inflation inertia is that it gives the rate of inflation that would occur purely as a result of the history of firms' pricing decisions, independently of any current or expected future fluctuations in aggregate economic activity or cost-push shocks.

A second interpretation is that intrinsic inflation inertia constitutes a constraint on what monetary policy can achieve in the short run. The inertial component ш_t is predetermined, and the cost-push component $\mathfrak{z}_t$ is exogenous and cannot be affected by monetary policy. The remaining variables in (3.2.4) are the policymaker's goals: current inflation π_t, and the current and expected future levels of the output gap y_t in the component y_t. Therefore (3.2.4) shows that the level of inertia ш_t dictates what inflation rate is consistent with the complete elimination of current and future output gaps in the absence of cost-push shocks, taking the history of firms' pricing decisions as given. In other words, in the case where $\mathfrak{z}_t = 0$, the goal $\mathrm{y}_t = 0$ can be achieved if and only if $\pi_t = \text{ш}_t$. Reducing inflation below the current level of intrinsic inertia would require $\mathrm{y}_t < 0$, and since all the μ_j coefficients in (3.2.3) are strictly positive, this can only happen if there is a currently a recession, or one is expected in the future. Hence there is a real cost of reducing inflation below the current level of intrinsic inflation inertia.[12]

[12]Note that the cost referred to here is the sacrifice ratio: the cumulated loss of economic activity needed to achieve a reduction in inflation. For a full analysis using a utility-based loss function which also takes into account price distortions, see Aoki (2001) and Sheedy (2007*b*).

Equation (3.2.1) gives the definition of intrinsic inflation inertia ш_t at a point in time. It is then important to know how the series $\{\text{ш}_t\}$ evolves over time, because the time path of ш_t is the sequence of inflation rates that must be accommodated if aggregate output is to be stabilized in all future periods, assuming no further cost-push disturbances. The time path of intrinsic inflation inertia is found by solving the system of difference equations that comprises $\boldsymbol{\pi}_t = -(\mathbf{I}-\boldsymbol{\Lambda})\boldsymbol{\rho}_{t-1}$ and $\boldsymbol{\rho}_t = \boldsymbol{\rho}_{t-1} + \boldsymbol{\mathcal{R}}\boldsymbol{\pi}_t$ (the first of these being equation (3.1.6) with the aggregate output and cost-push terms set to zero), starting from a given vector $\boldsymbol{\rho}_{t_0}$ of relative prices at time t_0. The solution of these equations is denoted by $\boldsymbol{\pi}_{t_0+j} = \boldsymbol{\text{Ш}}(j;\boldsymbol{\rho}_{t_0})$, and since $\boldsymbol{\rho}_{t_0+j-1} = \boldsymbol{\rho}_{t_0} + \boldsymbol{\mathcal{R}}\sum_{k=1}^{j-1}\boldsymbol{\pi}_{t_0+k}$, it can be constructed recursively as follows,

$$\boldsymbol{\text{Ш}}(j;\boldsymbol{\rho}_0) = -(\mathbf{I}-\boldsymbol{\Lambda})\left(\boldsymbol{\rho}_0 + \boldsymbol{\mathcal{R}}\left(\sum_{k=1}^{j-1}\boldsymbol{\text{Ш}}(k;\boldsymbol{\rho}_0)\right)\right) \tag{3.2.5}$$

for $j \geq 1$. The time path of intrinsic inflation inertia $\{\text{Ш}(j;\boldsymbol{\rho}_0)\}_{j=1}^{\infty}$ is defined by taking the average of the solution in (3.2.5) across the n industries, so $\text{Ш}(j;\boldsymbol{\rho}_0) \equiv \boldsymbol{\omega}'\boldsymbol{\text{Ш}}(j;\boldsymbol{\rho}_0)$. Hence, starting from any time period t_0 and taking $\boldsymbol{\rho}_{t_0}$ as given, the inertial component of inflation in period t_0+j is given by $\text{ш}_{t_0+j} = \text{Ш}(j;\boldsymbol{\rho}_{t_0})$ under the assumption that all intrinsic inertia has been accommodated in the interim periods and no further cost-push shocks have occurred. This means that $\{\text{Ш}(j;\boldsymbol{\rho}_{t_0})\}_{j=1}^{\infty}$ has the interpretation of being the only sequence of economy-wide inflation rates from time t_0+1 onwards that is consistent with the complete elimination of output gaps over the same horizon, assuming no further cost-push shocks occur after period t_0.

The nature of the time path of intrinsic inertia is of course sensitive to the initial relative price vector $\boldsymbol{\rho}_{t_0-1}$, which can be affected by any number of economy-wide or industry-specific disturbances. But it is possible to give a precise analytical characterization of this time path when the relative price vector is initially at its steady-state value, but is then perturbed by some temporary aggregate disturbance such as a cost-push shock that affects all industries or a change in the output gap brought about by monetary policy. Suppose the economy starts from its steady state at time t_0-1 and then an aggregate disturbance occurs in period t_0. If the disturbance lasts for only one period then it is clear from equation (3.1.6) that the vector of inflation rates $\boldsymbol{\pi}_{t_0}$ is proportional to $\boldsymbol{\Lambda\kappa}$, and economy-wide inflation at t_0 is then a multiple of $\boldsymbol{\omega}'\boldsymbol{\Lambda\kappa}$. Since the economy was in its steady state at t_0-1, the relative price vector $\boldsymbol{\rho}_{t_0}$ must be proportional to $\boldsymbol{\mathcal{R}\Lambda\kappa}$. To construct a time path using equation (3.2.5) that is independent of the magnitude of the original disturbance, the size of the temporary shock is normalized so that its initial impact is to raise economy-wide inflation by 1%. The normalized time path $\text{п}(j)$ is then obtained from (3.2.5) as follows,

$$\text{п}(j) \equiv (\boldsymbol{\omega}'\boldsymbol{\Lambda\kappa})^{-1}\text{Ш}(j;\boldsymbol{\mathcal{R}\Lambda\kappa}) \tag{3.2.6}$$

for $j \geq 1$ and with $\text{п}(0) \equiv 1$. This time path is referred to as the intrinsic impulse response function for inflation. Its construction makes it identical to the impulse response function of inflation to a white-noise aggregate cost-push shock in an economy where monetary policy completely stabilizes the output gap.

The following result shows that the model of heterogeneous price stickiness always generates a negative inertial component of inflation after a positive cost-push shock. More precisely, the intrinsic impulse response function for inflation is always negative except in the very first period when the shock occurs.

Theorem 1 *For any non-degenerate distribution of price-adjustment probabilities $\{\alpha_i\}_{i=1}^{n}$, the intrinsic impulse response function of inflation defined in (3.2.6) has the following properties,*

$$\text{п}(j) < 0 \quad , \quad |\text{п}(j)| < |\text{п}(j-1)| \quad , \quad \lim_{j\to\infty}\text{п}(j) = 0 \tag{3.2.7}$$

for all $j \geq 1$. While the intrinsic impulse response function is initially positive (it is normalized to 1%, so

$\pi(0) = 1$*), it becomes and stays negative in all future periods. It is everywhere decreasing in magnitude and eventually tends to zero.*

Proof See appendix A.10. ■

Theorem 1 has some surprising implications. First, temporary cost-push shocks create a tendency for inflation to switch from above-average to below-average (or vice versa) once a shock has dissipated. Thus the only way for above-average inflation to follow above-average (and below-average to follow below-average) in this model is to introduce some positive extrinsic persistence in either the cost-push shock or the output gap to offset the negative intrinsic inertia.

A second unusual implication concerns the cost of disinflation, or rather the absence of such a cost. One interpretation of the intrinsic impulse response function is that it is the time-path of inflation after a temporary cost-push shock that is consistent with the complete stabilization of the output gap in all current and future periods. While higher-than-average inflation must be tolerated in the period when the shock occurs, once the shock has gone, inflation can fall without cost. Indeed, it can actually fall below average immediately afterwards without any loss of output, and moreover if it merely returned to average, a boom would occur. Therefore the presence of heterogeneity actually makes the task of disinflation even easier than in an economy with a New Keynesian Phillips curve, which is itself widely believed to understate the actual cost of lowering inflation.

In the special case where industries are not subject to idiosyncratic shocks, there is a more direct way of seeing the presence of negative intrinsic inflation inertia:

Proposition 4 *Suppose that there are no industry-specific cost-push shocks, so* $\mathbf{z}_t = \boldsymbol{\iota}\mathsf{z}_t$*. Then there exists a sequence of coefficients* $\{\gamma_j\}_{j=1}^{\infty}$ *such that current inflation inertia* ш_t *defined in (3.2.1) can be exactly expressed in terms of the history of economy-wide inflation rates* $\{\ \pi_{t-1}\ ,\ \pi_{t-2}\ ,\ \ldots\ \}$*:*

$$\text{ш}_t = \sum_{j=1}^{\infty} \gamma_j \pi_{t-j} \tag{3.2.8}$$

Now suppose the dynamics of the driving variables $\eta_y \mathsf{y}_t + \eta_z \mathsf{z}_t$ *from (3.1.6) can be modelled using any stationary AR(1) process. For any non-degenerate distribution of price-adjustment probabilities* $\{\alpha_i\}_{i=1}^{n}$*, it follows that all the coefficients of lagged inflation in (3.2.8) must be negative, that is,* $\gamma_j < 0$ *for all* $j \geq 1$*.*

Proof See appendix A.9. ■

This result shows that the effect of heterogeneity in price stickiness on inflation dynamics is similar to that created by having lags of inflation in the Phillips curve with negative coefficients. This reinforces the finding that heterogeneity implies negative intrinsic inflation inertia.

3.3 Inflation dynamics

The previous section has studied the effect of heterogeneity on intrinsic inflation inertia. But actual inflation persistence also results from persistence that is already present in driving variables such as unit labour costs or the output gap. This section generalizes the earlier results by showing that for a given amount of extrinsic persistence, heterogeneity in the frequency of price adjustment reduces the overall amount of inflation persistence. Thus while overall persistence may be positive or negative depending on whether negative intrinsic inertia outweighs positive extrinsic persistence, it is always possible to conclude that heterogeneity unambiguously reduces total inflation persistence, ceteris paribus.

In what follows, the output gap y_t and the cost-push shocks $\mathbf{z}_t$ are consolidated into the $n \times 1$ vector $\mathbf{x}_t \equiv \eta_y \boldsymbol{\iota} \mathsf{y}_t + \eta_z \mathbf{z}_t$, which is the vector of real marginal costs for each industry. Furthermore, only shocks to the aggregate economy are considered, so that $\mathbf{x}_t = \boldsymbol{\iota}\mathsf{x}_t$. Consequently, the expression for the inflation rates $\boldsymbol{\pi}_t$ in equation (3.1.6) becomes:

$$\boldsymbol{\pi}_t = -(\mathbf{I} - \boldsymbol{\Lambda})\boldsymbol{\rho}_{t-1} + \boldsymbol{\Lambda}\sum_{j=0}^{\infty}(\beta\boldsymbol{\Lambda})^j\boldsymbol{\kappa}\mathbb{E}_t\mathsf{x}_{t+j} \tag{3.3.1}$$

In this framework, extrinsic persistence is defined as any serial correlation in the driving variable $\{\mathsf{x}_t\}$, and overall inflation persistence is serial correlation in the stochastic process for economy-wide inflation $\{\pi_t\}$. It is generally supposed that the driving variable will exhibit positive serial correlation.

A first step in understanding the effects of heterogeneity on overall inflation dynamics is obtained by making a comparison with the standard New Keynesian Phillips curve. Taking the time-series properties of $\{\mathsf{x}_t\}$ as given, the rate of inflation implied by a New Keynesian Phillips curve with discount factor $0 < \beta < 1$ and short-run slope $\kappa > 0$ is denoted by $\Pi_t(\beta, \kappa)$, and is obtained by solving:

$$\Pi_t(\beta, \kappa) = \beta\mathbb{E}_t\Pi_{t+1}(\beta, \kappa) + \kappa\mathsf{x}_t \tag{3.3.2}$$

It turns out that actual inflation π_t in an economy with heterogeneity can be expressed in terms of a combination of current and past inflation rates implied by n hypothetical New Keynesian Phillips curves for economies without heterogeneity. This is proved in the following theorem.

Theorem 2 *There exist a set of n hypothetical discount factors $0 < \tilde{\beta}_i < 1$ and short-run slopes $\tilde{\kappa}_i > 0$ such that actual inflation $\{\pi_t\}$ is obtained from the current and past inflation rates calculated using the corresponding hypothetical New Keynesian Phillips curves defined in equation (3.3.2):*

$$\pi_t = \Pi_t(\beta, \tilde{\kappa}_1) + \sum_{i=2}^{n}\left(\mathfrak{c}_{i0}\Pi_t(\tilde{\beta}_i, \tilde{\kappa}_i) - \sum_{j=1}^{\infty}\mathfrak{c}_{ij}\Pi_{t-j}(\tilde{\beta}_i, \tilde{\kappa}_i)\right) \tag{3.3.3}$$

For each non-degenerate distribution of price-adjustment frequencies $\{\alpha_i\}_{i=1}^{n}$, all the coefficients $\mathfrak{c}_{ij}$ are strictly positive. Hence all past inflation rates enter the equation (3.3.3) with negative coefficients.

Proof See appendix A.11. ∎

Each of the $\Pi_t(\tilde{\beta}_i, \tilde{\kappa}_i)$ is the inflation outcome in an economy with a New Keynesian Phillips curve, which imparts no intrinsic inertia to inflation. Thus the variables $\{\Pi_t(\tilde{\beta}_i, \tilde{\kappa}_i)\}$ display the same amount of overall persistence as the extrinsic persistence found in the driving variable $\{\mathsf{x}_t\}$. However, actual inflation in the economy with heterogeneity is a linear combination of current and past values of these series, with all the coefficients on past values being negative. This reduces overall (positive) inflation persistence relative to what would occur were no heterogeneity present. It could even lead to negative persistence overall were the positive persistence in the driving variable $\{\mathsf{x}_t\}$ sufficiently weak.

A more precise illustration of this result can be given in the case where the stochastic process for the driving variable $\{\mathsf{x}_t\}$ is modelled as a stationary AR(1) process with positive autocorrelation:

$$\mathsf{x}_t = \mathfrak{a}\mathsf{x}_{t-1} + \upsilon_t \quad , \quad \upsilon_t \sim \mathcal{IID}(0, \sigma_\upsilon^2) \tag{3.3.4}$$

The autoregressive coefficient satisfies $0 \leq \mathfrak{a} < 1$. Without heterogeneity, the evolution of inflation would be determined by a New Keynesian Phillips curve of the form (3.1.4). It is straightforward to show that the NKPC implies that the impulse response function of inflation to a cost-push shock is simply proportional to the impulse response function of x_t itself. Let $\mathscr{I}(j)$ denote the impulse response function for inflation in the economy without heterogeneity after a cost-push shock that initially increases inflation by 1%, where j

is the number of periods that have elapsed since the shock occurred. For such an AR(1) process as (3.3.4), this impulse response function is then given by $\mathscr{J}(j) = \mathfrak{a}^j$.

When heterogeneity is present, the stochastic process for inflation in terms of the shock υ_t, and thus inflation's impulse response function, can be obtained by substituting (3.3.4) into (3.3.1) to obtain a representation for $\{\pi_t\}$ of the form,

$$\pi_t = л \sum_{j=0}^{\infty} \mathscr{I}(j)\upsilon_{t-j} \tag{3.3.5}$$

where л is a constant introduced because the impulse response function $\mathscr{I}(j)$ is normalized so that $\mathscr{I}(0) = 1$.

A comparison of inflation persistence with and without heterogeneity in price stickiness can be done by studying the shape and relative rates of decay of the impulse response functions $\mathscr{I}(j)$ and $\mathscr{J}(j)$ for the same cost-push shock.

Theorem 3 *For any non-degenerate distribution of price-adjustment frequencies $\{\alpha_i\}_{i=1}^n$, when the driving variable $\{\mathsf{x}_t\}$ is given by (3.3.4) for any $0 \leq \mathfrak{a} < 1$, the impulse response function $\mathscr{I}(j)$ for inflation with heterogeneity necessarily decays more rapidly than the impulse response $\mathscr{J}(j)$ without heterogeneity,*

$$\mathscr{I}(j) < \mathscr{J}(j) \tag{3.3.6}$$

for all $j \geq 1$. Furthermore, there are only two possible patterns for the impulse response function $\mathscr{I}(j)$ of inflation in the case of heterogeneity:

(i) "Inverted hump-shaped" — $\mathscr{I}(j)$ starts positive; it then declines and becomes negative; it then declines further and has a turning point; finally it increases, but remains negative, while it tends to zero.

(ii) "Fast monotonic decay" — $\mathscr{I}(j)$ starts and remains positive, declines monotonically to zero, but at a faster rate than $\mathscr{J}(j)$.

There is a threshold for the extent of extrinsic persistence $\mathfrak{a}$ below which the economy is in case (i), and above which it is in case (ii).

Proof See appendix A.12 ■

In all cases, the impulse response function of inflation exhibits less persistence with heterogeneity than without it. This is manifested in the more rapid decay of the former relative to the latter. The faster decay occurs because the negative intrinsic inertia implied by heterogeneity cancels out some of the positive extrinsic persistence, leading to lower overall persistence. When the (positive) extrinsic persistence is sufficiently weak, the negative intrinsic inertia dominates and the impulse response function switches from positive to negative at some point, taking on an inverted hump shape. But when extrinsic persistence is dominant, the impulse response function remains monotonic, but still decays more rapidly compared to the case of homogeneity.

