Sales and Monetary Policy*

Bernardo Guimaraes Kevin D. Sheedy†
London School of Economics

First draft: 8th December 2007
This draft: 16th September 2008

Abstract

A striking fact about prices is the prevalence of "sales": large temporary price cuts followed by prices returning exactly to their former levels. This paper builds a macroeconomic model with a rationale for sales based on firms facing consumers with different price sensitivities. Even if firms can adjust sales without cost, monetary policy has large real effects owing to sales being strategic substitutes: a firm's incentive to have a sale is decreasing in the number of other firms having sales. Thus the flexibility seen in individual prices due to sales does not translate into flexibility of the aggregate price level.

JEL classifications: E3; E5.

Keywords: sales; monetary policy; nominal rigidities.

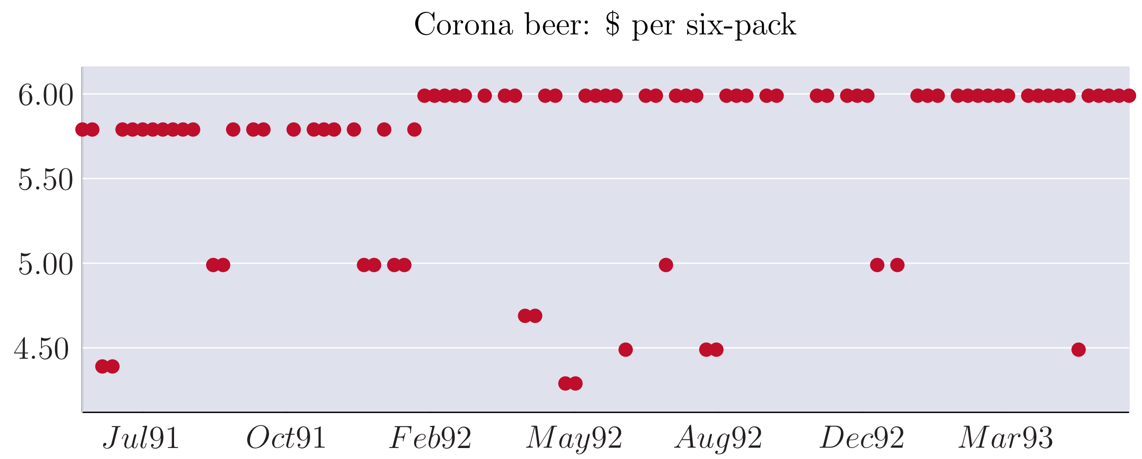
*We thank Andy Levin, Ananth Ramanarayanan, and especially Rachel Ngai for helpful discussions and suggestions, and seminar participants at Bristol University, the European Central Bank, Fundação Getúlio Vargas – Rio, the London School of Economics, the Paris School of Economics, the CEP/ESRC Monetary Policy conference, the 2008 Texas Monetary Conference, and the 2008 Money Macro and Finance Research Group conference for their comments.

†Corresponding author: 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, Website: http://personal.lse.ac.uk/sheedy.

1 Introduction

A striking fact about prices is that many price changes are "sales": large temporary price cuts followed by prices returning exactly to their former levels. Figure 1 shows a typical price path for a six-pack of Corona beer at an outlet of *Dominick's Finer Foods*, a U.S. supermarket. Sales are frequent; other types of price change are rare. This pattern is an archetype of retail pricing.[1]

Figure 1: *Example price path*



Notes: Weekly price observations from *Dominick's Finer Foods*, Oak Lawn, Illinois, U.S.A.
Source: James M. Kilts Center, GSB, University of Chicago (`http://research.chicagogsb.edu/marketing/databases/dominicks`).

Monetary policy's real effects on the economy depend crucially on the stickiness of prices. So Figure 1 poses a conundrum: viewed from different perspectives, the price path exhibits great flexibility on the one hand, but substantial stickiness on the other. While changes between some "normal" price and a temporary "sale" price are frequent, the "normal" price itself changes far less often.[2] Consequently, empirical estimates of price stickiness widely diverge when sales are treated differently. Bils and Klenow (2004) count sales as price changes and find that the median duration of a price spell across the whole consumer price index is around 4 months; by disregarding sales, Nakamura and Steinsson (2007) find a median duration of around 9 months.[3] Quantitative models deliver radically different estimates of the real effects of monetary policy depending on which of these two numbers is used. Hence an understanding of sales is needed to answer the question of how large those real effects should be.

Given the pattern of price adjustment documented in Figure 1, changes in the aggregate price level can come from three sources: changes in the "normal" price, changes in the size of the sale discount, and changes in the proportion of goods on sale. First consider a world where all changes in

[1]See Hosken and Reiffen (2004), Klenow and Kryvtsov (2005), Nakamura and Steinsson (2007), Kehoe and Midrigan (2007), Goldberg and Hellerstein (2007) and Eichenbaum, Jaimovich and Rebelo (2008) for recent studies.

[2]It is harder to make generalizations about sale prices. Some products feature a relatively constant sale discount; others display sizeable variation over time.

[3]Comparisons across euro-area countries also reveal that the treatment of sales has a significant bearing on the measured frequency of price adjustment, as discussed in Dhyne, Álvarez, Le Bihan, Veronese, Dias, Hoffmann, Jonker, Lünnemann, Rumler and Vilmunen (2006).

the aggregate price level are driven by variations in the fraction of goods on sale, and all individual price changes are associated with sales. This could be modelled by assuming firms have a fixed normal price, a fixed sale discount, and then optimally choose the fraction of time their good is on sale. If consumer preferences were standard, with all firms facing a constant price-elasticity demand function, then this paper shows firms' profit-maximizing choice of the frequency of sales would lead to approximate money neutrality. Even if the normal price and sale discount were fixed, optimizing variation in the fraction of goods on sale would mimic the price changes chosen by firms in a world of completely flexible prices.

This simple framework for analysing sales is not complete, though. No reason has yet been presented for why firms would want to choose a pricing strategy in which sales discounts play a significant role. In the IO literature, the most prominent theories of sales are based on incomplete information about prices and consumer preferences. Leading examples include Salop and Stiglitz (1977, 1982), Varian (1980), Sobel (1984) and Lazear (1986). This paper builds a general-equilibrium macroeconomic model with sales that draws upon the rationale proposed in the IO literature. Despite substantial heterogeneity at the microeconomic level, the model is easily aggregated to study macroeconomic questions.

The model presented here assumes consumers have different preferences over goods, and for each good, some consumers are more price sensitive than others. There are two types: loyal consumers with low price elasticities, and bargain hunters with high elasticities. Firms do not know the type of any individual customer, so they cannot practise price discrimination. One key finding of the paper is that when the difference between the price elasticities of loyal consumers and bargain hunters is sufficiently large, and there is a sufficient mixture of the two types, then in the unique equilibrium of the model, firms strictly prefer to sell their good at a high price at some moments and at a low sale price at other moments. The choice of different prices at different moments is a profit-maximizing strategy even in an entirely deterministic environment. Each of the two prices is aimed at a particular type of consumer, in spite of the fact that at a given moment all customers of a firm are actually paying the same price. Firms would like to price discriminate, but as this is impossible, their best alternative strategy is to target the two types of consumer at different moments by holding periodic sales.

The existence of consumers with different price elasticities leads to sales being strategic substitutes, or in other words, the more others have sales, the less any individual firm wants to have a sale. This is because the difficulty faced by a given firm in trying to win the custom of the more price-sensitive consumers is greatly dependent on the extent to which other firms are having sales. In contrast, a firm can rely more on its loyal customers, whose purchases are much less sensitive to other firms' pricing decisions. Thus the relative attractiveness of targeting the bargain hunters decreases when others are targeting them with sales. The resulting market equilibrium features a balance between the fractions of time a firm spends targeting the two groups of consumers, and hence large and frequent price changes occur even in the absence of any shocks.

Having built a model of sales, the key question to be answered is: for the purposes of monetary policy analysis, does it matter that the normal price is sticky amidst all the flexibility due to sales seen in Figure 1? In sharp contrast to the simple framework discussed first where flexibility in sales together with homogeneous and standard consumer preferences implied money was approximately neutral, monetary policy has strong real effects in the IO-based model of sales when the normal price is sticky but sales decisions are completely flexible.

The strong real effects of monetary policy follow from sales being strategic substitutes in the IO-based model. After an expansionary monetary policy shock, an individual firm has a direct incentive to hold fewer and less generous sales, thus increasing the price it sells at on average. However, if all other firms were to follow this course of action then any one firm would have a tempting opportunity to boost its market share among the bargain hunters by holding a sale — bargain hunters are much easier to attract if neglected by others. This leads firms in equilibrium not to adjust sales by much in response to a monetary shock. Although the shock is common to all firms, there are also strong incentives not to follow what others do. Consequently, the aggregate price level adjusts by less and monetary policy shocks have larger real effects.

The model is then calibrated to match some simple facts about sales and hence assess quantitatively the real effects of monetary policy. If the normal price is completely sticky and sales decisions are completely flexible then the elasticity of output with respect to an unanticipated change in the money supply is around 0.9, and the elasticity of the price level is around 0.1. The flexibility due to sales seen at the level of individual prices imparts little flexibility to the aggregate price level. Therefore, the real effects of monetary policy in a model with a sticky normal price and fully flexible sales are very similar to those found in a model with a single sticky price and no sales. These numerical results are not particularly sensitive to the calibration of the model.

This analysis treats the normal price as sticky, an assumption in line with the stylized facts as illustrated in Figure 1. A branch of the macroeconomics literature has proposed a range of justifications for price rigidity, some of which can be applied to explain why the normal price is not continuously readjusted. While these features are not directly incorporated into the model, there are three findings of this paper which add support to the assumption of a sticky normal price when firms simultaneously have the option of adjusting sales. First, the absolute size of any reoptimization costs needed to justify a constant normal price is much lower than in an otherwise comparable model where the normal price is the only price. Second, deviations of normal prices from profit-maximizing levels are small — even though the model features very large individual price changes — so the losses from failing to adjust the normal price are not as great as might be supposed simply by looking at the volatility of individual prices. Third, a firm has more to gain from ensuring its sale price is set optimally than it has from continuously monitoring the optimality of its normal price.

This paper then constructs a fully dynamic version of the model with sales where firms' normal prices are reoptimized at staggered intervals. Individual price paths generated by the model are similar to real-world examples such as that in Figure 1 even though no idiosyncratic shocks are assumed. This dynamic extension is also very tractable and an expression for the resulting Phillips curve is derived analytically. It is embedded into a dynamic stochastic general equilibrium model and the results of simulations are compared to those from the same DSGE model without sales incorporating a standard New Keynesian Phillips curve instead. A quantitative analysis reveals that the difference between the real effects of monetary policy in the two models is small, and in line with the findings of the simpler model with a fixed normal price.

Even though the recent empirical literature on price adjustment has highlighted the importance of sales, macroeconomic models have largely side-stepped the issue. The one exception is Kehoe and Midrigan (2007). In their model, firms face different menu costs depending on whether they make permanent or temporary (downward) price changes. Coupled with large but transitory idiosyncratic shocks, this mechanism gives rise to sales in equilibrium.

The plan of the paper is as follows. Section 2 sets out a simple model with a fixed normal price and sale discount, which provides a benchmark for subsequent analysis. The IO-based model of sales is introduced in section 3, and the equilibrium of the model is characterized in section 4. The response to monetary shocks is studied in section 5. Section 6 examines the robustness of the results to different assumptions about wage flexibility. Section 7 constructs the fully dynamic extension of the model. Section 8 draws some conclusions.

2 Benchmark model

As a first pass at exploring the implications of sales for monetary policy analysis, this section adds sales in the simplest possible way to an otherwise standard macroeconomic model. While ad hoc, this benchmark model is useful in shedding light on the potential of sales as an adjustment mechanism in response to shocks, and also provides a point of comparison for later results.

The economy contains a measure-one continuum $\mathscr{H}$ of households with utility:

$$U \equiv u\left(2C^{\frac{1}{2}}m^{\frac{1}{2}}\right) - \nu(H) \,, \qquad [2.1]$$

where C denotes consumption of a composite good, m is real money balances, and H is hours of labour supplied. The utility function $u(\cdot)$ is differentiable, strictly increasing and strictly concave; the disutility function $\nu(\cdot)$ is differentiable, strictly increasing and convex.

The composite good C is a Dixit-Stiglitz aggregator over a measure-one continuum $\mathscr{T}$ of types of differentiated goods:

$$C \equiv \left(\int_{\mathscr{T}} c(\tau)^{\frac{\varepsilon-1}{\varepsilon}} d\tau\right)^{\frac{\varepsilon}{\varepsilon-1}} \,,$$

where $c(\tau)$ is consumption of good type $\tau \in \mathscr{T}$ and ε is the elasticity of substitution, which satisfies $\varepsilon > 1$.

Each household makes all its consumption purchases at only one random point in time, however, in equilibrium it is actually indifferent about when it shops. At a given point in time suppose the price of good τ is $p(\tau)$. Minimizing the cost of purchasing composite good C implies the following demand functions for each individual good τ:

$$c(\tau) = \left(\frac{p(\tau)}{P}\right)^{-\varepsilon} C \,,$$

where the price level P is:

$$P = \left(\int_{\mathscr{T}} p(\tau)^{1-\varepsilon} d\tau\right)^{\frac{1}{1-\varepsilon}} \,.$$

Households may pay different prices for individual goods depending on when they do their shopping, but in equilibrium they all face the same cost P of buying one unit of the composite good. Households hold money balances M, or equivalently, real money balances $m \equiv M/P$. The money wage is W per hour of labour. Each household receives dividends $\mathfrak{D}$ from firms (households have equal equity stakes), and a lump-sum transfer $\mathfrak{T}$, both of which are specified in money terms. The household budget constraint is thus:

$$PC + M = WH + \mathfrak{D} + \mathfrak{T} . \tag{2.2}$$

The utility-maximizing choice of real money balances implies:

$$C = \frac{M}{P} , \tag{2.3}$$

and in equilibrium, M is equal to the monetary transfer $\mathfrak{T}$. This provides a simple specification of aggregate demand, similar to a cash-in-advance constraint. There is no capital accumulation, and no government or international sectors in the economy, so goods market equilibrium requires $C = Y$, and therefore:

$$\begin{aligned} c(\tau) &= \left(\frac{p(\tau)}{P}\right)^{-\varepsilon} Y , \\ Y &= \frac{M}{P} . \end{aligned} \tag{2.4}$$

Each good is made by a single firm. With H hours of labour, a firm can produce physical output Q of its good according to the production function:

$$Q = \mathcal{F}(H) , \tag{2.5}$$

where $\mathcal{F}(\cdot)$ is a differentiable, strictly increasing and strictly concave function with $\mathcal{F}(0) = 0$. Hence the minimum cost $\mathscr{C}(Q; W)$ of producing output Q for a given wage W is:

$$\mathscr{C}(Q; W) = W\mathcal{F}^{-1}(Q) . \tag{2.6}$$

The cost function $\mathscr{C}(Q; W)$ is differentiable, strictly increasing and strictly convex in Q, and satisfies $\mathscr{C}(0; W) = 0$ as a result of the properties of the production function [2.5].

Firms sell their goods at all points in time, and can choose to vary their prices. To isolate the effects of firms adjusting the fractions of time when their goods are on sale, the benchmark model assumes that firms start with two predetermined prices, taken as exogenous here, and can choose how often each price is charged. Denote the lower of the two prices by p_S, referred to as the "sale" price, and the other price by p_N, referred to as the "normal" price. A firm then chooses the fraction of time s when its good is on sale at price p_S, with its good sold at price p_N for the remaining fraction of time $1 - s$. Firms choose the timing of their sales randomly, which is an equilibrium strategy given that all other firms are doing so. This also confirms that consumers face the same price level P irrespective of when they do their shopping.

Since households select their shopping times at random, the total quantity Q sold by a firm across all shopping times is obtained from households' demand functions as follows:

$$Q = s\left(\frac{p_S}{P}\right)^{-\varepsilon} Y + (1-s)\left(\frac{p_N}{P}\right)^{-\varepsilon} Y \ ,$$

and thus total profits $\mathscr{P}$ are:

$$\mathscr{P} = sp_S\left(\frac{p_S}{P}\right)^{-\varepsilon} Y + (1-s)p_N\left(\frac{p_N}{P}\right)^{-\varepsilon} Y - \mathscr{C}\left(s\left(\frac{p_S}{P}\right)^{-\varepsilon} Y + (1-s)\left(\frac{p_N}{P}\right)^{-\varepsilon} Y; W\right) \ . \quad [2.7]$$

Firms choose their sales fraction s to maximize profits, taking predetermined prices p_S and p_N as given for now.

Suppose that prices p_S and p_N and wage W are fixed in money terms. Now consider a shock to the money supply M. Firms adjust s in response, which means that they can effectively choose the average price they sell at. This gives them considerable freedom to respond to shocks. The following proposition establishes that firms find it optimal to use this freedom to the full: in this setting, money is neutral.

Proposition 1 *Given predetermined prices p_S and p_N, and predetermined wage W, if firms choose their sales fraction s to maximize profits* [2.7], *as long as $s \in (0,1)$ before and after the monetary shock, firms' output Q is unaffected by the shock.*

Proof The first-order condition for s is:

$$(p_S - X)p_S^{-\varepsilon} P^{\varepsilon} Y = (p_N - X)p_N^{-\varepsilon} P^{\varepsilon} Y \ , \quad [2.8]$$

where $X \equiv \mathscr{C}'(Q;W)$ is the marginal cost of producing total output Q. Because the term $P^{\varepsilon}Y$ is common to both sides of the first-order condition, the equation reduces to one involving only p_S, p_N and X. As p_S and p_N are predetermined, X must be constant. Since W is also predetermined and the cost function $\mathscr{C}(Q;W)$ is strictly convex, Q must be constant as well. ■

This result shows that monetary policy does not affect the physical output Q of firms. A positive shock to M leads firms to sell fewer goods on sale. As the quantity produced is constant, an increase in the money supply has to be followed by a corresponding increase in the price level. The prices p_S and p_N are sticky; the proportion s of goods sold on sale is responsible for the adjustment.

As households buy different goods at different prices, aggregate output Y is not exactly equal to the physical quantity of output Q. Proposition 1 shows that Q is constant in the face of monetary shocks, and though aggregate output Y is affected by these shocks, the size of the effect is extremely small and its direction is necessarily ambiguous. Furthermore, if a shock resulted in the sales fraction s changing from (almost) zero to (almost) one, then output Y would be completely unaffected.

The result of Proposition 1 is even more surprising in light of the assumption of a predetermined money wage. Usually nominal rigidity need only be present in either prices or wages for monetary policy to have significant real effects.

3 The model of sales

The benchmark model assumes that firms start with two predetermined prices p_S and p_N. However, this is not a profit-maximizing strategy given household preferences in that setting. This section proposes a model in which firms want to choose a two-price distribution even in a deterministic environment.

3.1 Households

Each household $\imath \in \mathscr{H}$ has the same utility function [2.1] over its composite consumption good $C(\imath)$, real money balances $m(\imath)$, and hours worked $H(\imath)$, as is used in the benchmark model of section 2. The budget constraint [2.2] and aggregate demand [2.4] are also as before. Household $\imath$'s utility-maximizing trade-off between consumption and leisure is given by:

$$\frac{\nu_h(H(\imath))}{u_c(C(\imath))} = \frac{W}{P}\,, \tag{3.1}$$

making use of the first-order condition [2.3] for utility-maximizing money demand. The only change to the benchmark model introduced here is in the specification of each household's composite good.