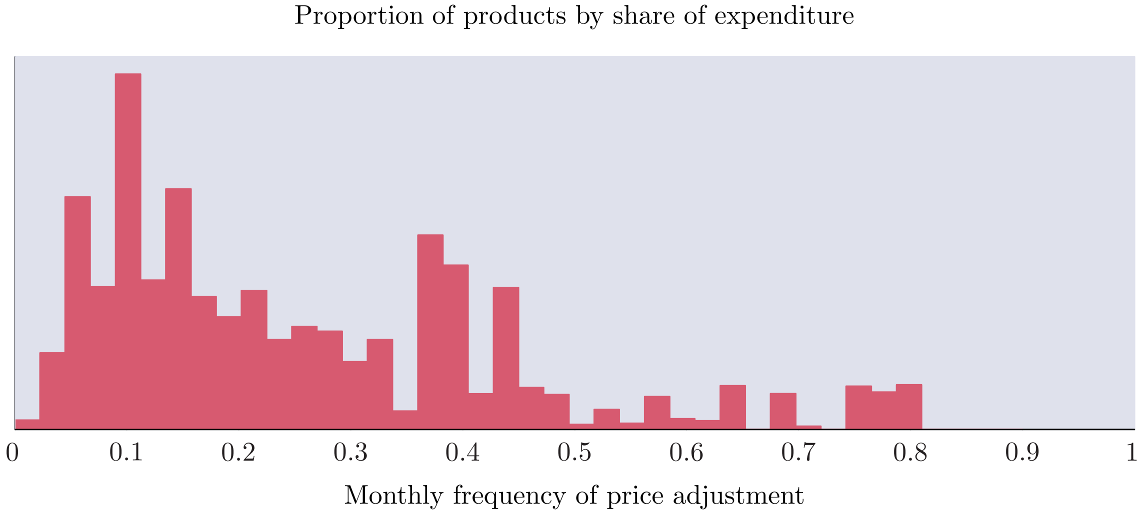
4 Discussion

4.1 Quantitative significance of the results

While the results of section 3 clearly establish the qualitative effects on inflation of introducing heterogeneity in price stickiness, the quantitative importance of the results remains to be assessed. This is done by calibrating a model with a range of industries that mimics the dispersion of price-adjustment frequencies found for the U.S. by Bils and Klenow (2004). The calibrated model is then used to obtain both the intrinsic impulse response function of inflation, and the overall impulse response function in the presence of some extrinsic persistence.

The Bils and Klenow dataset is derived from the U.S. Bureau of Labor Statistics (BLS) survey of individual prices, which underlies the construction of the Consumer Price Index (CPI). They present average monthly frequencies of price adjustment for 350 product categories (entry-level items, or ELIs), for the years between 1995 and 1997. Each of these categories is treated as a separate "industry" for the purposes of this paper. Hence n is set to 350, and the distribution $\{\alpha_i\}_{i=1}^n$ is taken directly from Bils and Klenow's results.[13] The distribution of industry sizes $\{\omega_i\}_{i=1}^n$ is derived from the share of each ELI in the Consumer Expenditure Survey (CEX). The ELIs in the Bils and Klenow dataset comprise 68.9% of the total consumer expenditure according to the CEX. Here it is assumed that their results are representative of the whole U.S. economy, so industry sizes ω_i are set as proportional to the CEX weights.[14] When this distribution is used, the weighted average of the monthly frequency of price adjustment across industries is 0.261, and the standard deviation is 0.189. A histogram of the distribution of price-adjustment frequencies is plotted in Figure 1.

Figure 1: Calibrated distribution of price-adjustment frequencies



Source: Bils and Klenow (2004).

The other parameters of the model are determined as follows. The discount factor β is set to 0.998 so that it is consistent with a 2% annual real interest rate when one time period is equal to a month (since the steady-state annual real interest rate is equal to $(1/\beta)^{12} - 1$). The intra-industry elasticity of substitution ε_f is set to 6, yielding an average markup of price on marginal cost of 20% (recall that equation (2.2.3) implies that the average (gross) markup is given by $\varepsilon_\mathsf{f}/(\varepsilon_\mathsf{f} - 1)$, which is equal to 1.2 when $\varepsilon_\mathsf{f} = 6$). The inter-industry elasticity of substitution ε_s is 1, so that consumers are much more likely to substitute within industries than across industries. The elasticity η_{cy} of a firm's real marginal cost $\mathcal{C}_Y\left(Y_t(\imath); Y_t^*, \mathcal{Z}_{it}\right)$ with respect to a change in its own output $Y_t(\imath)$ is one of the determinants of real rigidity in the model (real rigidity is decreasing in η_{cy}). In order to explain why firms keep their prices sticky when menu costs are small, Ball and Romer (1990) argue that a high degree of real rigidity is needed. The parameter η_{cy} is set to 0.1 here. Knowledge of η_{cy} together with ε_f and ε_s then allows the value of the coefficient η_ρ appearing in (2.2.8) and (3.1.2) to be obtained.[15] Finally, to calculate the overall impulse response function (in addition to just the intrinsic impulse response function) it is necessary to make an assumption about the extent of extrinsic persistence. Extrinsic persistence is modelled by assuming the driving variable x_t is an AR(1) process, as in equation (3.3.4). The autoregressive parameter $\mathfrak{a}$ is set to 0.75. The values of all the

[13]Column "Freq" of Table A1 in Bils and Klenow (2004).
[14]Column "Weight" of Table A1 in Bils and Klenow (2004), rescaled so that the weights sum to 100%.
[15]See equation (A.1.6) in appendix A.1 for details.

Table 1: Calibrated parameter values

Description	Parameter	Value
Number of industries	n	350
Mean price-adjustment frequency (monthly)	$\bar{\alpha}$	0.261
Standard deviation of price-adjustment frequencies	σ_α	0.189
Discount factor (monthly)	β	0.998
Intra-industry elasticity of substitution	ε_{f}	6
Inter-industry elasticity of substitution	ε_{s}	1
Elasticity of marginal cost w.r.t. output (real rigidity)	η_{cy}	0.1
Extrinsic persistence (monthly)	$\mathfrak{a}$	0.75

parameters used in the calibration are listed in Table 1.

The impulse response function of inflation is calculated both with and without heterogeneity present to isolate the impact of heterogeneity in price stickiness on inflation persistence. The model without heterogeneity assumes that there is just one common probability of price adjustment that applies to all firms (as in the standard Calvo model). This probability is taken to be the average $\bar{\alpha}$ of the distribution $\{\alpha_i\}_{i=1}^n$ used in the case of heterogeneity, as given in Table 1.

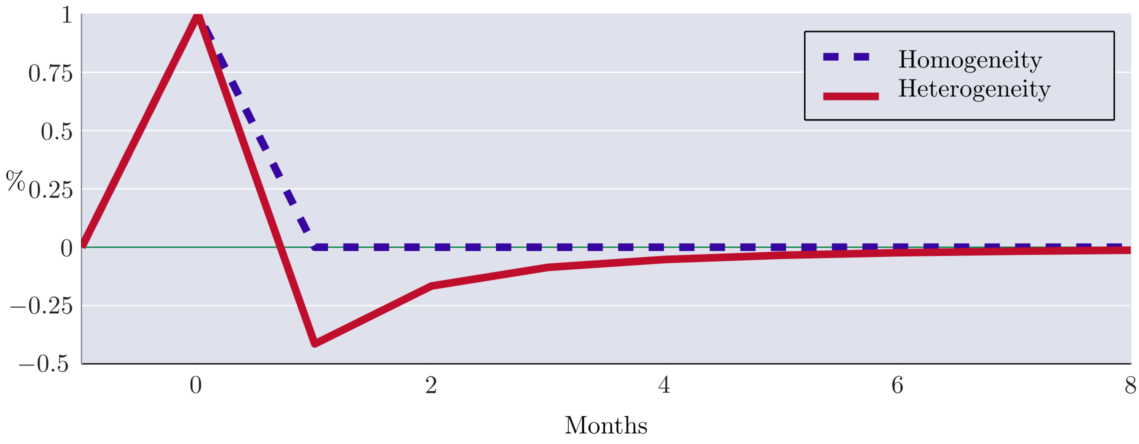
First consider intrinsic inflation inertia alone. The intrinsic impulse response function is the only path of inflation that is consistent with complete stabilization of the output gap after a (positive) white-noise cost-push shock has occurred. Figure 2 plots the intrinsic impulse response function for the two cases of homogeneous and heterogeneous price stickiness. To aid comparison, the impulse response functions are scaled so that they both begin at 1%. This can be done because measures of persistence are independent of scale. With homogeneity this impulse response function returns to zero in the period after the shock, illustrating the lack of intrinsic inflation inertia generated by the standard Calvo model. With heterogeneity, the effects of the shock do not die away immediately, but the subsequent effect is clearly negative because the intrinsic impulse response function falls and stays below zero from one period after the shock onwards. Immediately after this reversal, the negative effect on inflation is equivalent to more than 40% of the initial positive impact when the shock occurred. Thus there is a noticeable and significant change from positive to negative in the intrinsic impulse response function when heterogeneity is present, as compared to an immediate return to zero without heterogeneity.

Figure 3 makes the comparison between the cases of homogeneity and heterogeneity in the presence of some extrinsic persistence. As was the case with intrinsic persistence, the addition of heterogeneity makes a substantial difference to overall inflation persistence. The impulse response function with heterogeneity decays almost twice as fast than that with homogeneity, and switches from positive to negative before returning to zero in the long run. It thus has the "inverted hump shape" referred to in Theorem 3, even with a high degree of extrinsic persistence. For yet higher degrees of extrinsic persistence the pattern would eventually switch to one of monotonic decay in both cases, but the decay under heterogeneity would continue to be much more rapid. Therefore, Figures 2 and 3 together indicate that the qualitative features identified in section 3 are also likely to be quantitatively important with a distribution of price-adjustment frequencies that matches what was found by Bils and Klenow for the U.S. economy.

4.2 Hazard function analysis

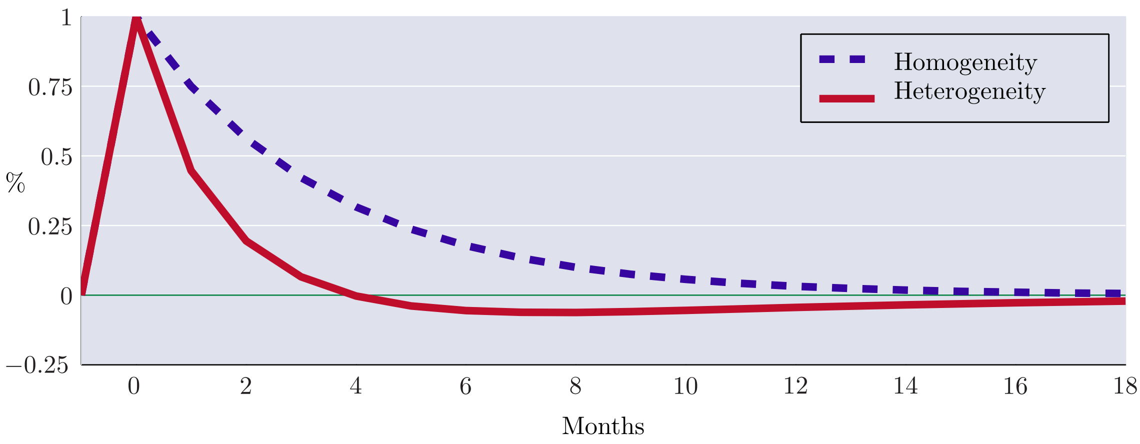
The underlying logic behind why heterogeneity reduces inflation persistence can be understood in terms of the hazard function for price adjustment. This hazard function gives the probability that a firm changes

Figure 2: Intrinsic impulse response functions for inflation with and without heterogeneity



Notes: The intrinsic impulse response function in the case of heterogeneity is obtained from equation (3.2.6) using the method described in section 3.2. The calibrated parameters of the model are given in Table 1 and the distribution of price-adjustment frequencies $\{\alpha_i\}_{i=1}^{n}$ is plotted in Figure 1. The intrinsic impulse response function with homogeneity is obtained immediately from the properties of the New Keynesian Phillips curve in (3.1.4) and (3.1.7). Both impulse response functions are normalized to 1% at time 0 to aid comparison.

Figure 3: Overall impulse response functions for inflation with and without heterogeneity



Notes: The impulse response function in the case of heterogeneity is obtained from equations (3.3.1) and (3.3.4) using the calibrated parameters from Table 1 and the distribution of price-adjustment frequencies $\{\alpha_i\}_{i=1}^{n}$ plotted in Figure 1. The impulse response function with homogeneity is obtained by solving (3.1.7) and (3.3.4). Both impulse response functions are normalized to 1% at time 0 to aid comparison.

its price as a function of the time that has elapsed since its previous price change. Conditional on a firm being in a particular industry, the hazard function used in this paper is flat because of the assumption of Calvo price setting. But when the frequency of price adjustment differs across industries, the economy-wide hazard function need not be flat even when each industry's is. Indeed, the economy-wide hazard function must be downwards sloping when there is such heterogeneity.[16] The following discussion shows how this result can be deduced as a simple consequence of Bayes' theorem, and then goes on to analyse why a downwards-sloping hazard function has the effect of reducing inflation persistence.

At any time, the set of all firms can be partitioned into those that make a price adjustment and those whose prices remain sticky. At time $t = 0$, these sets are denoted by $\mathcal{A}_0$ and $\mathcal{A}_0^c$ respectively, where ^{c} indicates the complement of a set. The industry-i probability of price adjustment α_i is by definition the probability of the event $\mathcal{A}_0$ given Ω_i, so $\alpha_i \equiv \mathbb{P}(\mathcal{A}_0|\Omega_i)$. The overall probability of price adjustment for all firms in the economy is denoted by $\alpha \equiv \mathbb{P}(\mathcal{A}_0)$, which can also be expressed as the average $\mathbb{E}_I[\mathbb{P}(\mathcal{A}_0|\Omega_i)]$, where $\mathbb{E}_I[\cdot]$ denotes the expectation taken over all n industries in the economy using industry weights ω_i.

Now consider the probability of price adjustment one period afterwards at time $t = 1$. In this time period, let $\mathcal{A}_1$ denote the set of firms that change price, and $\mathcal{A}_1^c$ the set that leave their prices fixed. To find the hazard function of price changes for the whole economy it is necessary to evaluate the probability of $\mathcal{A}_1$ conditional on whether $\mathcal{A}_0$ or $\mathcal{A}_0^c$ occurred at $t = 0$.[17] Denote the former probability by $\alpha_{\mathcal{A}} \equiv \mathbb{P}(\mathcal{A}_1|\mathcal{A}_0)$ and the latter by $\alpha_{\mathcal{A}^c} \equiv \mathbb{P}(\mathcal{A}_1|\mathcal{A}_0^c)$. Bayes' theorem is now invoked to show that heterogeneity always implies $\alpha_{\mathcal{A}} > \alpha_{\mathcal{A}^c}$.

The Calvo pricing assumption means that, conditional on a firm being in a particular industry, the probability of price adjustment is no different at time $t = 1$ than it was at time $t = 0$ irrespective of whether a price change occurred at $t = 0$. Hence, $\mathbb{P}(\mathcal{A}_1|\Omega_i) = \mathbb{P}(\mathcal{A}_0|\Omega_i) = \alpha_i$. Because the industry a firm is based in is the ultimate determinant of its price-adjustment probability, the probability of a firm from an unknown industry making a price change one period after an earlier price change can be written as $\alpha_{\mathcal{A}} = \mathbb{E}_{\mathcal{A}_0}[\mathbb{P}(\mathcal{A}_1|\Omega_i)]$, where $\mathbb{E}_{\mathcal{A}_0}[\cdot]$ denotes the expectation taken only over the set $\mathcal{A}_0$ of firms that change price at $t = 0$. Using the definition of conditional expectation, this can be rewritten as an expectation taken over the set I of all industries as follows:

$$\mathbb{E}_{\mathcal{A}_0}[\mathbb{P}(\mathcal{A}_1|\Omega_i)] = \mathbb{E}_I\left[\mathbb{P}(\mathcal{A}_1|\Omega_i)\frac{\mathbb{P}(\Omega_i|\mathcal{A}_0)}{\mathbb{P}(\Omega_i)}\right] \tag{4.2.1}$$

Bayes' theorem is then applied to replace the probability of a firm being in industry i conditional on its being observed to change price at $t = 0$ with a term involving instead the probability of observing a price change conditional on a firm being in industry i:

$$\mathbb{P}(\Omega_i|\mathcal{A}_0) = \frac{\mathbb{P}(\mathcal{A}_0|\Omega_i)\mathbb{P}(\Omega_i)}{\mathbb{P}(\mathcal{A}_0)} \tag{4.2.2}$$

By combining equations (4.2.1) and (4.2.2), and using the equivalent versions of these for the probability $\alpha_{\mathcal{A}^c} \equiv \mathbb{E}_{\mathcal{A}_0^c}[\mathbb{P}(\mathcal{A}_1|\Omega_i)]$ of a price change after a period of price stickiness, the following expression is obtained:

$$\alpha_{\mathcal{A}} - \alpha_{\mathcal{A}^c} = \mathbb{E}_I\left[\frac{\mathbb{P}(\mathcal{A}_1|\Omega_i)\mathbb{P}(\mathcal{A}_0|\Omega_i)}{\mathbb{P}(\mathcal{A}_0)} - \frac{\mathbb{P}(\mathcal{A}_1|\Omega_i)(1-\mathbb{P}(\mathcal{A}_0|\Omega_i))}{1-\mathbb{P}(\mathcal{A}_0)}\right] \tag{4.2.3}$$

[16] A similar argument is made by Álvarez *et al.* (2005).