3.2 Composite goods

Household $\imath$'s consumption $C(\imath)$ is a composite good comprising a large number of individual products. Individual goods are categorized as *brands* of particular product *types*. There is a measure-one continuum $\mathscr{T}$ of product types in the economy. For each product type $\tau \in \mathscr{T}$, there is a measure-one continuum $\mathscr{B}$ of brands, with individual brands indexed by $b \in \mathscr{B}$. Hence every good in the economy is uniquely referred to by a type-brand pair $(\tau, b) \in (\mathscr{T} \times \mathscr{B})$.

Households have different preferences over this set of goods. Taking a given household, there is a set of product types $\Lambda \subset \mathscr{T}$ for which that household is *loyal* to a particular brand of each product type $\tau \in \Lambda$ in the set. For product type $\tau \in \Lambda$, the brand receiving the household's loyalty is denoted by $\mathcal{B}(\tau)$, where $\mathcal{B} : \Lambda \to \mathscr{B}$ is a mapping from types to brands. Loyalty means the household gets no utility from consuming any other brands of that product type. When the household is not loyal to a particular brand of a product type τ, that is, $\tau \in \mathscr{T} \backslash \Lambda$, the household is said to be a *bargain hunter* for product type τ. This means the household gets utility from consuming any of the brands of that product type.

The composite consumption good C for a given household is first defined in terms of a Dixit-Stiglitz aggregator over product types with elasticity of substitution ϵ. For those product types where the household is a bargain hunter, there is an additional Dixit-Stiglitz aggregator defined over brands of that product type with elasticity of substitution η. The overall aggregator is:

$$C \equiv \left(\int_{\Lambda} c(\tau, \mathcal{B}(\tau))^{\frac{\epsilon-1}{\epsilon}} d\tau + \int_{\mathscr{T} \backslash \Lambda} \left(\int_{\mathscr{B}} c(\tau, b)^{\frac{\eta-1}{\eta}} db \right)^{\frac{\eta(\epsilon-1)}{\epsilon(\eta-1)}} d\tau \right)^{\frac{\epsilon}{\epsilon-1}}\,, \tag{3.2}$$

where $c(\tau, b)$ is the household's consumption of brand b of product type τ. It is assumed that

$\eta > \epsilon > 1$, so bargain hunters are more willing to substitute between different brands of a specific product type than households are to substitute between different product types. Households have a zero elasticity of substitution between brands of product types for which they are loyal to a particular brand.

There is a measure-one continuum $\mathscr{M}$ of shopping moments when goods are purchased and consumed. A household does all its shopping at a randomly chosen moment $\jmath \in \mathscr{M}$. Denote the set of households that shop at moment $\jmath$ by $\mathscr{H}(\jmath)$. As shown later, all households are indifferent in equilibrium between all shopping moments.

The price level P is the minimum cost to the household of buying one unit of its composite good [3.2]:

$$P \equiv \min_{\{c(\tau,b)\}} \int_{\mathscr{T}} \int_{\mathscr{B}} p(\tau, b) c(\tau, b) db d\tau \ \text{ s.t. } \ C \geq 1 \ ,$$

where $p(\tau, b)$ is the price of brand b of product type τ prevailing at the household's shopping moment. For the composite good defined in [3.2], the minimized level of expenditure is:

$$P = \left(\int_{\Lambda} p(\tau, \mathcal{B}(\tau))^{1-\epsilon} d\tau + \int_{\mathscr{T} \setminus \Lambda} \left(\int_{\mathscr{B}} p(\tau, b)^{1-\eta} db \right)^{\frac{1-\epsilon}{1-\eta}} d\tau \right)^{\frac{1}{1-\epsilon}} . \qquad [3.3]$$

The expenditure-minimizing demand functions are:

$$c(\tau, b) = \begin{cases} \left(\frac{p(\tau,b)}{p_B(\tau)}\right)^{-\eta} \left(\frac{p_B(\tau)}{P}\right)^{-\epsilon} C & \text{if } \tau \in \mathscr{T} \setminus \Lambda \ , \quad \text{where } \ p_B(\tau) \equiv \left(\int_{\mathscr{B}} p(\tau, b)^{1-\eta} db\right)^{\frac{1}{1-\eta}} \ , \\ \left(\frac{p(\tau,b)}{P}\right)^{-\epsilon} C & \text{if } \tau \in \Lambda \text{ and } b = \mathcal{B}(\tau) \ , \\ 0 & \text{if } \ \tau \in \Lambda \ \text{ and } \ b \neq \mathcal{B}(\tau) \ , \end{cases} \qquad [3.4]$$

where C is the amount of the composite good consumed and P is the price level given in [3.3]. The term $p_B(\tau)$ is an index of prices for all brands of type τ, as is relevant to those households who are bargain hunters for that product type. With these demand functions, total expenditure on all consumption goods is:

$$\int_{\mathscr{T}} \int_{\mathscr{B}} p(\tau, b) c(\tau, b) db d\tau = PC \ .$$

Household preferences over goods are completely characterized by the consumption aggregator in [3.2], the loyal set Λ, and the brands $\mathcal{B}(\tau)$ receiving the household's loyalty. All households share a consumption aggregator of the same form with the same elasticities of substitution ϵ and η, but Λ and $\mathcal{B}(\tau)$ differ across households, and are randomly drawn from a probability distribution.

For each product type $\tau \in \mathscr{T}$, a household has probability λ of including type τ in its loyal set Λ. This event is independent across households and product types. Formally:

$$\mathbb{P}_{\mathscr{H}} [\tau \in \Lambda] = \lambda \ , \quad \text{for all } \ \tau \in \mathscr{T} \ . \qquad [3.5]$$

Consequently, the loyal set Λ and the set of types $\mathscr{T} \setminus \Lambda$ for which a household is a bargain hunter have measures λ and $1 - \lambda$ respectively for all households. It is assumed that $0 < \lambda < 1$, thus each household is loyal to a brand for some product types and a bargain hunter for other product types. So when households are referred to as either loyal or bargain hunters, this designation is for a specific product type only.

Conditional on including product type τ in a household's loyal set Λ, all brands of that type have an equal chance of receiving the household's loyalty. The assignment of brands to loyal households is independent across households and product types. Formally:

$$\mathbb{P}_{\mathscr{H}}\left[\mathcal{B}(\tau) \in B \middle| \tau \in \Lambda\right] = \int_B db \,, \quad \text{for all } \; B \subseteq \mathscr{B} \; \text{ and any } \; \tau \in \mathscr{T} \,. \tag{3.6}$$

Viewed from the perspective of a firm, assumptions [3.5] and [3.6] imply that it operates in a market where a randomly selected fraction λ of consumers are loyal to its product, and a fraction $1-\lambda$ are bargain hunters for its product type.

3.3 Firms

Each brand b of each product type τ is produced by a single firm, indexed by $(\tau, b) \in (\mathscr{T} \times \mathscr{B})$. All firms have the same production function $\mathcal{F}(H)$ and cost function $\mathscr{C}(Q;W)$ as are given in [2.5] and [2.6] for the benchmark model of section 2.

Each firm sells its good at every shopping moment, but not necessarily at the same price at all moments. Consider a given firm producing brand b of product type τ, and a given moment $\jmath \in \mathscr{M}$ when households $\mathscr{H}(\jmath)$ are doing their shopping. Take a particular household $\imath \in \mathscr{H}(\jmath)$. If the household is loyal to this brand and the brand is sold at price p, equation [3.4] shows that $p^{-\epsilon}(P(\imath)^{\epsilon}C(\imath))$ units are demanded. But demand is zero if the household is loyal to another brand. If the household is a bargain hunter for this product type then demand is $p^{-\eta}P_B^{\eta-\epsilon}(P(\imath)^{\epsilon}C(\imath))$, where P_B is the bargain hunters' price index for type-τ brands, that is, the price index $p_B(\tau)$ from equation [3.4] constructed using prices posted at moment $\jmath$. P_B is the same for all bargain hunters for the same product type at the same shopping moment. The only component of demand that could be household specific is $P(\imath)^{\epsilon}C(\imath)$, which multiplies the amount demanded irrespective of whether the household is loyal or a bargain hunter, and determines the overall level of expenditure. Define $\mathcal{E}(\jmath)$ as the summation of this household-specific expenditure component taken over all shoppers at moment $\jmath$:

$$\mathcal{E}(\jmath) \equiv \int_{\mathscr{H}(\jmath)} P(\imath)^{\epsilon}C(\imath)d\imath \,. \tag{3.7}$$

Since the market for any good is contains a fraction λ of potential buyers who are loyal to the product and a fraction $1-\lambda$ of bargain hunters, and because the product types and brands benefiting from households' loyalty are selected randomly according to [3.5] and [3.6], and as households choose shopping moments at random, total demand for a good sold at price p is:

$$\mathscr{D}(p;P_B,\mathcal{E}) = (\lambda + (1-\lambda)v(p;P_B))\, p^{-\epsilon}\mathcal{E} \,, \quad \text{where } \; v(p;P_B) \equiv \left(\frac{p}{P_B}\right)^{-(\eta-\epsilon)} \,, \tag{3.8}$$

at a shopping moment with bargain hunters' price index P_B for brands of the same product type, and a total household expenditure level [3.7] equal to $\mathcal{E}$. Demand is specified in terms of a function $v(p;P_B)$, referred to as the *purchase multiplier*, which is defined as the ratio of the amounts sold at the same price to a given measure of bargain hunters relative to the same measure of loyal customers.

The demand function $\mathscr{D}(p; P_B, \mathcal{E})$ is used to calculate the total revenue $\mathscr{R}(q; P_B, \mathcal{E})$ received from selling quantity q at a particular shopping moment with P_B and $\mathcal{E}$ given:

$$\mathscr{R}(q; P_B, \mathcal{E}) \equiv q\mathscr{D}^{-1}(q; P_B, \mathcal{E}) \ , \quad \text{with} \ \ p = \mathscr{D}^{-1}(q; P_B, \mathcal{E}) \ , \tag{3.9}$$

where $\mathscr{D}^{-1}(q; P_B, \mathcal{E})$ is the inverse demand function corresponding to [3.8].

3.4 Price setting

Now consider the profit-maximization problem for a given firm, which chooses a price for its product at each shopping moment. As is shown below, the total household expenditure level $\mathcal{E}$, defined in [3.7], is the same at all moments in equilibrium. Furthermore, the bargain hunters' price index P_B appearing in the demand function [3.8] is the same for all product types and at all moments. Under these conditions, the profit-maximization problem reduces to choosing the distribution of prices used across shopping moments.

For the specification of demand found in the benchmark model of section 2, firms would choose a single price at all moments even if they were to have the option of choosing a general distribution. But this is not true when households have the heterogeneous preferences described in section 3.2.

Let $F(p)$ be a general distribution function for prices. This distribution function is chosen to maximize profits $\mathscr{P}$:

$$\mathscr{P} = \int_p \mathscr{R}\left(\mathscr{D}(p; P_B, \mathcal{E}); P_B, \mathcal{E}\right) dF(p) - \mathscr{C}\left(\int_p \mathscr{D}(p; P_B, \mathcal{E}) dF(p); W\right) \ , \tag{3.10}$$

where $\mathscr{R}(q; P_B, \mathcal{E})$ is total revenue from selling quantity q at one shopping moment, defined in [3.9], and $\mathscr{C}(Q; W)$ is the total cost [2.6] of producing the entire output Q of the good sold across all shopping moments.

Consider a discrete distribution of prices $\{p_i\}$ with weights $\{\omega_i\}$.[4] The first-order conditions for maximizing [3.10] with respect to prices p_i and weights ω_i are:

$$\mathscr{R}'\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) = \mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{and} \tag{3.11a}$$

$$\mathscr{R}\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) = \aleph + \mathscr{D}(p_i; P_B, \mathcal{E})\mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{if} \ \ \omega_i > 0 \ ; \quad \text{or} \tag{3.11b}$$

$$\mathscr{R}\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) \leq \aleph + \mathscr{D}(p_i; P_B, \mathcal{E})\mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{if} \ \ \omega_i = 0 \ , \tag{3.11c}$$

where $\aleph$ is the Lagrangian multiplier attached to the constraint $\sum_j \omega_j = 1$. Equation [3.11a] is the usual marginal revenue equals marginal cost condition, which must hold for any price that receives positive weight. The interpretation of first-order conditions [3.11b] and [3.11c] is discussed later.

[4] It is shown later that restricting attention to discrete distributions is without loss of generality.

3.5 Aggregation

Since all households are randomizing over their choice of shopping moment, and brand loyalty is drawn randomly according to [3.5] and [3.6], there is no intrinsic difference for firms between any two shopping moments. And as long as firms are selecting prices from their desired price distributions at random for particular shopping moments, the bargain hunters' price index P_B is the same at all moments and for all product types.[5] This also means that $P(\imath) = P$ for all households $\imath \in \mathscr{H}$, and it therefore makes sense to talk about *the* aggregate price level P, in spite of households' consumption baskets differing.

Given that households share a common price level and have the same preferences [2.1] over their composite goods, money balances and hours, it follows that all households choose the same levels of composite consumption, real money balances and hours. That is, $C(\imath) = C$, $m(\imath) = m$ and $H(\imath) = H$ for all $\imath \in \mathscr{H}$. Since consumption is the only source of demand in the economy, goods market equilibrium requires $C = Y$, where Y is aggregate income.

The common level of consumption $C = Y$ and the common price level P imply that the total household expenditure level [3.7] is the same across all shopping moments, as claimed earlier. This is equal to $\mathcal{E} = P^{\epsilon}Y$ at every moment $\jmath \in \mathscr{M}$. Together with the randomization assumptions for brand loyalty, this justifies the claim that all firms face the same demand function at any shopping moment, so a firm cannot improve upon randomly selecting the moments at which it charges particular prices from its desired price distribution. Finally, note that randomization by firms makes households indifferent between all shopping moments, as was claimed.

4 Equilibrium with flexible prices and wages

Consider the equilibrium of the economy when all exogenous variables are constant and prices can be freely adjusted as discussed in section 3.4, and wages adjust to clear the labour market. With a constant money supply, and no shifts in the production function [2.5], the aggregate price level and aggregate output are also constant.

The equilibrium pricing strategies chosen by firms depend on the nature of the demand function $\mathscr{D}(p; P_B, \mathcal{E})$ for a firm's product at a particular shopping moment, as given in equation [3.8]. What is crucial is that demand comes from two different sources: loyal customers and bargain hunters — and these groups have different price sensitivities. Loyal customers will not switch to other brands, so the only margin of substitution they have is shifting expenditure to other types of product in their consumption basket. On the other hand, bargain hunters want to find the brands offering the best deals for a particular product type. The price elasticities of these two groups are ϵ and η respectively, and it makes sense to assume $\eta > \epsilon$. This means that the overall demand function faced by a firm does not have a constant price elasticity: the elasticity changes with the composition of the firm's customers. High prices mean that most bargain hunters have deserted its brand, and the residual mass of loyal customers has a low price elasticity. Low prices put the firm in contention to win over the bargain hunters, but fierce competition among brands for these customers means

[5]This is true when the distribution of firms' desired price distributions is not different across product types. This requirement is satisfied at all points in the paper, including the fully dynamic extension of the model.

the price elasticity is high.

These arguments suggest that the price elasticity of demand decreases with price. This is confirmed by differentiating demand function $\mathscr{D}(p; P_B, \mathcal{E})$ from [3.8] to obtain the price elasticity $\zeta(p; P_B)$ (in absolute value):

$$\zeta(p; P_B) = \frac{\lambda\epsilon + (1-\lambda)\eta v(p; P_B)}{\lambda + (1-\lambda)v(p; P_B)} \, . \tag{4.1}$$

The elasticity is a weighted average of ϵ and η, with the weight on the larger elasticity η increasing with the purchase multiplier $v(p; P_B)$, as defined in [3.8]. The higher is the price p, the lower is the purchase multiplier, and the smaller is the price elasticity. Such a change in elasticity is simply a less extreme version of a "kinked" demand curve. For very low prices, the elasticity is approximately constant and equal to η because the bargain hunters are preponderant; for very high prices, it is approximately constant and equal to ϵ because only loyal customers remain. In the middle of the demand function there is a smooth transition between ϵ and η.

As is the case with a kinked demand curve, the varying price elasticity of demand means that the marginal revenue function is not necessarily downward sloping for all prices, even though the demand function [3.8] is downward sloping everywhere. To see this, note that marginal revenue $\mathscr{R}'(q; P_B, \mathcal{E})$ can be expressed in terms of price $p = \mathscr{D}^{-1}(q; P_B, \mathcal{E})$ and the price elasticity of demand $\zeta(p; P_B)$ as follows:

$$\mathscr{R}'(\mathscr{D}(p; P_B, \mathcal{E}); P_B, \mathcal{E}) = \left(1 - \frac{1}{\zeta(p; P_B)}\right) p \, . \tag{4.2}$$

It can be confirmed that if η is sufficiently large for a given ϵ then marginal revenue is indeed non-monotonic.

Proposition 2 *Consider the total revenue function $\mathscr{R}(q; P_B, \mathcal{E})$ defined in [3.9] and suppose that $\eta > \epsilon > 1$. Then marginal revenue $\mathscr{R}'(q; P_B, \mathcal{E})$ is non-monotonic (initially decreasing, then increasing on an interval, and then subsequently decreasing) if and only if $0 < \lambda < 1$ and*

$$\eta > (3\epsilon - 1) + 2\sqrt{2\epsilon(\epsilon - 1)} \tag{4.3}$$

hold, and everywhere decreasing otherwise.

Proof See appendix A.2. ■

Observe from [4.2] that to obtain non-monotonicity it is necessary to have a sufficiently large increase in the price elasticity $\zeta(p; P_B)$ outweighing a falling price in some range. From [4.1], this happens when the gap between ϵ and η is larger.

With a non-monotonic marginal revenue function $\mathscr{R}'(q; P_B, \mathcal{E})$, it is possible that two prices and quantities are associated with the same level of marginal revenue. First-order condition [3.11a] then suggests firms might want to charge different prices at different shopping moments.

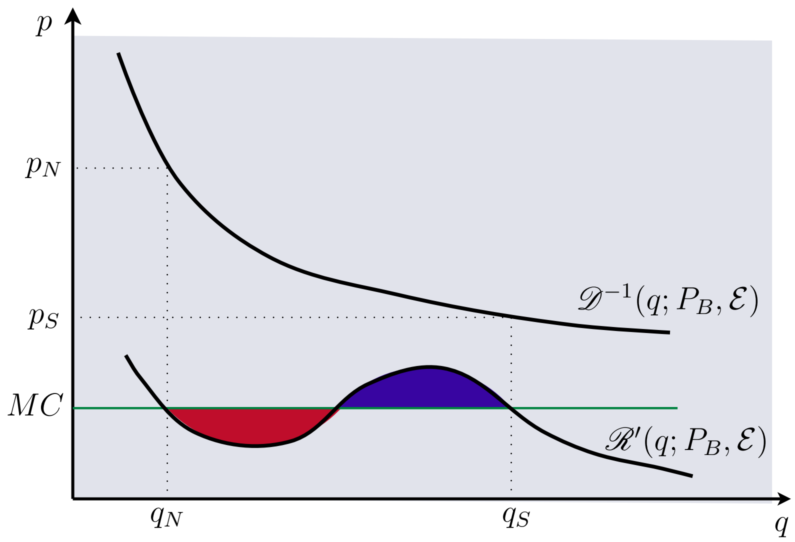
As was discussed in the introduction, an interesting case is where firms find it optimal to choose a distribution with two prices: a "normal" high price, and a low "sale" price. Denote these two prices respectively by p_N and p_S with $p_N > p_S$, and let $q_N = \mathscr{D}(p_N; P_B, \mathcal{E})$ and $q_S = \mathscr{D}(p_S; P_B; \mathcal{E})$ be the quantities demanded at a single shopping moment by all customers at these prices. The fraction of shopping moments when a firm's good is on sale is denoted by s. If $0 < s < 1$ then both prices must satisfy first-order conditions [3.11a]–[3.11b]. By eliminating the Lagrangian multiplier $\aleph$ from [3.11b], profit maximization requires:

$$\mathscr{R}'(q_N; P_B, \mathcal{E}) = \mathscr{R}'(q_S; P_B, \mathcal{E}) = \frac{\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}(q_N; P_B, \mathcal{E})}{q_S - q_N} = \mathscr{C}'\left(sq_S + (1-s)q_N; W\right) \ . \quad [4.4]$$

Hence there are three requirements for the optimality of this price distribution: marginal revenue must be equalized at both normal and sale prices, the extra revenue generated by having a sale at a particular shopping moment per extra unit sold must be equal to the common marginal revenue; and marginal revenue and average extra revenue must both equal marginal cost.