[17] This analysis distinguishes only the probability of changing a price which has been fixed for one period from that of changing a price fixed for more than one period. Calculating the full hazard function requires conditioning on the precise age of the price. To keep things simple, this analysis effectively calculates just the first point on the hazard function and a weighted average of the points that come afterwards. It is possible to extend the results to cover the full hazard function, but considering just these two points is enough to give the intuition.

The above expression can be simplified to show that

$$\alpha_{\mathcal{A}} - \alpha_{\mathcal{A}^c} = \frac{\mathbb{V}_I[\alpha_i]}{\alpha(1-\alpha)} \tag{4.2.4}$$

where $\mathbb{V}_I[\cdot]$ denotes the inter-industry variance operator. Since the variance of the distribution $\{\alpha_i\}_{i=1}^n$ is always strictly positive whenever there is heterogeneity, the probability of a firm changing price given that a price change has just been observed is higher than the probability of a price change conditional on none having been observed in the previous period. Therefore the economy-wide hazard function is downwards sloping if the industry-specific hazard functions are flat, but at levels that differ across industries.

The next step is to argue that a downwards-sloping hazard function implies negative intrinsic inflation inertia (Sheedy, 2007*a*), and thus reduces overall inflation persistence for a given level of extrinsic persistence. The analysis is clearest in the case where there is no extrinsic persistence, so consider again the example of a temporary shock to one of the determinants of inflation. To simplify matters further, suppose that the economy starts from a steady state in which inflation has reached zero prior to $t = 0$. At $t = 0$ the serially uncorrelated shock occurs, and suppose without loss of generality that its size is normalized so that the initial effect on inflation is to raise it by 1% ($\pi_0 = 1$). By $t = 1$ the shock itself has completely dissipated and the aim is to find out what now happens to inflation.

Suppose firms believe there will be no inflation persistence, or in other words, that they think inflation will return to its former level by $t = 1$ when the shock has gone (whether or not this expectation is rational is assessed below). Without the anticipation of further disturbances, each firm will want to match any changes in the average money prices of other firms to restore its original profit-maximizing relative price. For the firms in group $\mathcal{A}_0^c$ that left their prices fixed at time $t = 0$, this entails a 1% money price increase (the "catch-up" effect) because the general price level has risen by 1% since $t = 0$ and no further change is expected. Now consider the firms in group $\mathcal{A}_0$ that did choose new prices at time $t = 0$. Denote the average inflation rate at $t = 0$ for these firms only by $\pi_{\mathcal{A}_0}$. Since their price changes alone created overall inflation of 1%, and because these firms are a subset of all the firms in the economy ($\alpha < 1$), firms in the set $\mathcal{A}_0$ have prices with an average rate of change greater than 1%. In fact, this inflation rate must be equal to $\pi_{\mathcal{A}_0} = \alpha^{-1} > 1$. Therefore this group of firms desires a money price cut of $(\pi_{\mathcal{A}_0} - 1)$% on average to bring their prices back into line with others (the "roll-back" effect).

In summary, at time $t = 1$ there is a set $\mathcal{A}_0^c$ of firms of size $1 - \alpha$ with probability $\alpha_{\mathcal{A}^c}$ of increasing their money prices by 1% ("catch-up"). There is another set $\mathcal{A}_0$ of firms of size α with probability $\alpha_{\mathcal{A}}$ of cutting their prices in money terms by $(\pi_{\mathcal{A}_0} - 1)$% on average ("roll-back").[18] The overall impact on inflation π_1 at $t = 1$ is:

$$\pi_1 = (1-\alpha)\alpha_{\mathcal{A}^c} - \alpha\alpha_{\mathcal{A}}(\pi_{\mathcal{A}_0} - 1) \tag{4.2.5}$$

Whether inflation is positive (above average) or negative (below average) depends on which of the "catch-up" and "roll-back" effects is dominant. Because $\pi_{\mathcal{A}_0}$ is equal to α^{-1} to be consistent with $\pi_0 = 1$, the expression for inflation π_1 in (4.2.5) can be written in terms of the difference between the probabilities $\alpha_{\mathcal{A}}$ and $\alpha_{\mathcal{A}^c}$:

$$\pi_1 = -(1-\alpha)(\alpha_{\mathcal{A}} - \alpha_{\mathcal{A}^c}) \tag{4.2.6}$$

Therefore, the catch-up effect dominates if the hazard function is upwards sloping ($\alpha_{\mathcal{A}^c} > \alpha_{\mathcal{A}}$), leading to above-average inflation in the period $t = 1$ after the initial shock has gone. A flat hazard function ($\alpha_{\mathcal{A}^c} = \alpha_{\mathcal{A}}$) implies that the two effects exactly cancel out and inflation returns to its average value; and a downwards sloping hazard ($\alpha_{\mathcal{A}^c} < \alpha_{\mathcal{A}}$) means a dominant roll-back effect and consequently below-average

[18] The possibility that there are differences in the probability of rolling back associated with differences in the size of the desired roll-back within this group is analysed below.

inflation at $t = 1$. As equations (4.2.4) and (4.2.6) demonstrate, an economy with heterogeneous Calvo price setting necessarily falls within the latter category. Zero overall inflation persistence therefore cannot be a rational expectations equilibrium when there is heterogeneity — even with no extrinsic persistence. Moreover, there is a clear tendency towards negative persistence because the belief that inflation has no persistence results in actual inflation displaying negative persistence. A formal confirmation that the unique rational expectations equilibrium does indeed exhibit negative inflation persistence is provided by the results of section 3.

While the link between heterogeneity and downwards-sloping hazard functions is instructive, there is no formal equivalence between the two. In other words, a model where every firm shares the same downwards-sloping hazard function — and one that is the same as that generated by the model of heterogeneous Calvo pricing — does not have the same implications for inflation dynamics. In fact, the previous analysis in this section has actually understated the tendency towards negative inflation persistence implied by heterogeneity. The understatement has occurred because there is a positive correlation across industries between the size of the desired roll-back of prices and the probability of a roll-back actually occurring. This was neglected in the derivation of equations (4.2.5) and (4.2.6).

To incorporate this correlation into the analysis, let $\pi_{i,\mathcal{A}_0}$ denote the percentage change in prices of those firms from industry i that did choose new prices at time $t = 0$. The average percentage change over all firms that change price is denoted by $\pi_{\mathcal{A}_0}$ as before, and can be calculated from the industry-specific $\pi_{i,\mathcal{A}_0}$ using $\pi_{\mathcal{A}_0} = \mathbb{E}_{\mathcal{A}_0}[\pi_{i,\mathcal{A}_0}]$. Firms from industry i that did change price at $t = 0$ would like to roll back their prices by amount $(\pi_{i,\mathcal{A}_0} - 1)\%$ at $t = 1$, and have probability $\mathbb{P}[\mathcal{A}_1|\Omega_i]$ of doing so. Since the calculation of the catch-up effect is unchanged from before, equation (4.2.5) giving the overall effect on inflation π_1 at $t = 1$ is replaced by:

$$\pi_1 = (1-\alpha)\alpha_{\mathcal{A}^c} - \alpha\mathbb{E}_{\mathcal{A}_0}[\mathbb{P}(\mathcal{A}_1|\Omega_i)(\pi_{i,\mathcal{A}_0} - 1)] \tag{4.2.7}$$

Using the definition of the covariance operator $\mathbb{C}_{\mathcal{A}_0}[\cdot,\cdot]$ over the set of firms changing price at $t = 0$, equation (4.2.7) can be restated as follows:

$$\pi_1 = (1-\alpha)\alpha_{\mathcal{A}^c} - \alpha\mathbb{E}_{\mathcal{A}_0}[\mathbb{P}(\mathcal{A}_1|\Omega_i)]\mathbb{E}_{\mathcal{A}_0}[\pi_{i,\mathcal{A}_0} - 1] - \alpha\mathbb{C}_{\mathcal{A}_0}[\mathbb{P}(\mathcal{A}_1|\Omega_i), \pi_{i,\mathcal{A}_0} - 1] \tag{4.2.8}$$

Using the fact that $\mathbb{E}_{\mathcal{A}_0}[\mathbb{P}(\mathcal{A}_1|\Omega_i)] = \alpha_{\mathcal{A}}$ and $\mathbb{E}_{\mathcal{A}_0}[\pi_{i,\mathcal{A}_0}] = \pi_{\mathcal{A}_0} = \alpha^{-1}$, an equivalent expression for π_1 is:

$$\pi_1 = -(1-\alpha)(\alpha_{\mathcal{A}} - \alpha_{\mathcal{A}^c}) - \alpha\mathbb{C}_{\mathcal{A}_0}[\alpha_i, \pi_{i,\mathcal{A}_0}] \tag{4.2.9}$$

Finally, the expression in (4.2.4) for the difference between the probabilities $\alpha_{\mathcal{A}}$ and $\alpha_{\mathcal{A}^c}$ is substituted into the above:

$$\pi_1 = -\alpha^{-1}\mathbb{V}_I[\alpha_i] - \alpha\mathbb{C}_{\mathcal{A}_0}[\alpha_i, \pi_{i,\mathcal{A}_0}] \tag{4.2.10}$$

When there is a positive correlation between the probability of price adjustment and the size of the desired roll-back of prices, the resulting inflation rate π_1 from (4.2.9) or (4.2.10) is negative, and unambiguously more so than the inflation rate implied by (4.2.6). Thus by ignoring this positive correlation, the magnitude of the roll-back effect is biased downwards. And the correlation should be positive whenever there is heterogeneity in the frequency of price adjustment across industries. This is because it is precisely those firms from industries with more flexible prices that are willing to make larger temporary price increases following macroeconomic shocks (and thus need larger roll-backs in the future) because these price increases can be more quickly reversed. Therefore, the effect of heterogeneity on inflation persistence is actually more negative than the analysis of the implied downwards-sloping hazard function alone would suggest.

5 Conclusions

This paper has shown that differences in the frequency of price adjustment between industries unambiguously reduce overall inflation persistence relative to what would occur if all industries shared the same price-adjustment frequency. By viewing overall inflation persistence as deriving from two sources, intrinsic inertia from the Phillips curve itself, and extrinsic persistence from serial correlation in variables that affect inflation, the paper has shown that heterogeneity in price stickiness actually implies negative intrinsic inertia. This explains why for the given level of extrinsic persistence, heterogeneity lowers overall inflation persistence.

In addition to the analytical results establishing the direction of heterogeneity's effect on persistence, this paper also assesses its quantitative impact. A calibration of the model using U.S. microeconomic evidence on the dispersion of price-adjustment frequencies shows that the effect of heterogeneity on inflation dynamics is quantitatively important. The microeconomic evidence indicates a significant amount of negative intrinsic inertia, and the paper shows that this translates into a substantial reduction in overall inflation persistence relative to the case of no heterogeneity.

But while there is overwhelming microeconomic evidence supporting the model's assumption of a range of price-adjustment frequencies, much less support is found for the macroeconomic implications of this heterogeneity. The nature of this puzzle can be understood by going back to some of the criticisms of the New Keynesian Phillips curve posed by Mankiw (2001). If a non-degenerate distribution of price-adjustment frequencies is added to an otherwise standard New Keynesian model, these criticisms of the NKPC apply even more forcefully to the model with heterogeneity.

First, the New Keynesian Phillips curve is only consistent with a given amount of inflation persistence if a similar degree of persistence is found in the determinants of inflation. With heterogeneity, the situation is made worse because inflation is now less persistent than its determinants. Thus more serial correlation in inflation's driving variables must be explained to justify a given level of inflation persistence.

Second, the New Keynesian Phillips curve does not imply a cost of disinflation. When heterogeneity is introduced, the cost of disinflation is even lower than it otherwise would be under the NKPC. This means that a disinflation which is costless with the NKPC would actually stimulate economic activity in the model with heterogeneity.

Third, the New Keynesian Phillips curve cannot explain why monetary policy has its greatest effect on inflation with a longer lag than it does for real variables such as aggregate output.[19] Once heterogeneity is introduced, inflation's impulse response function to a monetary policy shock decays more rapidly than it would with the NKPC, and so the peak effect on inflation comes even sooner.

The intuition for these macroeconomic implications is straightforward. When inflation occurs as the result of a macroeconomic shock, the underlying price changes come from a group of firms which is not generally representative of all firms in the economy. Because some industries have more flexible prices than others, the group of firms changing price is likely to be drawn disproportionately from those industries with more flexible prices. But this means that the firms which did change their prices in response to a shock are also the most nimble in reversing any price changes once the shock has dissipated. This makes inflation less persistent and disinflation easier to achieve without cost than it otherwise would be.

The problem is that the basic New Keynesian model to which heterogeneity has been added contains no source of (positive) intrinsic inflation inertia that can outweigh the negative intrinsic inertia implied by heterogeneity. It is usually the case that at least some positive intrinsic inertia, such as that resulting from backwards-looking rules of thumb for price setting or indexation, must be added to the standard New

[19]See Christiano, Eichenbaum and Evans (1999) for evidence on this from a structural VAR study of monetary policy shocks. This stylized fact is also widely accepted by central banks.

Keynesian Phillips curve to fit the macroeconomic evidence on inflation dynamics. The results of this paper suggest that when heterogeneity is present, the need for these sources of intrinsic persistence is even greater still. Thus, more research on sources of positive intrinsic inflation inertia is required if the microeconomic evidence on price setting is to be reconciled with what is known about aggregate inflation dynamics.

References

Altissimo, F., Ehrmann, M. and Smets, F. (2006), "Inflation persistence and price setting behaviour in the Euro area: A summary of the IPN evidence", *Occasional paper 46*, European Central Bank. 1

Altissimo, F., Mojon, B. and Zaffaroni, P. (2007), "Fast micro and slow macro: Can aggregation explain the persistence of inflation?", *Working paper 729*, European Central Bank. 3

Álvarez, L. J., Burriel, P. and Hernando, I. (2005), "Do decreasing hazard functions for price changes make any sense?", *Working paper 461*, European Central Bank. 2, 18

Anton, H. (1994), *Elementary Linear Algebra*, John Wiley & Sons, New York, 7th edition. 29

Aoki, K. (2001), "Optimal monetary policy responses to relative price changes", *Journal of Monetary Economics*, **48**:55–80. 2, 10

Ball, L. and Romer, D. (1990), "Real rigidities and the non-neutrality of money", *Review of Economic Studies*, **57(2)**:183–203. 15

Bellman, R. (1960), *Introduction to Matrix Analysis*, McGraw-Hill, New York. 28

Benigno, P. (2004), "Optimal monetary policy in a currency area", *Journal of International Economics*, **63(2)**:293–320. 2

Bils, M. and Klenow, P. (2004), "Some evidence on the importance of sticky prices", *Journal of Political Economy*, **112**:947–985. 1, 14, 15

Blinder, A., Canetti, E. R. D., Lebow, D. E. and Rudd, J. B. (1998), *Asking About Prices: A New Approach to Understanding Price Stickiness*, Russell Sage Foundation, New York. 1

Calvo, G. A. (1983), "Staggered prices in a utility-maximizing framework", *Journal of Monetary Economics*, **12**:383–398. 1, 5

Cavalho, C. (2005), "Heterogeneity in price setting and the real effects of monetary shocks", *Working paper*, Princeton University. 2

Christiano, L. J., Eichenbaum, M. J. and Evans, C. L. (1999), "Monetary policy shocks: What have we learned and to what end?", in J. B. Taylor and M. Woodford (eds.), "Handbook of Macroeconomics", volume 1, Elsevier Science. 21

——— (2005), "Nominal rigidities and the dynamic effects of a shock to monetary policy", *Journal of Political Economy*, **113**:1–45. 1

de Walque, G., Smets, F. and Wouters, R. (2006), "Price shocks in general equilibrium: Alternatve specifications", *CESifo economic studies*, IFO Institute for Economic Research. 2

Dhyne, E., Álvarez, L., Le Bihan, H., Veronese, G., Dias, D., Hoffmann, J., Jonker, N., Lünnemann, P., Rumler, F. and Vilmunen, J. (2005), "Price-setting in the Euro area: Some stylised facts from individual consumer price data", *Working paper 524*, European Central Bank. 1

Dixon, H. and Kara, E. (2005), "Persistence and nominal inertia in a generalized Taylor economy: How longer contracts dominate shorter contracts", *Working paper 489*, European Central Bank. 2

Fuhrer, J. C. (2005), "Intrinsic and inherited inflation persistence", *Working paper 05-8*, Federal Reserve Bank of Boston. 1

Fuhrer, J. C. and Moore, G. R. (1995), "Inflation persistence", *Quarterly Journal of Economics*, **109**:197–230. 1

Galí, J. and Gertler, M. (1999), "Inflation dynamics: A structural econometric analysis", *Journal of Monetary Economics*, **44**:195–222. 1

Granger, C. W. J. (1980), "Long memory relationships and the aggregation of dynamic models", *Journal of Econometrics*, **14**:227–238. 3

Mankiw, N. G. (2001), "The inexorable and mysterious trade-off between inflation and unemployment", *Economic Journal*, **111**:C45–61. 21

Sheedy, K. D. (2007*a*), "Intrinsic inflation inertia", *Working paper*, London School of Economics. 1, 2, 3, 19

——— (2007*b*), "Resistance to inflation persistence: Optimal monetary policy commitment", *Working paper*, London School of Economics. 10

Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, New Jersey. 1, 6

A Technical appendix

A.1 Steady state and log linearizations

The steady state around which the first-order approximations of the model's equations are taken is characterized by a constant inflation rate, constant real output growth, a constant nominal interest rate, and the absence of cost-push shocks:

$$\mathit{\Pi}_t = \bar{\mathit{\Pi}} \quad , \quad G_t = \bar{G} \quad , \quad \mathcal{I}_t = \bar{\mathcal{I}} \quad , \quad \mathcal{Z}_{it} = \bar{\mathcal{Z}}_i = 1 \tag{A.1.1}$$

For simplicity it is assumed that the constant rates of inflation and real output growth are zero, so the steady-state gross inflation rate is $\bar{\mathit{\Pi}} = 1$, and the steady-state gross real growth rate is $\bar{G} = 1$. It is assumed that the steady-state real interest rate is positive, which requires that the steady-state gross nominal interest rate satisfies $\bar{\mathcal{I}} > 1$. It is more convenient to represent the interest rate as a discount factor, so define $\beta \equiv \bar{\mathcal{I}}^{-1}$, which must satisfy $0 < \beta < 1$.

The aim is to find the steady-state values of $r_{it} = \bar{r}_i$, $\varrho_{it} = \bar{\varrho}_i$ and $\mathcal{Y}_t = \bar{\mathcal{Y}}$. Since $\bar{\mathit{\Pi}} = 1$, the second part of (2.2.6) shows that $\bar{r}_i = \bar{\varrho}_i$ for all i. By evaluating (2.2.4) at the steady state (A.1.1) and using the fact that $\bar{r}_i = \bar{\varrho}_i$, it is seen that $\bar{r}_i = \bar{r}$ and $\bar{\varrho}_i = \bar{\varrho}$ for all i. Then by substituting $\bar{\varrho}_i = \bar{\varrho}$ into the first part of (2.2.6), it follows that $\bar{\varrho} = 1$, and hence $\bar{r} = 1$. Finally by substituting these results back into (2.2.4), the steady-state output gap is found. In summary, the steady state implied by the assumptions is:

$$\bar{r}_i = 1 \quad , \quad \bar{\varrho}_i = 1 \quad , \quad \bar{\mathcal{Y}} = \left(\frac{\varepsilon_{\mathsf{f}} - 1}{\varepsilon_{\mathsf{f}}}\right)^{\frac{1}{\eta_{cy}}} \tag{A.1.2}$$

The equations of the model are now log-linearized around the steady state defined by (A.1.1) and (A.1.2). All second- and higher-order terms are suppressed in the following equations and throughout the paper. By log-linearizing the price level equations in (2.2.6):

$$\sum_{i=1}^{n} \omega_i \rho_{it} = 0 \quad , \quad \rho_{it} = \sum_{j=0}^{\infty} \alpha_i (1-\alpha_i)^j \left\{ \mathsf{r}_{i,t-j} - \sum_{k=0}^{j-1} \pi_{t-k} \right\} \tag{A.1.3}$$

Using the properties of the steady state in (A.1.1) and (A.1.2), and the definitions of $\varrho_{it} \equiv P_{it}/P_t$ and $r_{it} \equiv R_{it}/P_t$, it follows that $\rho_{it} = \mathsf{P}_{it} - \mathsf{P}_t$, $\mathsf{r}_{it} = \mathsf{R}_{it} - \mathsf{P}_t$, and $\pi_t = \mathsf{P}_t - \mathsf{P}_{t-1}$. Hence the results in (2.2.7) can be deduced from equation (A.1.3).

Now consider a log linearization of the reset price equation (2.2.4),

$$\begin{aligned} \mathsf{r}_{it} = \frac{1-\beta(1-\alpha_i)}{1+\varepsilon_{\mathsf{f}}\eta_{cy}} \sum_{\tau=t}^{\infty} (\beta(1-\alpha_i))^{\tau-t} \mathbb{E}_t \Bigg[& \sum_{s=t+1}^{\tau} \{ \mathsf{G}_s + (\varepsilon_{\mathsf{f}} + (1+\varepsilon_{\mathsf{f}}\eta_{cy}))\, \pi_s - \mathsf{i}_s \} \\ & + (\varepsilon_{\mathsf{f}} - \varepsilon_{\mathsf{s}})(1+\eta_{cy})\rho_{i\tau} + \eta_{cy}\mathsf{y}_\tau + \mathsf{z}_{i\tau} - \sum_{s=t+1}^{\tau} \{ \mathsf{G}_s + \varepsilon_{\mathsf{f}}\pi_s - \mathsf{i}_s \} - (\varepsilon_{\mathsf{f}} - \varepsilon_{\mathsf{s}})\rho_{i\tau} \Bigg] \end{aligned} \tag{A.1.4}$$

where $\mathsf{i}_t \equiv \log \mathcal{I}_t - \log \bar{\mathcal{I}}$ denotes the log deviation of the gross nominal interest rate $\mathcal{I}_t$. This expression can be simplified as follows:

$$\mathsf{r}_{it} = \frac{1-\beta(1-\alpha_i)}{1+\varepsilon_{\mathsf{f}}\eta_{cy}} \sum_{\tau=t}^{\infty} (\beta(1-\alpha_i))^{\tau-t} \left((\varepsilon_{\mathsf{f}} - \varepsilon_{\mathsf{s}})\eta_{cy}\rho_{i\tau} + \eta_{cy}\mathsf{y}_\tau + \mathsf{z}_{i\tau} + (1+\varepsilon_{\mathsf{f}}\eta_{cy}) \sum_{s=t+1}^{\tau} \pi_s \right) \tag{A.1.5}$$

By substituting $\mathsf{r}_{it} = \mathsf{R}_{it} - \mathsf{P}_t$ and $\pi_t = \mathsf{P}_t - \mathsf{P}_{t-1}$ into (A.1.5) and rearranging, the expression for R_{it} given in (2.2.8) is obtained with the coefficients η_ρ, η_y and η_z defined as follows:

$$\eta_\rho \equiv \frac{1+\eta_{cy}\varepsilon_{\mathsf{s}}}{1+\eta_{cy}\varepsilon_{\mathsf{f}}} \quad , \quad \eta_y \equiv \frac{\eta_{cy}}{1+\eta_{cy}\varepsilon_{\mathsf{f}}} \quad , \quad \eta_z \equiv \frac{1}{1+\eta_{cy}\varepsilon_{\mathsf{f}}} \tag{A.1.6}$$

This completes the necessary log linearizations.

A.2 Proof of Lemma 1

Take the i-th eigenvalue ζ_i^S of $\boldsymbol{\mathcal{S}} \equiv \mathbf{K}\boldsymbol{\mathcal{R}}$ with corresponding eigenvector $\mathbf{v}_i \neq \mathbf{0}$. Since $\boldsymbol{\mathcal{R}} \equiv \mathbf{I} - \boldsymbol{\iota}\boldsymbol{\omega}'$, the requirement $\boldsymbol{\mathcal{S}}\mathbf{v}_i = \zeta_i^S \mathbf{v}_i$ is equivalent to:

$$\mathbf{K}\mathbf{v}_i - \mathbf{K}\boldsymbol{\iota}(\boldsymbol{\omega}'\mathbf{v}_i) = \zeta_i^S \mathbf{v}_i \tag{A.2.1}$$

Let v_{ji} denote the j-th element of the $n \times 1$ eigenvector $\mathbf{v}_i$, and $\bar{\mathsf{v}}_i \equiv \boldsymbol{\omega}'\mathbf{v}_i$ the weighted average of the elements of $\mathbf{v}_i$ using the industry sizes ω_j as weights. Because $\mathbf{K} \equiv \operatorname{diag}\{\kappa_i\}_{i=1}^n$ and $\mathbf{v}_i \equiv (\ \mathsf{v}_{1i}\ ,\ \ldots\ ,\ \mathsf{v}_{ni}\ )'$, equation (A.2.1) can be stated as $\kappa_j(\mathsf{v}_{ji} - \bar{\mathsf{v}}_i) = \zeta_i^S \mathsf{v}_{ji}$ for all $j = 1, \ldots, n$. By collecting the terms involving v_{ji} on the left-hand side, this becomes

$$(\kappa_j - \zeta_i^S)\mathsf{v}_{ji} = \kappa_j \bar{\mathsf{v}}_i \tag{A.2.2}$$

again for all $j = 1, \ldots, n$.

Now consider the special case where the eigenvalue ζ_i^S is exactly equal to one of the Phillips curve slopes, that is, $\zeta_i^S = \kappa_\ell$ for some ℓ. Since $\kappa_\ell > 0$ from the inequalities in (A.7.1), equation (A.2.2) implies that $\bar{\mathsf{v}}_i = 0$. Because (A.7.1) shows that all the Phillips curve slopes κ_j are distinct, it must be the case that $\kappa_j \neq \zeta_i^S$ for all $j \neq \ell$. It then follows from (A.2.2) that $\mathsf{v}_{ji} = 0$ for all $j \neq \ell$ because $\bar{\mathsf{v}}_i = 0$. Consequently, the weighted average $\bar{\mathsf{v}}_i \equiv \boldsymbol{\omega}'\mathbf{v}_i$ is simply equal to $\omega_\ell \mathsf{v}_{\ell i}$. And moreover since $\omega_\ell > 0$, the fact that $\bar{\mathsf{v}}_i = 0$

means that $\mathsf{v}_{\ell i}$ is also zero. Thus, all the elements of $\mathbf{v}_i$ must be zero if $\zeta_i^S = \kappa_\ell$ for some ℓ. But this would imply that $\mathbf{v}_i$ is the zero vector and hence cannot be an eigenvector, contrary to the supposition. Therefore, the case where the eigenvalue ζ_i^S is exactly equal to one of the Phillips curve slopes can be ruled out, so $\zeta_i^S \neq \kappa_j$ for all j. Thus, an expression for the elements of the eigenvector $\mathbf{v}_i$ can be obtained directly from (A.2.2)

$$\mathsf{v}_{ji} = \frac{\kappa_j}{\kappa_j - \zeta_i^S}\bar{\mathsf{v}}_i \tag{A.2.3}$$

for all j and i. From this formula it is immediately apparent that were the weighted average $\bar{\mathsf{v}}_i$ equal to zero then $\mathbf{v}_i$ would be the zero vector, and again could not be an eigenvector. Thus $\bar{\mathsf{v}}_i \neq 0$, and as the eigenvalues are only determined up to a scalar multiple, $\bar{\mathsf{v}}_i$ can be set to 1 without loss of generality. This ensures that all the eigenvectors can be normalized so that $\boldsymbol{\omega}'\mathbf{v}_i = 1$.

By taking a weighted sum using ω_j of the elements v_{ji} in formula (A.2.3) and making use of the normalization $\sum_{j=1}^n \omega_j \mathsf{v}_{ji} = 1$, the following necessary condition is obtained that must be satisfied by any eigenvalue ζ_i^S:

$$\sum_{j=1}^n \omega_j \frac{\kappa_j}{\kappa_j - \zeta_i^S} = 1 \tag{A.2.4}$$

Because the industry sizes ω_j sum to 1, (A.2.4) is equivalent to:

$$\sum_{j=1}^n \omega_j \frac{\zeta_i^S}{\kappa_j - \zeta_i^S} = 0 \tag{A.2.5}$$

It is clear that an eigenvalue of zero is always consistent with equation (A.2.5). Now consider any non-zero eigenvalue $\zeta_i^S \neq 0$. As the eigenvalue is non-zero and is known not to equal any of the Phillips curve slopes κ_j exactly, an equivalent version of equation (A.2.5) can be obtained by multiplying both sides by $\zeta_i^{S^{-1}} \prod_{\ell=1}^n (\kappa_\ell - \zeta_i^S)$ to obtain:

$$\sum_{j=1}^n \omega_j \left(\prod_{\substack{\ell=1 \\ \ell \neq j}}^n (\kappa_\ell - \zeta_i^S) \right) = 0 \tag{A.2.6}$$

Define the following scalar polynomials $\mathfrak{f}_j(z)$ and $\mathfrak{f}(z)$ with reference to the expression in (A.2.6):

$$\mathfrak{f}(z) \equiv \sum_{j=1}^n \omega_j \mathfrak{f}_j(z) \quad , \quad \mathfrak{f}_j(z) \equiv \prod_{\substack{\ell=1 \\ \ell \neq j}}^n (\kappa_\ell - z) \tag{A.2.7}$$

It clear from the construction in (A.2.7) that equation (A.2.6) is equivalent to $\mathfrak{f}(\zeta_i^S) = 0$, making this a necessary condition for any non-zero eigenvalue of $\boldsymbol{\mathcal{S}}$. It is also apparent that each $\mathfrak{f}_j(z)$ and hence $\mathfrak{f}(z)$ is a polynomial of degree $n-1$, with a corresponding set of $n-1$ roots. Because it is known that the $n \times n$ matrix $\boldsymbol{\mathcal{S}}$ has n eigenvalues, and that a zero eigenvalue is consistent with necessary condition (A.2.5), it follows that zero must always be an eigenvalue of $\boldsymbol{\mathcal{S}}$ and that the polynomial equation $\mathfrak{f}(\zeta_i^S) = 0$ is necessary and sufficient for the remaining $n-1$ eigenvalues.

Let the zero eigenvalue be ordered first so that $\zeta_1^S = 0$ without loss of generality. The other eigenvalues ζ_i^S for $i \geq 2$ are the roots of $\mathfrak{f}(z) = 0$. The definition of the polynomial $\mathfrak{f}_j(z)$ in (A.2.7) implies the following expression when it is evaluated at the Phillips curve slopes κ_i:

$$\mathfrak{f}_j(\kappa_i) = \begin{cases} \prod_{\substack{\ell=1 \\ \ell \neq j}}^n (\kappa_\ell - \kappa_j) & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases} \tag{A.2.8}$$

Therefore the polynomial $\mathfrak{f}_j(z)$ is zero when evaluated at any Phillips curve slope except that corresponding to the j-th industry, and hence $\mathfrak{f}(\kappa_i) = \omega_i \mathfrak{f}_i(\kappa_i)$. The sign of this expression can be obtained from (A.2.8) using the chain of inequalities for the Phillips curve slopes in (A.6.6):

$$(-1)^{j-1}\mathfrak{f}_j(\kappa_j) = \prod_{\ell=1}^{j-1}(\kappa_j - \kappa_\ell) \prod_{\ell=j+1}^{n}(\kappa_\ell - \kappa_j) > 0 \tag{A.2.9}$$

Since $\omega_i > 0$ it must be the case that $(-1)^{i-1}\mathfrak{f}(\kappa_i) > 0$ for all i. Thus the function $\mathfrak{f}(z)$ alternates in sign as it is evaluated at each of the Phillips curve slopes κ_i in sequence. Because the function $\mathfrak{f}(z)$ is a polynomial, it is continuous and hence the intermediate value theorem can be applied to the intervals of the real line in which $\mathfrak{f}(z)$ changes sign. For each $i = 2, \ldots, n$, there consequently exists a $\zeta_i^S \in \mathbb{R}$ with $\kappa_{i-1} < \zeta_i^S < \kappa_i$ such that $\mathfrak{f}(\zeta_i^S) = 0$. This yields the set of n real eigenvalues, and the chain of inequalities (A.7.1) follows from (A.6.6).

It is clear from (A.7.1) that all the eigenvalues are distinct, and because of this, the set of n eigenvectors is linearly independent. That the elements of these eigenvectors are real numbers can be seen from the formula in (A.2.3) and the fact that the eigenvalues themselves are real numbers. And since $\zeta_1^S = 0$, the expression for the elements of the eigenvectors in (A.2.3) implies that $\mathsf{v}_{j1} = 1$ for all j, so the vector of 1s is the eigenvector corresponding to the zero eigenvalue. Finally, note that $\mathbf{v}_1 = \boldsymbol{\iota}$ is consistent with the normalization $\boldsymbol{\omega}'\mathbf{v}_1 = 1$. This completes the proof of the lemma.