Firms have a choice at which shopping moment they sell each unit of their output, so switching a unit from one moment to another must not increase total revenue, thus marginal revenue must be equalized at all prices that are used at some shopping moment. Furthermore, firms must be indifferent between having a sale or not at one particular moment. This requires that the extra revenue generated by the sale per extra unit sold must equal marginal cost.

Figure 2: *Demand function and non-monotonic marginal revenue function*



Notes: Schematic representation of demand function [3.8] and marginal revenue function [4.2]. Shown for the case where elasticities ϵ and η satisfy [4.3].

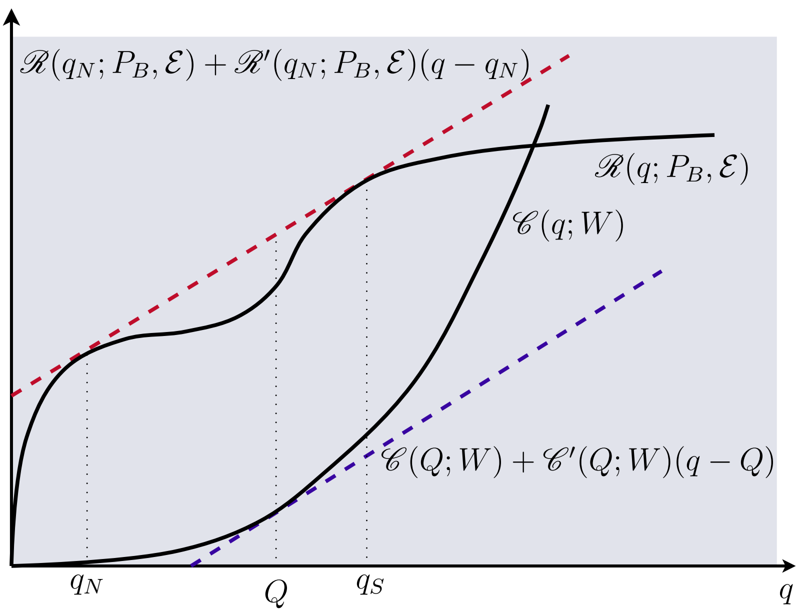
A graphical interpretation of the first two optimality conditions from [4.4] is shown in Figure 2. Marginal revenue is initially downward sloping, then becomes upward sloping, before finally changing direction once more to become downward sloping again. Both quantities q_N and q_S are associated with the same marginal revenue, which is in turn equal to the marginal cost MC of producing total output $Q = sq_S + (1-s)q_N$ (note that the marginal cost function is not shown).[6] The average extra revenue condition is represented in the diagram by the equality of the two shaded areas bounded

[6]There is a third point between q_N and q_S also associated with the same marginal revenue, but including this point in a firm's price distribution would violate the second-order conditions for profit maximization.

between the marginal revenue function and the horizontal line MC, and between the quantities q_N and q_S.

The full set of first-order conditions [4.4] is depicted using the total revenue and total cost functions shown in Figure 3. The first point to note is that because firms can charge different prices at different shopping moments, the set of achievable total revenues is convexified.[7] This raises attainable total revenue in the range between q_N and q_S. The first two conditions for profit maximization in [4.4] require that the total revenue function has a common tangent line at both quantities q_N and q_S, which is equivalent to the slope of the chord being the same as that of the tangent itself. This slope then determines the unique point where marginal cost equals marginal revenue, which yields the total quantity sold and hence the proportion s of sales.

Figure 3: *Total revenue and total cost functions with first-order conditions*



Notes: Schematic representation of total revenue function $\mathscr{R}(q; P_B, \mathcal{E})$ from [3.9] and total cost function $\mathscr{C}(Q; W)$ from [2.6], shown for elasticities ϵ and η satisfying [4.3].

Conditions for the existence and uniqueness of the two-price equilibrium are given below.

Theorem 1 *Suppose firms choose distributions of prices to maximize profits $\mathscr{P}$ in* [3.10].

(i) If elasticities ϵ and η are such that the non-monotonicity condition [4.3] *holds then there exist $\underline{\lambda}(\epsilon, \eta)$ and $\overline{\lambda}(\epsilon, \eta)$ such that $0 < \underline{\lambda}(\epsilon, \eta) < \overline{\lambda}(\epsilon, \eta) < 1$, and if $\lambda \in (\underline{\lambda}(\epsilon, \eta), \overline{\lambda}(\epsilon, \eta))$ then there exists a two-price equilibrium, and no other equilibria exist.*

(ii) If either of these conditions fails then there exists a one-price equilibrium, and no other equilibria exist.

Proof See appendix A.3 ■

[7] This also implies that the first-order conditions for profit maximization are sufficient as well as necessary.

This result indicates that two conditions need to be fulfilled if two prices are to be the unique equilibrium. First, marginal revenue must be non-monotonic, which requires a sufficiently large difference between the elasticities ϵ and η, as has been discussed above and analysed in Proposition 2. Second, neither loyal consumers nor bargain hunters should predominate among a firm's potential customers: there must be a sufficient mixture of the two types. This justifies a firm having a high price at some moments aimed at its loyal customers, and a low one at others aimed at bargain hunters, even though no actual price discrimination is practised since it is not possible for the firm to distinguish the two types prior to the moment of purchase.

Note that since there is no reference to the degree of convexity of the cost function in Theorem 1, it is only the preference parameters ϵ, η and λ that determine whether the unique equilibrium features two prices or not. This is because even if marginal cost were constant, the actions of other firms affect the total revenue function, in particular the slope of the chord in Figure 3, as is discussed in full later. Convexity of the cost function ensures individual firms strictly prefer a two-price distribution across shopping moments when this is the equilibrium price distribution for brands of their product type. Hence the two-price equilibrium cannot arise from firms choosing different one-price distributions.

Although this analysis considers just two types of consumer, adding more types does not necessarily make additional prices sustainable in equilibrium. There are two reasons: more prices in equilibrium requires more undulations in the marginal revenue function, and a common tangent line of the total revenue function at more than two points. Neither of these two configurations follows automatically on augmenting the model with extra types of consumer.

Given the stylized facts discussed in the introduction, this paper focuses on parameters ϵ, η and λ for which there is a unique two-price equilibrium. The total physical quantity of output sold by firms is $Q = sq_S + (1-s)q_N$. Using the price elasticity $\zeta(p; P_B)$ from [4.1] and the relationship between price and marginal revenue in [4.2], the marginal revenue equals marginal cost conditions for each price are expressed in terms of desired markups on marginal cost $X \equiv \mathscr{C}'(Q; W)$:

$$p_S = \mu(p_S; P_B)X \ , \quad p_N = \mu(p_N; P_B)X \ , \quad \text{with} \ \ \mu(p; P_B) = \frac{\lambda\epsilon + (1-\lambda)\eta v(p; P_B)}{\lambda(\epsilon-1) + (1-\lambda)(\eta-1)v(p; P_B)} \ . \tag{4.5}$$

The desired markup function $\mu(p; P_B)$ depends on the parameters ϵ, η and λ, and the purchase multiplier $v(p; P_B)$ from [3.8]. Let $v_S \equiv v(p_S; P_B)$ and $v_N \equiv v(p_N; P_B)$ denote the two purchase multipliers, and $\mu_S \equiv \mu(p_S; P_B)$ and $\mu_N \equiv \mu(p_N; P_B)$ the associated desired markups:

$$\mu_S = \frac{\lambda\epsilon + (1-\lambda)\eta v_S}{\lambda(\epsilon-1) + (1-\lambda)(\eta-1)v_S} \ , \quad \mu_N = \frac{\lambda\epsilon + (1-\lambda)\eta v_N}{\lambda(\epsilon-1) + (1-\lambda)(\eta-1)v_N} \ . \tag{4.6}$$

By using demand function [3.8], the first-order condition in [4.4] linking average extra revenue to marginal cost is expressed as:

$$\frac{\mu_S - 1}{\mu_N - 1} = \frac{(\lambda + (1-\lambda)v_N)\mu_N^{-\epsilon}}{(\lambda + (1-\lambda)v_S)\mu_S^{-\epsilon}} \ . \tag{4.7}$$

Given that a fraction s of all prices are set at p_S and the remaining $1-s$ are set at p_N, equation

[3.4] implies the bargain hunters' price index P_B is:

$$P_B = \left(sp_S^{1-\eta} + (1-s)p_N^{1-\eta}\right)^{\frac{1}{1-\eta}} \ , \qquad [4.8]$$

which is used to calculate the purchase multipliers and characterize the desired markups μ_S and μ_N.

In finding the equilibrium, the model has a convenient block-recursive structure, that is, the microeconomic aspects of the equilibrium can be characterized independently of the macroeconomic equilibrium, which is then determined afterwards. The key micro variables are the sales fraction s, the markups μ_S and μ_N, the markup ratio $\mu \equiv \mu_S/\mu_N$, and the ratio of the quantities sold at the sale and normal prices, denoted by $\chi \equiv q_S/q_N$. Using the demand function [3.8], equation [4.7] yields a relationship between the quantity ratio χ and desired markups μ_S and μ_N:

$$\chi = \frac{\mu_N - 1}{\mu_S - 1} \ .$$

The following proposition verifies the block-recursive structure of the model and derives some comparative statics results.

Proposition 3 *Suppose parameters ϵ, η and λ are such that there is a unique two-price equilibrium.*

(i) The equilibrium values of μ, χ, s, v_S, v_N, μ_S and μ_N are functions only of the parameters ϵ, η and λ.

(ii) The equilibrium values of μ, χ, μ_S and μ_N are functions only of the parameters ϵ and η.

(iii) Let $\underline{\lambda}(\epsilon, \eta)$ and $\overline{\lambda}(\epsilon, \eta)$ be as defined in Theorem 1. Then:

$$\lim_{\lambda \to \underline{\lambda}(\epsilon,\eta)^+} s = 1 \ , \qquad \lim_{\lambda \to \overline{\lambda}(\epsilon,\eta)^-} s = 0 \ , \qquad \frac{\partial s}{\partial \lambda} < 0 \ .$$

(iv) The markup and quantity ratios μ and χ are continuous functions of ϵ and η, and:

$$\lim_{\epsilon \to 1^+} \mu = 0 \ , \quad \lim_{\epsilon \to 1^+} \chi = \infty \ , \quad \lim_{\eta \to \eta^*(\epsilon)^+} \mu = 1 \ , \quad \lim_{\eta \to \eta^*(\epsilon)^+} \chi = 1 \ ,$$

where $\eta^*(\epsilon) \equiv (3\epsilon - 1) + 2\sqrt{2\epsilon(\epsilon - 1)}$ is the lower bound for η which ensures non-monotonicity of the marginal revenue function according to Proposition 2.

PROOF See appendix A.4. ■

The first part of the proposition establishes the separation of the equilibrium for the microeconomic variables from the broader macroeconomic equilibrium. Furthermore, the second part shows that only ϵ and η are needed to pin down the markup ratio μ and the quantity ratio χ, which together with λ then determine the sales fraction s. The equilibrium sales fraction s is strictly decreasing in λ and varies from one to zero as λ spans its interval of values consistent with a two-price equilibrium. The final part shows there is a smooth transition between the two-price and the one-price equilibria, and that the markup and quantity ratios span their natural ranges for admissible parameter values.

Given the purchase multipliers v_S and v_N and markups μ_S and μ_N, finding the macroeconomic equilibrium is straightforward. The aggregate price level P is obtained by combining [3.3] and demand function [3.4], and making use of the definition of the purchase multiplier $v(p; P_B)$ from [3.8]:

$$P = \left(s(\lambda + (1-\lambda)v_S)p_S^{1-\epsilon} + (1-s)(\lambda + (1-\lambda)v_N)p_N^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}} \ .$$

This allows the level of real marginal cost $x \equiv X/P$ to be deduced as follows:

$$x = \left(s(\lambda + (1-\lambda)v_S)\mu_S^{1-\epsilon} + (1-s)(\lambda + (1-\lambda)v_N)\mu_N^{1-\epsilon}\right)^{\frac{1}{\epsilon-1}} \ . \qquad [4.9]$$

With real marginal cost and the desired markups, relative prices $\varrho_S \equiv p_S/P$ and $\varrho_N \equiv p_N/P$ are determined. These yield the amounts sold at the two prices relative to aggregate output:

$$q_S = (\lambda + (1-\lambda)v_S)\, \varrho_S^{-\epsilon} Y \ , \quad q_N = (\lambda + (1-\lambda)v_N)\, \varrho_N^{-\epsilon} Y \ . \qquad [4.10]$$

Given that total physical output is $Q = sq_S + (1-s)q_N$, the ratio of Y to Q, denoted by δ, is:

$$\delta \equiv \frac{1}{s(\lambda + (1-\lambda)v_S)\varrho_S^{-\epsilon} + (1-s)(\lambda + (1-\lambda)v_N)\varrho_N^{-\epsilon}} \ ,$$

which satisfies $0 < \delta < 1$. The production function [2.5], cost function [2.6], and labour supply function [3.1] imply a positive relationship between real marginal cost x and aggregate output Y:

$$x = \frac{\nu_h\left(\mathcal{F}^{-1}(Y/\delta)\right)}{u_c(Y)\mathcal{F}'\left(\mathcal{F}^{-1}(Y/\delta)\right)} \ . \qquad [4.11]$$

Since the equilibrium real marginal cost x is already known from [4.9], the equation above uniquely determines output Y. Using equation [2.4], the aggregate price level P is then given by $P = M/Y$.

5 Monetary shocks in a model of sales

The benchmark model of section 2 analysed the effects of a monetary shock with predetermined prices p_S and p_N and wage W alongside flexibility in the sales fraction s, but critically, the reason why firms had a two-price distribution rather than just a single price was left unexplained. The sales model introduced in section 3 provides precisely such a reason, and this section performs a similar experiment when sales are flexible.[8]

Starting from the flexible-price equilibrium characterized in section 4, suppose that prices p_S and p_N and wage W are initially set at levels consistent with the expected money supply $\bar{M}$. Following the realization of the actual money supply M, firms can adjust their sales by changing both the frequency s and the price p_S. The normal price p_N remains at its predetermined level, and for now, the money wage W also remains constant.

The freedom to adjust sales through s and p_S, but not the normal price p_N, means that of the

[8] Actually the exercise here gives firms greater freedom than in the benchmark model by allowing the size of the sale discount to be adjusted. In the benchmark model, if p_S could be changed then money would be automatically neutral because the profit-maximizing strategy in that setting is to charge a single price.

first-order conditions in [4.4], only the second and third equalities hold:

$$\frac{p_S q_S - p_N q_N}{q_S - q_N} = X \ , \quad p_S = \mu(p_S; P_B)X \ , \qquad [5.1]$$

where X denotes marginal cost. Note that achieving the optimal markup $\mu(p_S; P_B)$ at the sale price is equivalent to equalizing marginal revenue at p_S and marginal cost.

The use of the sales margin of adjustment in the benchmark model led to money neutrality. But it turns out that the answer to the question of whether monetary shocks have real effects is radically different once a reason for sales is built into the model: monetary shocks now have large real effects. The crux of the result is that sales are strategic substitutes: a firm finds a sale more attractive when other firms are holding fewer sales.

Monetary shocks are analysed by considering a situation where the money supply is in a neighbourhood of its expected value. Denote log deviations of variables from their corresponding flexible-price equilibrium value using sans serif letters, and from here onwards, the flexible-price equilibrium values themselves with a bar over the variable.

Theorem 2 *Consider parameter values ϵ, η and λ for which the economy has a unique two-price equilibrium, as described in Theorem 1.*

(i) If the sales fraction s is adjusted optimally according to the first equation in [5.1] then the elasticity of marginal cost X with respect to P_B is unity, and no other variables have first-order effects on marginal cost:

$$\mathsf{X} = \mathsf{P}_B \ .$$

(ii) If both the sales fraction s and the sale price p_S are adjusted optimally according to [5.1] then the elasticity of the optimal sale price p_S with respect to marginal cost X is unity, and no other variables have first-order effects on the optimal sale price:

$$\mathsf{p}_S = \mathsf{X} \ .$$

Proof See appendix A.6. ■

The first part of the theorem demonstrates that sales are strategic substitutes. As other firms cut back on sales either by reducing s or increasing p_S, the bargain hunters' price index P_B in [4.8] increases. Theorem 2 shows this leads a given firm optimally to increase its total quantity sold to the point where marginal cost X has risen one-for-one in percentage terms with P_B. As the firm's normal price is not adjusted, the increase in quantity sold is brought about by holding more sales.

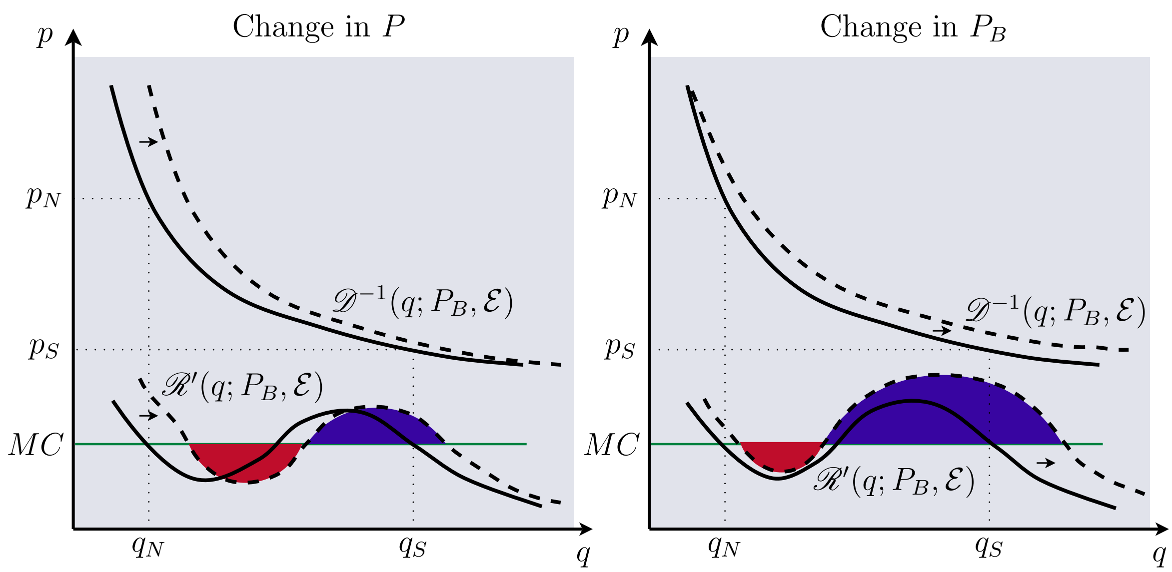
The problem of choosing the profit-maximizing adjustment of sales is essentially one of a firm deciding how much to target its loyal customers versus bargain hunters for its product type. Because competition for bargain hunters is more intense than for loyal customers, the incentive to target the bargain hunters is much more sensitive to the extent that other firms are targeting them as well. Thus, a firm's desire to target the bargain hunters with sales is decreasing in the extent to which others are doing the same.