A.3 Proof of Lemma 2

Let $\mathscr{D}(z) \equiv \left|\boldsymbol{\Psi}(z^{-1})\right|$ be the determinantal equation of the matrix polynomial $\boldsymbol{\Psi}(z)$. Using the definition of $\boldsymbol{\Psi}(z)$ in (A.7.4), if z_0 is a root of the determinantal polynomial, that is $\mathscr{D}(z_0) = 0$, then it is also true that:

$$\left|\boldsymbol{\mathcal{M}} - \left(z_0^{-1} + \beta z_0\right)\mathbf{I}\right| = 0 \tag{A.3.1}$$

Therefore, z_0 is a root of $\mathscr{D}(z) = 0$ if and only if $\zeta^M = z_0^{-1} + \beta z_0$ is an eigenvalue of $\boldsymbol{\mathcal{M}}$. When $z \neq 0$ the equation $z^{-1} + \beta z = \zeta^M$ is equivalent to the following quadratic equation:

$$\beta z^2 - \zeta^M z + 1 = 0 \tag{A.3.2}$$

For a given value of ζ^M, the quadratic equation (A.3.2) has two roots. The lower branch of the quadratic root function is denoted by $\mathscr{Q}(\zeta)$:

$$\mathscr{Q}(\zeta) = \frac{\zeta - \sqrt{\zeta^2 - 4\beta}}{2\beta} \tag{A.3.3}$$

If $\zeta^M \geq 1 + \beta$ then it follows that ${\zeta^M}^2 - 4\beta \geq (1-\beta)^2 > 0$, and so the roots of the quadratic equation (A.3.2) are both real numbers. It is also clear that $\mathscr{Q}(1+\beta) = 1$, and $\zeta^M \geq 1 + \beta$ implies $0 < \mathscr{Q}(\zeta^M) \leq 1$. Thus, the lower branch (A.3.3) is chosen because the root is always on or inside the unit circle when the inequalities in (A.7.7) are satisfied. The first derivative of the quadratic root function $\mathscr{Q}(\zeta)$ in (A.3.3) is

$$\mathscr{Q}'(\zeta) = \frac{1}{2\beta}\left(1 - \frac{\zeta}{\sqrt{\zeta^2 - 4\beta}}\right) \tag{A.3.4}$$

and $\mathscr{Q}'(\zeta^M)$ is strictly negative whenever $\zeta^M \geq 1 + \beta$. By defining $\zeta_i^\Lambda \equiv \mathscr{Q}(\zeta_i^M)$, these properties together with the inequalities for ζ_i^M in (A.7.7) establish the corresponding chain of inequalities (A.7.9) for the real numbers ζ_i^Λ.

Let $\mathbf{D}^\Lambda \equiv \operatorname{diag}\{\zeta_i^\Lambda\}_{i=1}^n$ be the $n \times n$ diagonal matrix of the ζ_i^Λ values. The matrix $\boldsymbol{\Lambda}$ is constructed so that it has n eigenvalue–eigenvector pairs ζ_i^Λ and v_i. This is done by defining $\mathbf{\Lambda} \equiv \mathsf{V}\mathsf{D}^\Lambda\mathsf{V}^{-1}$. Because (A.7.9) implies that $\mathbf{\Lambda}$ has no zero eigenvalue, $\mathbf{\Lambda}$ is certain to be invertible. The definition of the eigenvalues ζ_i^Λ as roots of the quadratic equation in (A.3.2) also ensures that $\beta{\zeta_i^\Lambda}^2 - \zeta_i^M\zeta_i^\Lambda + 1 = 0$ for all i. Because D^Λ and D^M are both diagonal matrices, this set of n scalar quadratic equations can be stated as the following matrix quadratic equation:

$$\beta{\mathsf{D}^\Lambda}^2 - \mathsf{D}^M\mathsf{D}^\Lambda + \mathbf{I} = \mathbf{0} \tag{A.3.5}$$

The matrices $\mathbf{\Lambda}$ and $\boldsymbol{\mathcal{M}}$ share the same set of eigenvectors v_i, or in other words, they are simultaneously diagonalizable. Premultiplication of (A.3.5) by V and postmultiplication by V^{-1} thus demonstrates that the following matrix quadratic equation is always satisfied by $\boldsymbol{\mathcal{M}}$ and $\mathbf{\Lambda}$:

$$\beta\mathbf{\Lambda}^2 - \boldsymbol{\mathcal{M}}\mathbf{\Lambda} + \mathbf{I} = \mathbf{0} \tag{A.3.6}$$

Because $\mathbf{\Lambda}$ is non-singular, equation (A.3.6) implies that $\boldsymbol{\mathcal{M}} = \mathbf{\Lambda}^{-1} + \beta\mathbf{\Lambda}$.

Define the $n \times n$ linear matrix function $\boldsymbol{\mathit{\Lambda}}(z) \equiv \mathbf{I} - \mathbf{\Lambda}z$ using the matrix $\mathbf{\Lambda}$ as constructed above. Then the brackets of the matrix product $\boldsymbol{\mathit{\Lambda}}(\beta z^{-1})\mathbf{\Lambda}^{-1}\boldsymbol{\mathit{\Lambda}}(z)$ in (A.7.8) are multiplied out as follows:

$$(\mathbf{I} - \beta\mathbf{\Lambda}z^{-1})\mathbf{\Lambda}^{-1}(\mathbf{I} - \mathbf{\Lambda}z) = (\mathbf{\Lambda}^{-1} + \beta\mathbf{\Lambda}) - \mathbf{I}z - \beta\mathbf{I}z^{-1} \tag{A.3.7}$$

By comparing the coefficients of each power of z in (A.3.7) with the those of the matrix function $\boldsymbol{\mathit{\Psi}}(z)$ in (A.7.4), and using the expression for $\boldsymbol{\mathcal{M}}$ in (A.3.6), it is clear that $\boldsymbol{\mathit{\Psi}}(z)$ and $\boldsymbol{\mathit{\Lambda}}(\beta z^{-1})\mathbf{\Lambda}^{-1}\boldsymbol{\mathit{\Lambda}}(z)$ are the same matrix function. Therefore, all the claims of the lemma are verified.

A.4 Proof of Lemma 3

The matrix of eigenvectors V is invertible according to Lemma 1, so there is a unique solution given by $\boldsymbol{\varkappa} = \mathsf{V}^{-1}\boldsymbol{\kappa}$. Since this $\boldsymbol{\varkappa}$ satisfies $\mathsf{V}\boldsymbol{\varkappa} = \boldsymbol{\kappa}$, the expression for the elements v_{ji} of matrix V given in (A.2.3) can be used to write out the system of equations determining $\boldsymbol{\varkappa}$ explicitly,

$$\sum_{j=1}^{n} \frac{\kappa_i\varkappa_j}{\kappa_i - \zeta_j^S} = \kappa_i \tag{A.4.1}$$

where the above holds for all $i = 1, \ldots, n$ and recalling that the eigenvectors have been normalized in accordance with Lemma 1 so that $\bar{\mathsf{v}}_i = 1$. Because (A.6.6) implies that κ_i is strictly positive, nothing is lost by cancelling it from both sides of (A.4.1):

$$\sum_{j=1}^{n} \frac{\varkappa_j}{\kappa_i - \zeta_j^S} = 1 \tag{A.4.2}$$

The following identity is used to solve the system of equations in (A.4.2):

$$\sum_{j=1}^{n} \frac{\prod_{\substack{h=1\\h\neq i}}^{n}(\kappa_h - \zeta_j^S)}{\prod_{\substack{k=1\\k\neq j}}^{n}(\zeta_k^S - \zeta_j^S)} \equiv 1 \tag{A.4.3}$$

The above holds for all $i = 1, \ldots, n$, and the first step is to verify this identity before using it to find an explicit formula for $\varkappa_i$. Since it is known from (A.7.1) that all the eigenvalues ζ_j^S of $\boldsymbol{\mathcal{S}}$ are distinct, the identity in (A.4.3) can be multiplied by the non-zero product $\prod_{k=1}^{n}\prod_{l=1}^{k-1}(\zeta_k^S - \zeta_l^S)$ to obtain an equivalent expression (which is also required to hold for all i):

$$\sum_{j=1}^{n}(-1)^{j-1}\prod_{\substack{h=1\\h\neq i}}^{n}\prod_{\substack{k=1\\k\neq j}}^{n}\prod_{\substack{l=1\\l\neq j}}^{k-1}(\kappa_h-\zeta_j^S)(\zeta_k^S-\zeta_l^S)\equiv\prod_{k=1}^{n}\prod_{l=1}^{k-1}(\zeta_k^S-\zeta_l^S) \tag{A.4.4}$$

A special type of matrix known as a Vandermonde matrix is very useful in verifying the identity (A.4.4). For a given sequence of n numbers $\{\zeta_i^S\}_{i=1}^n$, the $n\times n$ Vandermonde matrix $\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\right)$ is defined as,

$$\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\right)\equiv\begin{pmatrix}1 & \zeta_1^S & \cdots & {\zeta_1^S}^{n-1}\\ 1 & \zeta_2^S & \cdots & {\zeta_2^S}^{n-1}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & \zeta_n^S & \cdots & {\zeta_n^S}^{n-1}\end{pmatrix} \tag{A.4.5}$$

where the i-th row of the matrix is a geometric progression in ζ_i^S. The determinant of the Vandermonde matrix in (A.4.5) is equal to the following expression:[20]

$$\left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\}\right)\right|=\prod_{k=1}^{n}\prod_{l=1}^{k-1}(\zeta_k^S-\zeta_l^S) \tag{A.4.6}$$

It is clear from this formula that the identity in (A.4.4) can be restated in terms of determinants of Vandermonde matrices,

$$\left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\right)\right|\equiv\sum_{j=1}^{n}(-1)^{j-1}\left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\backslash\{\zeta_j^S\}\right)\right|\prod_{\substack{h=1\\h\neq i}}^{n}(\kappa_h-\zeta_j^S) \tag{A.4.7}$$

where the above must hold for all i. Let $\mathscr{C}_j$ denote the cofactor of $\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\right)$ in (A.4.5) corresponding to the n-th column and the j-th row:

$$\mathscr{C}_j\equiv(-1)^{n+j}\begin{vmatrix}1 & \zeta_1^S & \cdots & {\zeta_1^S}^{n-2}\\ \vdots & \vdots & & \vdots\\ 1 & \zeta_{j-1}^S & \cdots & {\zeta_{j-1}^S}^{n-2}\\ 1 & \zeta_{j+1}^S & \cdots & {\zeta_{j+1}^S}^{n-2}\\ \vdots & \vdots & & \vdots\\ 1 & \zeta_n^S & \cdots & {\zeta_n^S}^{n-2}\end{vmatrix} \tag{A.4.8}$$

It is immediately apparent from (A.4.5) and (A.4.8) that $\mathscr{C}_j$ is equal to $(-1)^{n+j}$ multiplied by the determinant of the Vandermonde matrix $\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\backslash\{\zeta_j^S\}\right)$, which is generated from the sequence $\{\zeta_k^S\}_{k=1}^n$ with the j-th element deleted:

$$\mathscr{C}_j=(-1)^{n+j}\left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\backslash\{\zeta_j^S\}\right)\right| \tag{A.4.9}$$

The cofactors $\mathscr{C}_j$ of the matrix $\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\right)$ have the property that the determinant $\left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\right)\right|$ can be obtained by multiplying each $\mathscr{C}_j$ by the j-th element in the n-th column of $\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\right)$ and summing along the n-th column. But when the cofactors are multiplied by elements from a different row, the sum is

[20] See Bellman (1960) for a proof.

equal to zero:[21]

$$\sum_{j=1}^{n} \zeta_j^{S^h} \mathscr{C}_j = \begin{cases} 0 & \text{if} \quad h = 0, 1, \ldots, n-2 \\ \left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\right)\right| & \text{if} \quad h = n-1 \end{cases} \tag{A.4.10}$$

To make use of this result, the product appearing in equation (A.4.7) is expanded as a sum of powers of ζ_j^S,

$$\prod_{h=1, h\neq i}^{n} (\kappa_h - \zeta_j^S) = \sum_{h=0}^{n-1} (-1)^h \mathfrak{K}_{i,h} \zeta_j^{S^h} \tag{A.4.11}$$

where the coefficients $\mathfrak{K}_{i,h}$ are given by sums of products of the Phillips curve slopes $\{\kappa_i\}_{i=1}^n$:

$$\mathfrak{K}_{i,h} \equiv \sum_{\substack{\forall(\ell_1,\ldots,\ell_{n-1-h}) \\ \ell_k \in \{1,\ldots,i-1,i+1,\ldots,n\}}} \prod_{k=1}^{n-1-h} \kappa_{\ell_k} \tag{A.4.12}$$

Note in particular that $\mathfrak{K}_{i,n-1} = 1$ for all i. By substituting the expression for the product in (A.4.11) into (A.4.7), that identity is now equivalent to:

$$\left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\}\right)\right| = \sum_{h=0}^{n-1} (-1)^h \mathfrak{K}_{i,h} \sum_{j=1}^{n} (-1)^{j-1} \zeta_j^{S^h} \left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\} \backslash \{\zeta_j^S\}\right)\right| \tag{A.4.13}$$

Using equation (A.4.9), the determinants on the right-hand side of the identity (A.4.13) can be replaced by terms involving the cofactors $\mathscr{C}_j$:

$$\left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\}\right)\right| \equiv \sum_{h=0}^{n-1} (-1)^{h-(n-1)} \mathfrak{K}_{i,h} \sum_{j=1}^{n} \zeta_j^{S^h} \mathscr{C}_j \tag{A.4.14}$$

Then the results for the sums of cofactors stated in (A.4.10) imply that the identity in (A.4.14) is equivalent to:

$$\left|\boldsymbol{\mathcal{V}}\left(\{\zeta_k^S\}_{k=1}^n\}\right)\right| \equiv \mathfrak{K}_{i,n-1} \left|\boldsymbol{\mathcal{V}}\left(\{\zeta_i^S\}_{i=1}^n\}\right)\right| \tag{A.4.15}$$

But this statement is clearly true since the coefficient $\mathfrak{K}_{i,n-1}$ in (A.4.12) is known to equal one for all i. Therefore, the original identity (A.4.3) must be true.

The identity (A.4.3) is now used to verify that the following proposed solution to the system of equations in (A.4.2) is correct:

$$\varkappa_j = \frac{\prod_{h=1}^{n} (\kappa_h - \zeta_j^S)}{\prod_{\substack{k=1 \\ k\neq j}}^{n} (\zeta_k^S - \zeta_j^S)} \tag{A.4.16}$$

By substituting this claim into equation (A.4.2) and cancelling the term $(\kappa_h - \zeta_i^S)$ from both numerator and denominator, the identity (A.4.3) is obtained. Thus the solution given in (A.4.16) must be correct for all values of κ_i and ζ_i^S.

Finally, by using the chain of inequalities for the sequences $\{\kappa_i\}_{i=1}^n$ and $\{\zeta_i^S\}_{i=1}^n$ in (A.7.1), it can be seen that the numerator of (A.4.16) contains $j-1$ negative terms and $n-j+1$ positive, and the denominator contains $j-1$ negative and $n-j$ positive. Hence, it is shown that $\varkappa_j > 0$ for all j, completing the proof.

[21] See any text on linear algebra, for example, Anton (1994).

A.5 Proof of Lemma 4

Define the $n \times 1$ vector $\mathbf{x}_t$ using the expected values of a scalar time-series $\{\mathsf{x}_t\}$:

$$\mathbf{x}_t = \sum_{k=0}^{\infty} (\beta\mathbf{\Lambda})^k \boldsymbol{\kappa} \mathbb{E}_t \mathsf{x}_{t+k} \tag{A.5.1}$$

If $\mathbf{x}_t = \boldsymbol{\iota}\mathsf{x}_t$ then equation (3.1.5) implies that the vector of price levels $\mathbf{P}_t$ can be expressed in terms of $\mathbf{x}_t$:

$$\mathbf{P}_t = \mathbf{\Lambda}\mathbf{P}_{t-1} + \mathbf{\Lambda}\mathbf{x}_t \tag{A.5.2}$$

Repeated backwards substitution of (A.5.2) shows that $\mathbf{P}_t$ can be written in terms of a sum involving current and past values of $\mathbf{x}_t$:

$$\mathbf{P}_t = \sum_{j=0}^{\infty} \mathbf{\Lambda}^{j+1} \mathbf{x}_{t-j} \tag{A.5.3}$$

The expression for $\mathbf{x}_t$ is then substituted into (A.5.3) to obtain:

$$\mathbf{P}_t = \sum_{j=0}^{\infty} \mathbf{\Lambda}^{j+1} \sum_{k=0}^{\infty} (\beta\mathbf{\Lambda})^k \boldsymbol{\kappa} \mathbb{E}_{t-j} \mathsf{x}_{t-j+k} \tag{A.5.4}$$

It is shown in Lemma 2 that the matrix $\mathbf{V}$ diagonalizes $\mathbf{\Lambda}$, which means that $\mathbf{\Lambda} = \mathbf{V}\mathbf{D}^{\Lambda}\mathbf{V}^{-1}$, where $\mathbf{D}^{\Lambda}$ is the matrix of eigenvalues of $\mathbf{\Lambda}$. The matrix $\mathbf{V}$ also diagonalizes all powers of $\mathbf{\Lambda}$ because $\mathbf{\Lambda}^k = \mathbf{V}{\mathbf{D}^{\Lambda}}^k\mathbf{V}^{-1}$. By using this fact and the definition $\boldsymbol{\varkappa} \equiv \mathbf{V}^{-1}\boldsymbol{\kappa}$, equation (A.5.4) is seen to be equivalent to (A.9.1a). Equation (A.9.1b) is then obtained by first-differencing (A.9.1a).

The equations for the aggregate price level $\mathsf{P}_t = \boldsymbol{\omega}'\mathbf{P}_t$ and inflation rate $\pi_t = \boldsymbol{\omega}'\boldsymbol{\pi}_t$ are deduced from their counterparts (A.9.1a) and (A.9.1b) by first noting that Lemma 1 implies that each of the columns $\mathbf{v}_i$ of $\mathbf{V}$ is normalized so that $\boldsymbol{\omega}'\mathbf{v}_i = 1$ and hence $\boldsymbol{\omega}'\mathbf{V} = \boldsymbol{\iota}'$. This together with the fact that $\mathbf{D}^{\Lambda}$ is a diagonal matrix, and $\zeta_1^{\Lambda} = 1$, yields equations (A.9.2a) and (A.9.2b), completing the proof.

A.6 Proof of Proposition 1

The first step is to use the assumption of Calvo pricing for each industry to obtain recursive versions of the price level and profit-maximizing reset price equations. As can be checked by repeated backwards substitution, the following equation is equivalent to the expression for the industry i price level P_{it} given in (2.2.7):

$$\mathsf{P}_{it} = (1-\alpha_i)\mathsf{P}_{i,t-1} + \alpha_i\mathsf{R}_{it} \tag{A.6.1}$$

Likewise, repeated forwards substitution shows that the following is a recursive version of the equation for the reset price R_{it} in (2.2.8):

$$\mathsf{R}_{it} = \beta(1-\alpha_i)\mathbb{E}_t\mathsf{R}_{i,t+1} + (1-\beta(1-\alpha_i))(\mathsf{P}_{it} - \eta_\rho\rho_{it} + \eta_y\mathsf{y}_t + \eta_z\mathsf{z}_{it}) \tag{A.6.2}$$

Substitute equation (A.6.1) into (A.6.2) to eliminate the terms involving the reset price R_{it}:

$$(1+\beta)\mathsf{P}_{it} = \mathsf{P}_{i,t-1} + \beta\mathbb{E}_t\mathsf{P}_{i,t+1} + \frac{\alpha_i(1-\beta(1-\alpha_i))}{1-\alpha_i}(-\eta_\rho\rho_{it} + \eta_y\mathsf{y}_t + \eta_z\mathsf{z}_{it}) \tag{A.6.3}$$

The coefficient of the term in parentheses on the right-hand side of (A.6.3) is the industry-specific component of the slope of the short-run Phillips curve. This depends on the steady-state discount factor β and the probability of price adjustment α_i in industry i. Hence define the following function $\mathscr{K}(\alpha)$ giving the industry-specific slope in terms of a particular price-adjustment frequency α:

$$\mathscr{K}(\alpha) \equiv \frac{\alpha(1-\beta(1-\alpha))}{1-\alpha} \tag{A.6.4}$$

This function has the property that if $0 < \alpha < 1$ then $0 < \mathscr{K}(\alpha) < \infty$. It has derivative

$$\mathscr{K}'(\alpha) = \frac{1-\beta(1-\alpha)^2}{(1-\alpha)^2} \tag{A.6.5}$$

which satisfies $\mathscr{K}'(\alpha) > 0$ for $0 < \alpha < 1$, implying that $\mathscr{K}(\alpha)$ is a strictly increasing function. So if $\kappa_i \equiv \mathscr{K}(\alpha_i)$ denotes the slope of the industry i Phillips curve then from the chain of inequalities for α_i in (2.1.9), a similar chain of inequalities for κ_i is obtained:

$$0 < \kappa_1 < \kappa_2 < \cdots < \kappa_{n-1} < \kappa_n < \infty \tag{A.6.6}$$

Let $\boldsymbol{\kappa}$ be the $n \times 1$ vector containing these industry-specific slopes, and let $\mathbf{K} \equiv \mathrm{diag}\{\kappa_i\}_{i=1}^n$ be the $n \times n$ diagonal matrix containing these slopes along its principal diagonal.

Using the definition of the matrix $\mathbf{K}$ and the vectors of price levels P_t and reset prices R_t, the n equations in (A.6.3) can be stated as

$$(1+\beta)\mathsf{P}_t + \eta_\rho \mathbf{K}\boldsymbol{\rho}_t = \mathsf{P}_{t-1} + \beta\mathbb{E}_t\mathsf{P}_{t+1} + \mathbf{K}(\eta_y\boldsymbol{\iota}\mathsf{y}_t + \eta_z\mathsf{z}_t) \tag{A.6.7}$$

where $\boldsymbol{\iota}$ is a $n \times 1$ vector of 1s. Since the relative price vector $\boldsymbol{\rho}_t$ is given by $\boldsymbol{\rho}_t = \boldsymbol{\mathcal{R}}\mathsf{P}_t$, the left-hand side of (A.6.7) is equivalent to $\boldsymbol{\mathcal{M}}\mathsf{P}_t$, with $\boldsymbol{\mathcal{M}} \equiv (1+\beta)\mathbf{I} + \eta_\rho\mathbf{K}\boldsymbol{\mathcal{R}}$ being a $n \times n$ matrix. It is easily checked that the matrix $\boldsymbol{\mathcal{R}} \equiv \mathbf{I} - \boldsymbol{\iota}\boldsymbol{\omega}'$ has the property that $\boldsymbol{\mathcal{R}}^2 = \boldsymbol{\mathcal{R}}$. Hence $\boldsymbol{\mathcal{R}}$ is idempotent and must therefore also be positive semi-definite. Furthermore, the parameter η_ρ is strictly positive and the matrix $(1+\beta)\mathbf{I}$ is positive definite, as is $\mathbf{K}$ from (A.6.6). By taking these facts together, the matrix $\boldsymbol{\mathcal{M}}$ must be positive definite. Multiplying both sides of (A.6.7) by $\boldsymbol{\mathcal{M}}^{-1}$ yields equation (3.1.1).

To state the pricing equations in terms of inflation rates and relative prices, note that the coefficients of the money price levels on both sides of equation (A.6.3) have the same sum. This means that by cancelling a unit root in the money price level, the equation can be restated in terms of the industry-specific inflation rate $\pi_{it} \equiv \mathsf{P}_{it} - \mathsf{P}_{i,t-1}$ and the relative price ρ_{it} as follows:

$$\pi_{it} = \beta\mathbb{E}_t\pi_{i,t+1} + \kappa_i(-\eta_\rho\rho_{it} + \eta_y\mathsf{y}_t + \eta_z\mathsf{z}_{it}) \tag{A.6.8}$$

Using the definitions of the matrix $\mathbf{K}$ and the vectors of inflation rates $\boldsymbol{\pi}_t$ and relative prices $\boldsymbol{\rho}_t$, equation (3.1.2) is immediately obtained from (A.6.8). This establishes all the claims of the proposition.

A.7 Proof of Proposition 2

The first step in obtaining the forwards- and backwards-looking components of the Phillips curve is to analyse the properties of the $n \times n$ matrix $\boldsymbol{\mathcal{S}} \equiv \mathbf{K}\boldsymbol{\mathcal{R}}$, in particular, its eigenvalues and eigenvectors. A scalar $\zeta^S \in \mathbb{C}$ is said to be an eigenvalue of $\boldsymbol{\mathcal{S}}$ if there exists a non-zero $n \times 1$ vector $\mathsf{v} \in \mathbb{C}^n$ such that $\boldsymbol{\mathcal{S}}\mathsf{v} = \zeta^S\mathsf{v}$. The i-th eigenvalue and eigenvector are denoted by ζ_i^S and v_i. The following result characterizes the properties of these eigenvalues and eigenvectors.

Lemma 1 *The matrix $\boldsymbol{\mathcal{S}}$ has n distinct, real, and non-negative eigenvalues $\zeta_i^S \in \mathbb{R}$, which are without loss of generality ordered to form an increasing sequence. Exactly one eigenvalue is zero and the others are* *interlaced with the sequence of Phillips curve slopes κ_i according to the following chain of inequalities:*

$$0 = \zeta_1^S < \kappa_1 < \zeta_2^S < \kappa_2 < \zeta_3^S < \cdots < \kappa_{n-1} < \zeta_n^S < \kappa_n < \infty \tag{A.7.1}$$

There also exists a corresponding set of n linearly independent and real-valued eigenvectors $\mathbf{v}_i \in \mathbb{R}^n$, and the eigenvector associated with the zero eigenvalue is a vector of 1s. The eigenvectors can be normalized so that $\boldsymbol{\omega}'\mathbf{v}_i = 1$ for all i.

Proof See appendix A.2. ■

The eigenvectors of $\boldsymbol{\mathcal{S}}$ are collected into a $n \times n$ matrix $\mathbf{V} \equiv (\ \mathbf{v}_1\ ,\ \cdots\ ,\ \mathbf{v}_n\ )$. The linear independence of the set of eigenvectors guaranteed by Lemma 1 ensures that $\mathbf{V}$ is non-singular. If $\mathbf{D}^S \equiv \operatorname{diag}\{\zeta_i^S\}_{i=1}^n$ is the diagonal matrix of eigenvalues of $\boldsymbol{\mathcal{S}}$, then eigenvalue-eigenvector relationship can be stated as $\boldsymbol{\mathcal{S}}\mathbf{V} = \mathbf{V}\mathbf{D}^S$. Because $\mathbf{V}$ is invertible, this means that the matrix $\boldsymbol{\mathcal{S}}$ can be diagonalized as follows:

$$\mathbf{V}^{-1}\boldsymbol{\mathcal{S}}\mathbf{V} = \mathbf{D}^S \tag{A.7.2}$$

The equations for the price level vector $\mathbf{P}_t$ in (3.1.1) are equivalent to the following expression since the matrix $\boldsymbol{\mathcal{M}}$ is non-singular:

$$\boldsymbol{\mathcal{M}}\mathbf{P}_t = \mathbf{P}_{t-1} + \beta\mathbb{E}_t\mathbf{P}_{t+1} + \boldsymbol{\mathcal{K}}(\eta_y\boldsymbol{\iota}\mathsf{y}_t + \eta_z\mathbf{z}_t) \tag{A.7.3}$$

By introducing the lag operator $\mathbb{L}$, the forward operator $\mathbb{F}$ (where $\mathbb{F} \equiv \mathbb{L}^{-1}$), and the $n \times n$ matrix function $\boldsymbol{\Psi}(z)$ defined by,

$$\boldsymbol{\Psi}(z) \equiv \boldsymbol{\mathcal{M}} - \mathbf{I}z - \beta\mathbf{I}z^{-1} \tag{A.7.4}$$

the pricing equations in (A.7.3) can be expressed as follows:

$$\mathbb{E}_t[\boldsymbol{\Psi}(\mathbb{L})\mathbf{P}_t] = \boldsymbol{\mathcal{K}}(\eta_y\boldsymbol{\iota}\mathsf{y}_t + \eta_z\mathbf{z}_t) \tag{A.7.5}$$

The matrix function $\boldsymbol{\Psi}(z)$ is factorized using the diagonalization of the $n \times n$ matrix $\boldsymbol{\mathcal{M}}$. Since $\boldsymbol{\mathcal{S}} \equiv \boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{R}}$, the definition of $\boldsymbol{\mathcal{M}}$ given in Proposition 1 is equivalent to $\boldsymbol{\mathcal{M}} = (1+\beta)\mathbf{I} + \eta_\rho\boldsymbol{\mathcal{S}}$. This allows the diagonalization of $\boldsymbol{\mathcal{M}}$ to be obtained easily from that of $\boldsymbol{\mathcal{S}}$, which was found in Lemma 1. Note that (A.7.2) implies that

$$\mathbf{V}^{-1}\boldsymbol{\mathcal{M}}\mathbf{V} = (1+\beta)\mathbf{I} + \eta_\rho\mathbf{D}^S \tag{A.7.6}$$

and since the right-hand side is a diagonal matrix, the same matrix of eigenvectors $\mathbf{V}$ diagonalizes both $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{M}}$. The matrix of eigenvalues of $\boldsymbol{\mathcal{M}}$ is thus obtained from the right-hand side of (A.7.6), and is denoted by the diagonal matrix $\mathbf{D}^M \equiv (1+\beta)\mathbf{I} + \eta_\rho\mathbf{D}^S$. If ζ_i^M is the i-th eigenvalue of $\boldsymbol{\mathcal{M}}$ then $\mathbf{D}^M \equiv \operatorname{diag}\{\zeta_i^M\}_{i=1}^n$ and $\zeta_i^M = (1+\beta) + \eta_\rho\zeta_i^S$. Because η_ρ is a positive constant, the inequalities for ζ_i^S in (A.7.1) imply the corresponding chain of inequalities for the ζ_i^M:

$$1 + \beta = \zeta_1^M < \zeta_2^M < \cdots < \zeta_{n-1}^M < \zeta_n^M < \infty \tag{A.7.7}$$

The next result constructs a factorization of the matrix function $\boldsymbol{\Psi}(z)$ using this diagonalization.

Lemma 2 *There exists a $n \times n$ non-singular matrix $\boldsymbol{\Lambda}$ such that the linear matrix function $\boldsymbol{\Lambda}(z) \equiv \mathbf{I} - \boldsymbol{\Lambda}z$ factorizes the matrix function $\boldsymbol{\Psi}(z)$ defined in (A.7.4) as follows*

$$\boldsymbol{\Psi}(z) = \boldsymbol{\Lambda}(\beta z^{-1})\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}(z) \tag{A.7.8}$$

for all $z \in \mathbb{C}\backslash\{0\}$. The matrix $\mathbf{\Lambda}$ has n distinct, real, and positive eigenvalues $\zeta_i^{\Lambda} \in \mathbb{R}$ satisfying the following chain of inequalities:

$$1 = \zeta_1^{\Lambda} > \zeta_2^{\Lambda} > \cdots > \zeta_{n-1}^{\Lambda} > \zeta_n^{\Lambda} > 0 \tag{A.7.9}$$

There are $n-1$ eigenvalues inside the unit circle and one eigenvalue equal to unity. The matrix $\mathbf{\Lambda}$ shares the same eigenvectors as $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{M}}$.

Proof See appendix A.3. ■

By substituting the factorization (A.7.8) of $\boldsymbol{\Psi}(z)$ into (A.7.5), the following expectational difference equation is obtained:

$$\mathbb{E}_t\left[(\mathbb{I} - \beta\mathbf{\Lambda}\mathbb{F})\mathbf{\Lambda}^{-1}(\mathbb{I} - \mathbf{\Lambda}\mathbb{L})\mathbf{P}_t\right] = \mathbf{K}(\eta_y\boldsymbol{\iota}\mathrm{y}_t + \eta_z\mathbf{z}_t) \tag{A.7.10}$$

Because Lemma 2 demonstrates that matrix $\mathbf{\Lambda}$ has no eigenvalues outside the unit circle and since $0 < \beta < 1$, the matrix $\beta\mathbf{\Lambda}$ has only eigenvalues strictly inside the unit circle. Thus the inverse of $(\mathbb{I} - \beta\mathbf{\Lambda}\mathbb{F})$ has the following convergent Taylor series expansion:

$$(\mathbb{I} - \beta\mathbf{\Lambda}\mathbb{F})^{-1} = \sum_{j=0}^{\infty}\beta^j\mathbf{\Lambda}^j\mathbb{F}^j \tag{A.7.11}$$

Hence, multiplication of both sides of (A.7.10) by $(\mathbb{I} - \beta\mathbf{\Lambda}\mathbb{F})^{-1}\mathbf{\Lambda}$ yields the following expression, which is equivalent to the set of pricing equations in (3.1.5):

$$\mathbf{P}_t = \mathbf{\Lambda}\mathbf{P}_{t-1} + \mathbf{\Lambda}\sum_{j=0}^{\infty}(\beta\mathbf{\Lambda})^j\mathbf{K}\mathbb{E}_t[\eta_y\boldsymbol{\iota}\mathrm{y}_{t+j} + \eta_z\mathbf{z}_{t+j}] \tag{A.7.12}$$

Next, note that because $\boldsymbol{\mathcal{S}}$ and $\mathbf{\Lambda}$ are simultaneously diagonalizable (sharing the same matrix of eigenvectors $\mathbf{V}$), the results of Lemmas 1 and 2 imply that $\boldsymbol{\iota}$ is an eigenvector of $\mathbf{\Lambda}$ corresponding to the eigenvalue of unity. Finally, observe that the price level vector $\mathbf{P}_t$ can be decomposed into a relative price vector $\boldsymbol{\rho}_t$ and an overall price level component as $\mathbf{P}_t = \boldsymbol{\rho}_t + \boldsymbol{\iota}\mathrm{P}_t$. It follows that $(\mathbf{I} - \mathbf{\Lambda})\mathbf{P}_{t-1} = (\mathbf{I} - \mathbf{\Lambda})\boldsymbol{\rho}_{t-1}$, and therefore

$$\mathbf{P}_t - \mathbf{\Lambda}\mathbf{P}_{t-1} = \boldsymbol{\pi}_t + (\mathbf{I} - \mathbf{\Lambda})\boldsymbol{\rho}_{t-1} \tag{A.7.13}$$

where $\boldsymbol{\pi}_t = \mathbf{P}_t - \mathbf{P}_{t-1}$ has been used. By substituting (A.7.13) into (A.7.12), the set of pricing equations (3.1.6) in terms of inflation rates and relative prices is obtained. This completes the proof.