Recall that in the benchmark model, firms have an incentive to reduce sales in response to a positive monetary shock, essentially mimicking an increase in price. The same incentive exists here, but is counteracted by another effect. As firms reduce their sales, an individual firm has a strong incentive to target the bargain hunters, who are being neglected by others. Consequently, the fall in sales is smaller, and so the price level rises by less. Therefore, output increases.

This analysis demonstrates that there are two conflicting effects on sales and the price level after a monetary shock. One tends toward money neutrality, while the other tends toward money having real effects. It is therefore a quantitative question how strong monetary policy's real effects will be.

The effects of others' actions on an individual firm's incentives to hold sales are shown in Figure 4. Others' price changes affect the demand function through both P and P_B. A rise in P shifts demand outward, with a proportional effect at every point. In contrast, a rise in P_B has a much more marked effect on demand at lower prices and higher quantities where the bargain hunters are found. Such a change upsets the balance between profits from selling at the sale and normal prices, boosting profits from selling on sale. This is seen in the differential between the shaded areas bounded between marginal revenue and marginal cost. The imbalance does not occur following a change in P, which is the only operative channel in the benchmark model.

Figure 4: *Impact of other firms' price changes on the demand and marginal revenue functions*



Notes: Schematic representation of shifts of demand and marginal revenue functions [3.8] and [4.2]. The aggregate price level P affects demand through $\mathcal{E} = P^{\epsilon}Y$ according to [3.7].

The discussion above explains why there must be a positive relationship between P_B and marginal cost X (working through changes in the desired sales fraction and thus a firm's total quantity sold), but the result of Theorem 2 is stronger: the elasticity must be unity. This follows from some elementary properties of the profit function. Define $\mathscr{P}(p; P_B, X, P, Y)$ to be a firm's profits at the margin from setting price p at one shopping moment:

$$\mathscr{P}(p; P_B, X, P, Y) \equiv (p - X)\mathscr{D}(p; P_B, P^{\epsilon}Y) \,, \qquad [5.2]$$

where $\mathcal{E} = P^{\epsilon}Y$ is used, in accordance with [3.7]. The ratio of marginal profits from selling at the sale and normal prices is:

$$\wp(p_S, p_N, P_B, X, P, Y) \equiv \frac{\mathscr{P}(p_S; P_B, X, P, Y)}{\mathscr{P}(p_N; P_B, X, P, Y)} \,, \qquad [5.3]$$

and the equation $\wp(p_S, p_N, P_B, X, P, Y) = 1$ is equivalent to the first-order condition for the optimal sales fraction in [5.1].

The demand function $\mathscr{D}(p; P_B, P^{\epsilon}Y)$ is homogeneous of degree zero in all prices, so the profit function [5.2] must be homogeneous of degree one in p, P_B, P and X, and hence the profit ratio $\wp(p_S, p_N, P_B, X, P, Y)$ is homogeneous of degree zero in p_S, p_N, P_B, P and X. The form of the demand function $\mathscr{D}(p; P_B, P^{\epsilon}Y)$ in [3.8] implies that P and Y proportionately affect profits at both the sale and normal prices and so have no influence on relative profits, thus $\wp(p_S, p_N, P_B, X, P, Y) = \wp(p_S, p_N, P_B, X, 1, 1)$ for all P and Y. Consequently, relative profits [5.3] must be homogeneous of degree zero in p_S, p_N, P_B and X alone. Since p_S and the predetermined value of p_N are chosen optimally, neither p_S nor p_N has a first-order effect on either profits or relative profits. Therefore, the profit ratio $\wp(p_S, p_N, P_B, X, 1, 1)$ must be locally homogeneous of degree zero in just P_B and X. Hence to ensure the ratio remains equal to one, P_B and X must change by the same proportion.

The option of adjusting the sales fraction s was also open to firms in the benchmark model, but here the use of this margin has important implications for competition among firms. By comparing the sales first-order condition [2.8] in the benchmark model with the analogous condition $\wp(p_S, p_N, P_B, X, P, Y) = 1$ in the model of sales, the key difference is the presence of P_B, the bargain hunters' price index, which influences demand differently at prices p_S and p_N.

The second part of Theorem 2 states that when both the sales fraction and sale price are optimally adjusted, the chosen sale price features a constant markup on marginal cost, at least locally. The first-order condition for the profit-maximizing sale price is $p_S/(\mu(p_S, P_B)X) = 1$, and this equation is homogeneous of degree zero in p_S, P_B and X because the optimal markup function $\mu(p; P_B)$ in [4.5] is homogeneous of degree zero in prices. As P_B and X must move proportionately to be consistent with the optimal choice of the sales fraction, a movement of p_S in the same proportion is required to satisfy the first-order condition for the sale price.

5.1 Calibration

The distinguishing parameters of the sales model are the two elasticities ϵ and η and the fraction λ of loyal consumers. As is shown in section 4, these parameters are directly related to observable prices and quantities: the markup ratio μ, which gives the size of the discount offered when a good is on sale; the quantity ratio χ, which measures proportionately how much more is purchased when a good is on sale; and the fraction s of goods sold at sale prices. There are thus three parameters that can be matched to data on three observables.

There is a growing empirical literature examining price adjustment patterns at the microeconomic level. This literature provides information about the markup ratio μ and the sales fraction s. The baseline values of these variables are taken from Nakamura and Steinsson (2007). Their study draws on individual price data from the BLS CPI research database, which covers approximately 70% of

U.S. consumer expenditure. They report that the fraction of price quotes which are sales (weighted by expenditure) is 7.4%. They also report that the median difference between $\log(p_S)$ and $\log(p_N)$ is 0.295, which yields $\mu = 0.745$.

In the retail and marketing literature, there has been for a long time a discussion of the effects of price promotions on demand. This literature provides information about the quantity ratio. Papers typically report a range of estimates conditional on factors other than price that affect the impact of a price promotion, for example, advertising. The baseline value of the quantity ratio is obtained from the study by Chakravarthi, Neslin and Sen (1996). Their results are based on scanner data from a large number of U.S. supermarkets. According to the elasticities they report, a temporary price cut of the size consistent with the markup ratio taken from Nakamura and Steinsson (2007) implies a quantity ratio of between approximately 4 and 6 if retailers draw their sale to the attention of customers. The baseline number used here is the midpoint of this range, so $\chi = 5$.

The three facts about sales, summarized in Table 1, are then used to find matching values of the three unknown parameters. This first requires a method for obtaining the unique equilibrium values of μ, χ and s. Proposition 3 shows that these depend only on the parameters ϵ, η and λ, and furthermore, that μ and χ are functions of ϵ and η alone. Procedures for calculating the equilibrium values of μ and χ, and then s, are described in appendix A.1 and appendix A.3 respectively.

Table 1: *Stylized facts about sales*

Description	Parameter	Value
Ratio of the sale markup to the normal markup (μ_S/μ_N)	μ	0.745*
Ratio of quantities sold at the sale price and the normal price (q_S/q_N)	χ	5†
Fraction of goods sold at the sale price	s	0.074*

* *Source*: Nakamura and Steinsson (2007)
† *Source*: Chakravarthi, Neslin and Sen (1996)

Given the mapping from parameters to the equilibrium of the model, parameters matching the three stylized facts were found by applying the Nelder-Mead simplex algorithm. The results are shown in Table 2. An extensive grid search over the parameters ϵ and η was used to verify that no other values are consistent with the target values of μ and χ. Proposition 3 demonstrates that given ϵ and η, there is always one and only one value of λ matching the target sales fraction s.

Table 2: *Parameters matching the stylized facts about sales*

Description	Parameter	Value
Elasticity of substitution between product types	ϵ	3.01
Elasticity of substitution between brands for a bargain hunter	η	19.7
Fraction of product types for which a household is loyal to a brand	λ	0.901

Notes: These parameters are the only values exactly consistent with the three stylized facts about sales given in Table 1.

In order to compute the effects of a monetary policy shock, the elasticity of marginal cost with respect to output must also be known, which requires one further parameter to be calibrated. This is done by specifying a production function:

$$\mathcal{F}(H) = AH^{\alpha} , \tag{5.4}$$

and setting $\alpha = 2/3$ to match the labour share of income.

5.2 Results

This section calculates the elasticities of aggregate output and the price level with respect to a monetary surprise, evaluated at the flexible-price equilibrium described in section 4. The results draw on the first-order Taylor approximation of the model presented in appendix A.7. The equilibrium values of output and the price level are determined under the assumption that the sales fraction and the sale price are optimally adjusted, but the normal price and money wage remain at their predetermined equilibrium values. The equations that characterize the equilibrium after a monetary shock are as in section 4, except that the first-order conditions for price p_N in [4.4] and wage W in [4.11] are dropped. The first-order conditions for profit-maximizing sales are summarized in [5.1].