A.8 Proof of Proposition 3

The aggregate demand component y_t from (3.2.3) is constructed using (3.2.2). Since Lemma 2 shows that $\mathbf{\Lambda}$ and $\boldsymbol{\mathcal{S}}$ are simultaneously diagonalizable, the matrix $\mathbf{V}$ of eigenvectors of $\boldsymbol{\mathcal{S}}$ can also be used to diagonalize powers of $\mathbf{\Lambda}$, and so $\mathbf{\Lambda}^j = \mathbf{V}\mathbf{D}^{\Lambda j}\mathbf{V}^{-1}$. By substituting this into the definition of y_t from (3.2.2):

$$\mathrm{y}_t = \eta_y\boldsymbol{\omega}'\mathbf{V}\sum_{j=0}^{\infty}\beta^j\mathbf{D}^{\Lambda j+1}\mathbf{V}^{-1}\boldsymbol{\kappa}\mathbb{E}_t\mathrm{y}_{t+j} \tag{A.8.1}$$

Using the result of Lemma 1 that $\boldsymbol{\omega}'\mathbf{V} = \boldsymbol{\iota}'$ and the definition $\boldsymbol{\varkappa} \equiv \mathbf{V}^{-1}\boldsymbol{\kappa}$, equation (A.8.1) becomes:

$$\mathrm{y}_t = \eta_y\boldsymbol{\iota}'\sum_{j=0}^{\infty}\beta^j\mathbf{D}^{\Lambda j+1}\boldsymbol{\varkappa}\mathbb{E}_t\mathrm{y}_{t+j} \tag{A.8.2}$$

Since $\mathbf{D}^{\Lambda}$ is a diagonal matrix, equation (A.8.2) can be written explicitly as follows:

$$\mathsf{y}_t = \eta_y \sum_{j=0}^{\infty} \beta^j \left(\sum_{i=1}^{n} \varkappa_i \zeta_i^{\Lambda j+1} \right) \mathbb{E}_t \mathsf{y}_{t+j} \tag{A.8.3}$$

By comparing this with (3.2.3), it is clear that μ_j is given by:

$$\mu_j = \eta_y \beta^j \sum_{i=1}^{n} \varkappa_i \zeta_i^{\Lambda j+1} \tag{A.8.4}$$

To establish the sign of this expression, the following result is needed:

Lemma 3 *The system of equations* $\mathbf{V}\boldsymbol{\varkappa} = \boldsymbol{\kappa}$ *has a unique solution, and each element of the* $n \times 1$ *solution vector* $\boldsymbol{\varkappa}$ *is strictly positive.*

Proof See appendix A.4. ■

Together, equation (A.8.4), the inequalities in (A.7.9), the result of Lemma 3, and the fact that $0 < \beta < 1$ and $\eta_y > 0$ imply that $\mu_j > 0$ for all j. To establish the claim about the rate of decay of the sequence $\{\mu_j\}_{j=0}^{\infty}$, note that (A.8.4), (A.7.9) and Lemma 3 imply:

$$\beta - \frac{\mu_{j+1}}{\mu_j} = \beta \left(\frac{\sum_{i=2}^{n} \varkappa_i (1 - \zeta_i^{\Lambda}) \zeta_i^{\Lambda j+1}}{\sum_{i=1}^{n} \varkappa_i \zeta_i^{\Lambda j+1}} \right) > 0 \tag{A.8.5}$$

This implies that $0 < \mu_{j+1} < \beta \mu_j$ for all $j \geq 0$, completing the proof of the proposition.

A.9 Proof of Proposition 4

Suppose that there are only aggregate cost-push shocks, that is, $\mathbf{z}_t = \boldsymbol{\iota} \mathsf{z}_t$, and that the overall set of determinants of inflation $\mathsf{x}_t \equiv \eta_y \mathsf{y}_t + \eta_z \mathsf{z}_t$ follows a stationary AR(1) process, so $\mathsf{x}_t = \mathfrak{a} \mathsf{x}_{t-1} + \upsilon_t$. The coefficient $\mathfrak{a}$ satisfies $|\mathfrak{a}| < 1$ and $\{\upsilon_t\}$ is a white noise shock, $\upsilon_t \sim \mathcal{IID}(0, \sigma_{\upsilon}^2)$.

The following result is useful in proving this proposition:

Lemma 4 *Suppose* $\mathbf{x}_t$ *is the* $n \times 1$ *vector defined by* $\mathbf{x}_t \equiv \eta_y \boldsymbol{\iota} \mathsf{y}_t + \eta_z \mathbf{z}_t$. *If all elements of the vector* $\mathbf{x}_t$ *are identical and equal to* x_t *so that* $\mathbf{x}_t = \boldsymbol{\iota} \mathsf{x}_t$ *then the following expressions for the vectors of price levels* $\mathbf{P}_t$ *and inflation rates* $\boldsymbol{\pi}_t$ *can be obtained,*

$$\mathbf{P}_t = \mathbf{V} \sum_{j=0}^{\infty} \mathbf{D}^{\Lambda j+1} \sum_{k=0}^{\infty} (\beta \mathbf{D}^{\Lambda})^k \boldsymbol{\varkappa} \mathbb{E}_{t-j} \mathsf{x}_{t-j+k} \tag{A.9.1a}$$

$$\boldsymbol{\pi}_t = \mathbf{V}\mathbf{D}^{\Lambda} \sum_{k=0}^{\infty} (\beta \mathbf{D}^{\Lambda})^k \boldsymbol{\varkappa} \mathbb{E}_t \mathsf{x}_{t+k} - \mathbf{V} \sum_{j=1}^{\infty} (\mathbf{I} - \mathbf{D}^{\Lambda}) \mathbf{D}^{\Lambda j} \sum_{k=0}^{\infty} (\beta \mathbf{D}^{\Lambda})^k \boldsymbol{\varkappa} \mathbb{E}_{t-j} \mathsf{x}_{t-j+k} \tag{A.9.1b}$$

where the $n \times 1$ *vector* $\boldsymbol{\varkappa} \equiv \mathbf{V}^{-1} \boldsymbol{\kappa}$ *has been defined. Similarly, expressions for the aggregate price level* P_t *and inflation rate* π_t *can be deduced,*

$$\mathsf{P}_t = \sum_{i=1}^{n} \varkappa_i \sum_{j=0}^{\infty} \zeta_i^{\Lambda j+1} \sum_{k=0}^{\infty} (\beta \zeta_i^{\Lambda})^k \mathbb{E}_{t-j} \mathsf{x}_{t-j+k} \tag{A.9.2a}$$

$$\pi_t = \sum_{i=1}^{n} \varkappa_i \zeta_i^{\Lambda} \sum_{k=0}^{\infty} (\beta \zeta_i^{\Lambda})^k \mathbb{E}_t \mathsf{x}_{t+k} - \sum_{i=2}^{n} \varkappa_i \sum_{j=1}^{\infty} (1 - \zeta_i^{\Lambda}) \zeta_i^{\Lambda j} \sum_{k=0}^{\infty} (\beta \zeta_i^{\Lambda})^k \mathbb{E}_{t-j} \mathsf{x}_{t-j+k} \tag{A.9.2b}$$

where $\varkappa_i$ *is the* i*-th element of the vector* $\boldsymbol{\varkappa}$.

Proof See appendix A.5. ■

Since $\{\mathsf{x}_t\}$ is an AR(1) process, $\mathbb{E}_t\mathsf{x}_{t+k} = \mathfrak{a}^k\mathsf{x}_t$, and hence:

$$\sum_{k=0}^{\infty}(\beta\zeta_i^\Lambda)^k\mathbb{E}_t\mathsf{x}_{t+k} = (1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}\mathsf{x}_t \tag{A.9.3}$$

Note that since $0 < \zeta_i^\Lambda \leq 1$, $0 < \beta < 1$ and $|\mathfrak{a}| < 1$, the term $(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}$ is strictly positive. Using (A.9.3), equation (A.9.2b) implies that the stochastic process for economy-wide inflation $\{\pi_t\}$ is given by:

$$\pi_t = \sum_{i=1}^{n}\varkappa_i\zeta_i^\Lambda(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}\mathsf{x}_t - \sum_{i=2}^{n}\varkappa_i\sum_{j=1}^{\infty}(1-\zeta_i^\Lambda){\zeta_i^\Lambda}^j(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}\mathsf{x}_{t-j} \tag{A.9.4}$$

As $\boldsymbol{\rho}_{t-1}$ is a relative price vector, it must be the case that $\boldsymbol{\omega}'\boldsymbol{\rho}_{t-1} = 0$, and so the definition of the current level of intrinsic inertia ш_t in (3.2.1) is equivalent to $\text{ш}_t = -\boldsymbol{\omega}'(\mathbf{I}-\boldsymbol{\Lambda})\boldsymbol{\rho}_{t-1}$. Since Lemmas 1 and 2 show that $\boldsymbol{\iota}$ is an eigenvector of $\boldsymbol{\Lambda}$ with a corresponding eigenvalue of one, it follows that $(\mathbf{I}-\boldsymbol{\Lambda})\boldsymbol{\rho}_{t-1} = (\mathbf{I}-\boldsymbol{\Lambda})\mathbf{P}_{t-1}$. And because $\mathbf{V}$ diagonalizes $\boldsymbol{\Lambda}$, the matrix $\mathbf{I}-\boldsymbol{\Lambda}$ can be written as $(\mathbf{I}-\boldsymbol{\Lambda}) = \mathbf{V}(\mathbf{I}-\mathbf{D}^\Lambda)\mathbf{V}^{-1}$. Together with the normalization of the eigenvectors $\boldsymbol{\omega}'\mathbf{V} = \boldsymbol{\iota}'$, the definition $\boldsymbol{\varkappa} \equiv \mathbf{V}^{-1}\boldsymbol{\kappa}$, and equation (A.9.1a), the current inertial component of inflation ш_t is equal to:

$$\text{ш}_t = -\sum_{i=1}^{n}\varkappa_i(1-\zeta_i^\Lambda)\sum_{j=0}^{\infty}{\zeta_i^\Lambda}^{j+1}\sum_{k=0}^{\infty}(\beta\zeta_i^\Lambda)^k\mathbb{E}_{t-1-j}\mathsf{x}_{t-1-j+k} \tag{A.9.5}$$

By substituting the expression for the sum from (A.9.3) into (A.9.5) and noting that $\zeta_1^\Lambda = 1$, intrinsic inertia ш_t can be written as:

$$\text{ш}_t = \sum_{j=1}^{\infty}\left\{-\sum_{i=2}^{n}\varkappa_i(1-\zeta_i^\Lambda){\zeta_i^\Lambda}^j(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}\right\}\mathsf{x}_{t-j} \tag{A.9.6}$$

The expression for inflation from (A.9.4) in terms of the history of $\{\mathsf{x}_t\}$ is then substituted into the alternative definition of intrinsic inertia from (3.2.8) to give:

$$\text{ш}_t = \sum_{j=1}^{\infty}\left\{\gamma_j\left(\sum_{i=1}^{n}\varkappa_i\zeta_i^\Lambda(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}\right) - \sum_{k=1}^{j-1}\gamma_{j-k}\left(\sum_{i=2}^{n}\varkappa_i(1-\zeta_i^\Lambda){\zeta_i^\Lambda}^k(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}\right)\right\}\mathsf{x}_{t-j} \tag{A.9.7}$$

The two expressions for ш_t in (A.9.6) and (A.9.7) are equivalent when all the coefficients of the history $\{\mathsf{x}_{t-1}, \mathsf{x}_{t-2}, \ldots\}$ are the same in both equations. A recursive formula for the sequence $\{\gamma_j\}_{j=1}^{\infty}$ is obtained by equating coefficients:

$$\gamma_j = -\frac{\sum_{i=2}^{n}\varkappa_i(1-\zeta_i^\Lambda){\zeta_i^\Lambda}^j(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}}{\sum_{i=1}^{n}\varkappa_i\zeta_i^\Lambda(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}} + \sum_{k=1}^{j-1}\left(\frac{\sum_{i=2}^{n}\varkappa_i(1-\zeta_i^\Lambda){\zeta_i^\Lambda}^k(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}}{\sum_{i=1}^{n}\varkappa_i\zeta_i^\Lambda(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1}}\right)\gamma_{j-k} \tag{A.9.8}$$

Since $(1-\beta\mathfrak{a}\zeta_i^\Lambda)^{-1} > 0$, $\varkappa_i > 0$ and $\zeta_i^\Lambda > 0$ for all i, and $\zeta_i^\Lambda < 1$ for $i \geq 2$ are obtained from the inequalities in (A.7.9) and Lemma 3, the value of γ_1 is negative, and by induction, so are all the other coefficients γ_j for $j \geq 1$. This completes the proof of the proposition.

A.10 Proof of Theorem 1

The intrinsic impulse response function for inflation can be obtained as the usual impulse response function under the assumption of a common white-noise cost-push shock for all industries if the shock and any resulting intrinsic inflation inertia are fully accommodated, so that there are no output gap fluctuations. Formally, this means that $\mathsf{y}_t = 0$ and $\mathbf{z}_t = \boldsymbol{\iota}\nu_t$, where $\nu_t \sim \mathcal{IID}(0, \sigma_\nu^2)$ is a white-noise shock. The intrinsic impulse response function $\{\text{п}(j)\}_{j=0}^{\infty}$ can then be obtained from the coefficients of the MA(∞) representation of inflation π_t in terms of the shock ν_t,

$$\pi_t = \text{л} \sum_{j=0}^{\infty} \text{п}(j)\nu_{t-j} \tag{A.10.1}$$

where the multiplicative factor л is introduced because the intrinsic impulse response function is normalized so that $\text{п}(0) = 1$. The MA(∞) representation of inflation can be obtained by making use of the results in Lemma 4.

The complete accommodation of the white-noise cost-push shock and of any resulting intrinsic inflation inertia that characterizes the intrinsic impulse response function formally requires that $\pi_t = \text{ш}_t + \mathfrak{z}_t$ at all times. From (3.2.4) this is clearly equivalent to $\mathsf{y}_t = 0$ in all time periods, which in turn by using equation (3.2.3) means that y_t is always zero. Thus if $\mathbf{x}_t \equiv \eta_y \boldsymbol{\iota} \mathsf{y}_t + \eta_z \mathbf{z}_t$ then $\mathbf{x}_t = \eta_z \boldsymbol{\iota}\nu_t$, and so Lemma 4 can be applied with $\mathsf{x}_t = \eta_z \nu_t$. Since $\{\nu_t\}$ is a white noise process, $\mathbb{E}_t \mathsf{x}_{t+k} = 0$ for all $k \geq 1$. Hence equation (A.9.2b) implies the following MA(∞) representation for inflation:

$$\pi_t = \left(\eta_z \sum_{i=1}^{n} \varkappa_i \zeta_i^{\Lambda}\right)\nu_t - \sum_{j=1}^{\infty}\left(\eta_z \sum_{i=2}^{n} \varkappa_i (1-\zeta_j^{\Lambda})\zeta_i^{\Lambda j}\right)\nu_{t-j} \tag{A.10.2}$$

By comparing the above with equation (A.10.1) and equating coefficients of ν_t, the intrinsic impulse response function is given by:

$$\text{п}(j) = \begin{cases} 1 & \text{if} \quad j = 0 \\ -\frac{\sum_{i=2}^{n} \varkappa_i (1-\zeta_i^{\Lambda})\zeta_i^{\Lambda j}}{\sum_{i=1}^{n} \varkappa_i \zeta_i^{\Lambda}} & \text{if} \quad j = 1, 2, \ldots \end{cases} \tag{A.10.3}$$

The multiplicative constant in (A.10.1) is set to $\text{л} = \eta_z \sum_{i=1}^{n} \varkappa_i \zeta_i^{\Lambda}$ because the normalization $\text{п}(0) = 1$ has been adopted. Since all the $\varkappa_i$ are strictly positive according to Lemma 3, and as (A.7.9) shows that $\zeta_i^{\Lambda} > 0$ for all i and $\zeta_i^{\Lambda} < 1$ for $i \geq 2$, the fact that $\text{п}(j) < 0$ for all $j \geq 1$ can be deduced from the expression for the intrinsic impulse response function in (A.10.3). This completes the proof of the theorem.

A.11 Proof of Theorem 2

If the expression for the hypothetical New Keynesian Phillips curve in (3.3.2) with discount factor β and short-run slope κ is iterated forwards, then the following equation for the inflation rate $\Pi_t(\beta, \kappa)$ is obtained:

$$\Pi_t(\beta, \kappa) = \kappa \sum_{j=0}^{\infty} \beta^j \mathbb{E}_t \mathsf{x}_{t+j} \tag{A.11.1}$$

By substituting equation (A.11.1) into the result (A.9.2b) from Lemma 4, the actual inflation rate can be written in terms of the current and past inflation rates generated by n hypothetical New Keynesian Phillips curves:

$$\pi_t = \Pi_t(\beta, \varkappa_1) + \sum_{i=2}^{n} \zeta_i^{\Lambda}\left(\Pi_t(\beta\zeta_i^{\Lambda}, \varkappa_i) - (1-\zeta_i^{\Lambda})\sum_{j=1}^{\infty} \zeta_i^{\Lambda j-1} \Pi_{t-j}(\beta\zeta_i^{\Lambda}, \varkappa_i)\right) \tag{A.11.2}$$

To verify the claim in (3.3.3), the discount factors used in the hypothetical NKPCs are set to $\tilde{\beta}_i \equiv \beta\zeta_i^{\Lambda}$, and the slopes to $\tilde{\kappa}_i \equiv \varkappa_i$. The results from (A.7.9) and Lemma 3 ensure that the inequalities $0 < \tilde{\beta}_i \leq \beta < 1$ and $0 < \tilde{\kappa}_i < \infty$ are satisfied. The coefficients $\mathfrak{c}_{ij}$ from (3.3.3) are then given by $\mathfrak{c}_{i0} \equiv \zeta_i^{\Lambda} > 0$, and

$\mathfrak{c}_{ij} \equiv (1-\zeta_i^\Lambda)\zeta_i^{\Lambda j-1} > 0$ for $j \geq 1$ and $i \geq 2$. This completes the proof.

A.12 Proof of Theorem 3

All the results of this theorem are derived under the assumption that the aggregate forcing variable x_t follows a stationary AR(1) process, as given in equation (3.3.4), with non-negative serial correlation ($0 \leq \mathfrak{a} < 1$). By iterating (3.3.4) backwards, x_t is expressed as a sum of current and past white-noise shocks υ_t:

$$\mathsf{x}_t = \sum_{l=0}^{\infty} \mathfrak{a}^l \upsilon_{t-l} \tag{A.12.1}$$

Thus the impulse response function of x_t to a shock υ_t is the geometric series $\{\mathfrak{a}^j\}_{j=0}^{\infty}$. The corresponding impulse response function $\mathscr{J}(j)$ of inflation in the case of homogeneity is simply proportional to this. As $\mathscr{J}(j)$ is normalized so that $\mathscr{J}(0) = 1$, it follows from (A.12.1) that $\mathscr{J}(j) = \mathfrak{a}^j$. The analysis below derives the corresponding impulse response function $\mathscr{I}(j)$ of inflation with heterogeneous price stickiness for the same stochastic process (3.3.4) of the cost-push shock.

Equation (3.3.4) implies that the conditional expectation of future x_t is given by $\mathbb{E}_t\mathsf{x}_{t+k} = \mathfrak{a}^k\mathsf{x}_t$ for all $k \geq 0$. This formula for the conditional expectation can be used together with (A.12.1) and equation (A.9.2a) from Lemma 4 to obtain an expression for the aggregate price level P_t in terms of the history of white noise shocks $\{\upsilon_t, \upsilon_{t-1}, \ldots\}$:

$$\mathsf{P}_t = \sum_{i=1}^{n} \varkappa_i \sum_{j=0}^{\infty} \zeta_i^{\Lambda j+1} \sum_{k=0}^{\infty} (\beta\zeta_i^\Lambda)^k \sum_{l=0}^{\infty} \mathfrak{a}^{k+l} \upsilon_{t-j-l} \tag{A.12.2}$$

By changing the order of summation in the above, the following alternative formula for P_t is found:

$$\mathsf{P}_t = \sum_{i=1}^{n} \varkappa_i \left(\sum_{k=0}^{\infty} (\beta\mathfrak{a}\zeta_i^\Lambda)^k \right) \sum_{j=0}^{\infty} \left(\zeta_i^{\Lambda j+1} \sum_{l=0}^{j} (\mathfrak{a}/\zeta_i^\Lambda)^l \right) \upsilon_{t-j} \tag{A.12.3}$$

The geometric sums appearing in (A.12.3) can be eliminated from the expression for the price level as follows:

$$\mathsf{P}_t = \sum_{j=0}^{\infty} \left(\sum_{i=1}^{n} \frac{\varkappa_i\zeta_i^\Lambda}{1-\beta\mathfrak{a}\zeta_i^\Lambda} \left(\frac{\zeta_i^{\Lambda j+1} - \mathfrak{a}^{j+1}}{\zeta_i^\Lambda - \mathfrak{a}} \right) \right) \upsilon_{t-j} \tag{A.12.4}$$

The stochastic process for economy-wide inflation $\pi_t = \mathsf{P}_t - \mathsf{P}_{t-1}$ is then obtained by first-differencing (A.12.4):

$$\pi_t = \sum_{j=0}^{\infty} \left(\sum_{i=1}^{n} \frac{\varkappa_i\zeta_i^\Lambda}{1-\beta\mathfrak{a}\zeta_i^\Lambda} \left(\frac{(1-\mathfrak{a})\mathfrak{a}^j - (1-\zeta_i^\Lambda)\zeta_i^{\Lambda j}}{\zeta_i^\Lambda - \mathfrak{a}} \right) \right) \upsilon_{t-j} \tag{A.12.5}$$

In equation (3.3.5), the MA(∞) representation of inflation is denoted by $\pi_t = \text{л} \sum_{j=0}^{\infty} \mathscr{I}(j)\upsilon_{t-j}$, where the positive constant л is introduced to ensure that $\mathscr{I}(0) = 1$. By comparing this with (A.12.5), the coefficient $\mathscr{I}(j)$ and the constant л are given by:

$$\mathscr{I}(j) = \frac{1}{\text{л}} \sum_{i=1}^{n} \frac{\varkappa_i\zeta_i^\Lambda}{1-\beta\mathfrak{a}\zeta_i^\Lambda} \left(\frac{(1-\mathfrak{a})\mathfrak{a}^j - (1-\zeta_i^\Lambda)\zeta_i^{\Lambda j}}{\zeta_i^\Lambda - \mathfrak{a}} \right) \quad , \quad \text{л} \equiv \sum_{i=1}^{n} \frac{\varkappa_i\zeta_i^\Lambda}{1-\beta\mathfrak{a}\zeta_i^\Lambda} \tag{A.12.6}$$

The first claim to prove is that $\mathscr{I}(j) < \mathscr{J}(j)$ for all $j \geq 1$. Since $\mathscr{J}(j) = \mathfrak{a}^j$ and $\zeta_1^\Lambda = 1$, the formula for $\mathscr{I}(j)$ in equation (A.12.6) can be used to show that this inequality holds if and only if:

$$\sum_{i=2}^{n} \frac{\varkappa_i \zeta_i^{\Lambda}(1-\zeta_i^{\Lambda})}{1-\beta\mathfrak{a}\zeta_i^{\Lambda}} \left(\frac{\zeta_i^{\Lambda j} - \mathfrak{a}^j}{\zeta_i^{\Lambda} - \mathfrak{a}} \right) > 0 \tag{A.12.7}$$

This expression is indeed positive for all $j \geq 1$ because $0 < \beta < 1$ and $0 \leq \mathfrak{a} < 1$ hold by assumption, the inequalities in (A.7.9) demonstrate that $0 < \zeta_i^{\Lambda} < 1$ for all $i \geq 2$, and Lemma 3 shows that $\varkappa_i > 0$ for all i. Finally, $\zeta_i^{\Lambda j} - \mathfrak{a}^j$ and $\zeta_i^{\Lambda} - \mathfrak{a}$ must always have the same sign for $j \geq 1$ because both ζ_i^{Λ} and $\mathfrak{a}$ are non-negative and less than one for $i \geq 2$. This establishes that $\mathscr{I}(j)$ decays more rapidly than $\mathscr{J}(j)$.

The next part of the theorem concerns the shape of the impulse response function $\mathscr{I}(j)$ with heterogeneity. Define the following function $\mathfrak{f}(\tau;\zeta,\mathfrak{a})$ of continuous time $\tau \geq 0$ with parameters $0 < \zeta \leq 1$ and $0 \leq \mathfrak{a} < 1$:

$$\mathfrak{f}(\tau;\zeta,\mathfrak{a}) \equiv \frac{(1-\mathfrak{a})\mathfrak{a}^{\tau} - (1-\zeta)\zeta^{\tau}}{\zeta - \mathfrak{a}} \tag{A.12.8}$$

The coefficient $\mathscr{I}(j)$ from (A.12.6) can be written as a sum of terms involving $\mathfrak{f}(j;\zeta_i^{\Lambda},\mathfrak{a})$ for each eigenvalue ζ_i^{Λ} of $\mathbf{\Lambda}$,

$$\mathscr{I}(\tau) = \frac{1}{\text{л}} \sum_{i=1}^{n} \frac{\varkappa_i \zeta_i^{\Lambda}}{1-\beta\mathfrak{a}\zeta_i^{\Lambda}} \mathfrak{f}(\tau;\zeta_i^{\Lambda},\mathfrak{a}) \tag{A.12.9}$$

where $\mathscr{I}(\tau)$ is treated as a function of continuous time for convenience, even though the results only involve $\mathscr{I}(\tau)$ evaluated at a discrete set of points. Note that the inequalities $0 < \beta < 1$, $0 \leq \mathfrak{a} < 1$, together with those in (A.7.9) and Lemma 3 imply that the coefficients of the functions $\mathfrak{f}(\tau;\zeta_i^{\Lambda},\mathfrak{a})$ in (A.12.9) are strictly positive.

By repeatedly differentiating the function $\mathfrak{f}(\tau;\zeta,\mathfrak{a})$ in (A.12.8) with respect to time τ, the following expression is found for the k-th order derivative, denoted by $\mathfrak{f}^{(k)}(\tau;\zeta,\mathfrak{a})$:

$$\mathfrak{f}^{(k)}(\tau;\zeta,\mathfrak{a}) = (-1)^k \frac{(\log \mathfrak{a}^{-1})^k (1-\mathfrak{a})\mathfrak{a}^{\tau} - (\log \zeta^{-1})^k (1-\zeta)\zeta^{\tau}}{\zeta - \mathfrak{a}} \tag{A.12.10}$$

Note that $\mathfrak{f}^{(k)}(\tau;\zeta,\mathfrak{a})$ and all its derivatives are continuous functions of time τ. It can be seen from (A.12.10) that $(-1)^k \mathfrak{f}^{(k)}(0) > 0$ for all k, and $\lim_{\tau\to\infty} \mathfrak{f}^{(k)}(\tau) = 0$, given the parameter restrictions $0 < \zeta \leq 1$ and $0 \leq \mathfrak{a} < 1$. Equation (A.12.9) implies that the time derivatives of $\mathscr{I}(\tau)$ can be obtained from those of $\mathfrak{f}(\tau;\zeta,\mathfrak{a})$, again with a sum involving the derivatives evaluated at all n eigenvalues ζ_i^{Λ}:

$$\mathscr{I}^{(k)}(\tau) = \frac{1}{\text{л}} \sum_{i=1}^{n} \frac{\varkappa_i \zeta_i^{\Lambda}}{1-\beta\mathfrak{a}\zeta_i^{\Lambda}} \mathfrak{f}^{(k)}(\tau;\zeta_i^{\Lambda},\mathfrak{a}) \tag{A.12.11}$$

Thus all the derivatives of $\mathscr{I}(\tau)$ inherit continuity from $\mathfrak{f}(\tau;\zeta,\mathfrak{a})$. And using the equivalent results for $\mathfrak{f}^{(k)}(\tau;\zeta,\mathfrak{a})$ derived above, equation (A.12.11) implies that $(-1)^k \mathscr{I}^{(k)}(0) > 0$ for all k and $\lim_{\tau\to\infty} \mathscr{I}^{(k)}(\tau) = 0$. By substituting (A.12.10) into (A.12.11) and rearranging:

$$\mathscr{I}^{(k)}(\tau) = (-1)^k (\log \mathfrak{a}^{-1})^k \frac{1}{\text{л}} \sum_{i=1}^{n} \frac{\varkappa_i \zeta_i^{\Lambda}}{1-\beta\mathfrak{a}\zeta_i^{\Lambda}} \left(\frac{(1-\mathfrak{a})\mathfrak{a}^{\tau} - \left(\frac{\log \zeta_i^{\Lambda^{-1}}}{\log \mathfrak{a}^{-1}} \right)^k (1-\zeta^{\Lambda})\zeta_i^{\Lambda\tau}}{\zeta_i^{\Lambda} - \mathfrak{a}} \right) \tag{A.12.12}$$

This expression for $\mathscr{I}^{(k)}(\tau)$ can be used to deduce the following inequalities involving the k-th and $(k+1)$-th order derivatives of $\mathscr{I}(\tau)$,

$$\frac{\mathscr{I}^{(k+1)}(\tau)}{(\log \mathfrak{a}^{-1})^{k+1}} \begin{cases} < -\frac{\mathscr{I}^{(k)}(\tau)}{(\log \mathfrak{a}^{-1})^k} & \text{(k even)} \\ > -\frac{\mathscr{I}^{(k)}(\tau)}{(\log \mathfrak{a}^{-1})^k} & \text{(k odd)} \end{cases} \tag{A.12.13}$$

where the direction of the inequality depends on whether k is odd or even.

Since $\mathscr{I}(0)$ is known to be positive, there are two mutually exclusive and exhaustive possibilities. First, that $\mathscr{I}(\tau)$ remains strictly positive for all $\tau \geq 0$. Second, that there is at least one point in finite time at which $\mathscr{I}(\tau)$ is non-positive.

First consider the case where $\mathscr{I}(\tau)$ is everywhere positive. The inequality in (A.12.13) then implies that $\mathscr{I}'(\tau)$ is negative for all τ, which in turn implies $\mathscr{I}''(\tau)$ is positive everywhere, and so on. So in this case, $(-1)^k \mathscr{I}^{(k)}(\tau) > 0$ for all $\tau \geq 0$ and all k. Thus all even-order derivatives of $\mathscr{I}(\tau)$ are positive everywhere, and all odd orders are negative everywhere. This means that $\mathscr{I}(\tau)$ is everywhere positive and decreasing, which corresponds to "case (ii)" in the statement of the theorem.

Now consider the case where $\mathscr{I}(\tau)$ is non-positive somewhere. Since $\mathscr{I}(\tau)$ is a continuous function, there must exist a smallest $\tau_0 > 0$ where the function is first equal to zero. It can then be deduced that $\mathscr{I}(\tau)$ must be negative in a neighbourhood to the right of τ_0, because inequality (A.12.13) implies that were the function not to become negative immediately after passing τ_0, then it would necessarily be decreasing in this range, which is not possible since it has already reached zero at τ_0.

Observe that once $\mathscr{I}(\tau)$ has become negative after τ_0, it cannot become positive again for larger values of τ. Were this to happen, because $\mathscr{I}(\tau)$ is a continuous function there would have to be a point where $\mathscr{I}(\tau)$ cuts the horizontal axis from below. However, the inequality (A.12.13) shows that as soon as the function becomes positive, it would immediately become decreasing, which is not possible for a continuously differentiable function. Thus $\mathscr{I}(\tau)$ cutting the horizontal axis from below can be ruled out, and hence $\mathscr{I}(\tau)$ must remain negative for all $\tau > \tau_0$. Finally, because $\mathscr{I}(\tau)$ is negative after τ_0, the fact that it is a continuously differentiable function which tends to zero as $\tau \to \infty$ means there must exist a first turning point $\tau_1 > \tau_0$ where $\mathscr{I}'(\tau_1) = 0$. Hence, while $\mathscr{I}'(\tau)$ is initially negative, it must become positive at some point.

These arguments are now generalized to apply to all the derivatives of $\mathscr{I}(\tau)$. Start with the k-th derivative $\mathscr{I}^{(k)}(\tau)$, where k is odd [even]. This derivative is known to be initially negative [positive], but suppose that it becomes positive [negative] for the first time immediately after point $\tau_k > 0$. Using a version of the earlier argument, the inequalities in (A.12.13) imply that $\mathscr{I}^{(k)}(\tau)$ is increasing [decreasing] in a neighbourhood to the right of τ_k, and must remain positive [negative] for all $\tau > \tau_k$. Because the k-th derivative tends to zero as $\tau \to \infty$, and since it is a continuous function, there must exist a first point $\tau_{k+1} > \tau_k$ where $\mathscr{I}^{(k+1)}(\tau_{k+1}) = 0$. Thus the $(k+1)$-th derivative of $\mathscr{I}(\tau)$ starts positive [negative], but becomes negative [positive] for the first time after τ_{k+1}.

This argument can be applied inductively to deduce that there exists a sequence of points $0 < \tau_0 < \tau_1 < \tau_2 < \cdots < \infty$ such that $(-1)^k \mathscr{I}^{(k)}(\tau) > 0$ if and only if $\tau < \tau_k$. Hence the function $\mathscr{I}(\tau)$ is positive and decreasing before τ_0, negative and decreasing between τ_0 and τ_1, and negative and increasing after τ_1. This corresponds to "case (i)" in the statement of the theorem, and necessarily occurs whenever "case (ii)" does not.

Finally, note that when $\mathfrak{a} \to 1$, the function $\mathfrak{f}(\tau; \zeta, \mathfrak{a})$ in (A.12.8) becomes:

$$\lim_{\mathfrak{a} \to 1} \mathfrak{f}(\tau; \zeta, \mathfrak{a}) = \zeta^\tau \tag{A.12.14}$$

This is positive for all τ, and since (A.12.9) shows that $\mathscr{I}(\tau)$ is a linear combination of the functions $\mathfrak{f}(\tau; \zeta_i^\Lambda, \mathfrak{a})$ with positive coefficients, $\mathscr{I}(\tau)$ must also be positive everywhere in this limiting case. Thus for $\mathfrak{a}$ sufficiently close to 1, the impulse response function $\mathscr{I}(j)$ is always in "case (ii)". When $\mathfrak{a} \to 0$, the extrinsic persistence in the shock disappears, and the actual impulse response function $\mathscr{I}(j)$ tends to the intrinsic impulse response function $\pi(j)$. But the properties of $\pi(j)$ derived in Theorem 1 demonstrate that it falls within "case (i)". Thus all the claims of the theorem are proved.