The results for the baseline calibration are examined first.[9] Using the parameters from Table 2 and $\alpha = 2/3$, the elasticities of output and the price level with respect to a monetary shock are:

$$\boxed{\frac{d\log Y}{d\log M} = 0.895 , \quad \frac{d\log P}{d\log M} = 0.105 .}$$

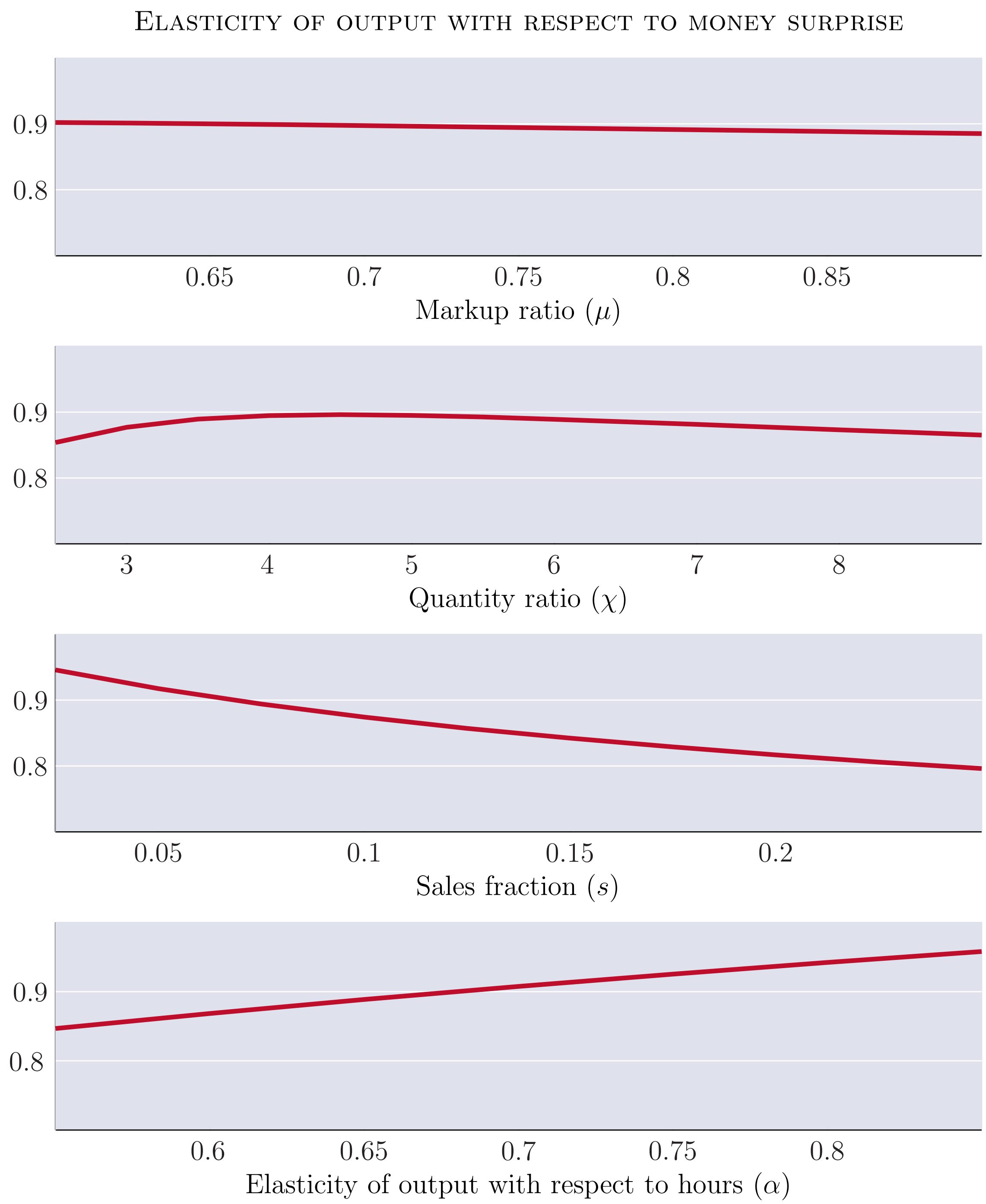
For a 1% surprise increase in the money supply, output rises by 0.895%. The sensitivity analysis in Figure 5 shows that this finding is not very sensitive to the stylized facts about sales and the production function used to calibrate the model. Of the target values, the real effects of money surprises are most sensitive to the sales fraction s. In the range of empirically plausible values of s (5% – 15%), monetary policy has substantial real effects: the elasticity of output with respect to a money surprise is found between 0.84 and 0.92.

The quantity ratio χ is the target for which the literature yields the widest range of estimates. But nonetheless, varying χ from 3 to 8 implies that the elasticity lies between 0.87 and 0.90. The target value of the elasticity of output with respect to hours α has some influence on the size of the real effects of money surprises, with the output response lying in the range 0.85–0.95 for reasonable choices of this parameter, but all values in this range imply substantial real effects. Finally, the target value of the markup ratio μ makes essentially no difference to the results.

These findings are in sharp contrast to the money neutrality result of the experiment performed using the benchmark model of section 2, where there was no rationale for firms having a two-price distribution. In the new model, consumer preferences are such that sales are an equilibrium phenomenon. In both cases, firms have an incentive to revise the frequency of sales and the size of

[9] Although determining the flexible-price equilibrium requires specifying the utility function [2.1], this information is not needed to compute the elasticities of output and the price level. This is seen by examining the first-order Taylor approximation of the model in appendix A.7.

Figure 5: *Sensitivity analysis for the real effects of monetary shocks*



Notes: For each graph, the results are obtained by fixing the other targets at their baseline values as given in Table 1 (together with $\alpha = 2/3$) and choosing matching values of the parameters ϵ, η and λ as explained in section 5.1.

sale discounts following a monetary shock. But the consumer preferences introduced to explain sales also give rise to strategic substitutability in sales decisions. Strategic substitutability is so strong that flexibility in sales imparts very little flexibility to the aggregate price level.

5.3 Justification for "sticky" normal prices

The previous analysis treated p_N as fixed, and s and p_S as completely flexible. In reality, there may be costs of readjusting s and p_S, but this paper shows that even without such costs, the possibility of continuously adjusting sales decisions has only a small impact on the real effects of monetary policy. Thus stickiness in p_N suffices to explain why monetary policy has real effects.

Recent micro evidence on price setting has highlighted the relative stickiness of so-called "reference" prices (Eichenbaum, Jaimovich and Rebelo, 2008), which include the "normal" prices of this paper. The model developed here is consistent with the finding of sticky reference prices, and moreover, in the setting of the model, it makes sense that the reference/normal price is relatively sticky. Three arguments are offered in support of this claim: (i) in the context of the model, the extra gain from adjusting the normal price after a firm has optimally chosen s and p_S is only 14% of the corresponding gain found in a standard model from adjusting a firm's single price; (ii) by adjusting its sales fraction s, a firm already reaps most of the overall benefits from price adjustment; and (iii) after adjusting s, the total gains from repeatedly adjusting the normal price (which is used 92.6% of the time in the baseline calibration) are actually very close to the total gains obtained from optimally setting the price used only when a good is on sale (that is, only 7.4% of the time).

These results build on the following proposition:

Proposition 4 *Consider arbitrary distributions of p_N and p_S around their flexible-price equilibrium values from section 4. Suppose all firms optimally choose their sales fraction s according to the first equation in [5.1].*

(i) Nominal marginal cost X is the same for all firms irrespective of their individual prices p_S and p_N, and moreover, $\mathsf{X} = \mathsf{P}_B$.

(ii) The total quantity sold Q is the same for all firms irrespective of their individual prices p_S and p_N.

(iii) If p_S^* and p_N^* denote the log-deviations of the desired sale and normal prices then $\mathsf{p}_S^* = \mathsf{p}_N^* = \mathsf{X}$.

(iv) A second-order approximation of the gains (expressed as a fraction of steady-state total revenue) from adjusting the sale and normal prices from p_S and p_N to p_S^* and p_N^* is:

$$\text{Gains} = \frac{1}{2}\bar{s}\frac{\bar{q}_S}{\bar{Q}}\bar{x}\left(\bar{\zeta}_S - \frac{(\eta-\epsilon)^2\lambda(1-\lambda)\bar{\upsilon}_S(\bar{\mu}_S - 1)}{(\lambda + (1-\lambda)\bar{\upsilon}_S)^2}\right)(\mathsf{p}_S - \mathsf{X})^2 + \frac{1}{2}(1-\bar{s})\frac{\bar{q}_N}{\bar{Q}}\bar{x}\left(\bar{\zeta}_N - \frac{(\eta-\epsilon)^2\lambda(1-\lambda)\bar{\upsilon}_N(\bar{\mu}_N - 1)}{(\lambda + (1-\lambda)\bar{\upsilon}_N)^2}\right)(\mathsf{p}_N - \mathsf{X})^2 \quad [5.5]$$

Proof See appendix A.8. ■

Corollary *If both s and p_S are optimally chosen, so $\mathsf{p}_S = \mathsf{p}_S^* = \mathsf{P}_B = \mathsf{X}$, then the gain from adjusting the normal price p_N to p_N^* is:*

$$\text{Gain} = \frac{1}{2}(1-\bar{s})\frac{\bar{q}_N}{\bar{Q}}\bar{x}\left(\bar{\zeta}_N - \frac{(\eta-\epsilon)^2\lambda(1-\lambda)\bar{\upsilon}_N(\bar{\mu}_N - 1)}{(\lambda + (1-\lambda)\bar{\upsilon}_N)^2}\right)(\mathsf{p}_N - \mathsf{X})^2 \quad [5.6]$$

□

The proposition considers the implications of firms optimally adjusting s, while the corollary also supposes that p_S is chosen to maximize profits.

Proposition 4 shows that the optimal choice of the sales fraction s already implies an optimal choice of total quantity sold, in the sense that if a firm were also to adjust optimally either its normal price or its sale price (or both) then this would make no difference to its total quantity sold. The implication is that most of the gains from price adjustment are already exhausted by choosing the sales fraction optimally. Quantitatively, the size of any remaining gains from changing the sale and normal prices themselves is assessed using the formula [5.5] set out in the proposition.

To understand the differences introduced by sales when compared to standard analyses of sticky prices, the expression in [5.5] for the gains in profits is contrasted with that which obtains in a model with entirely standard Dixit-Stiglitz preferences, and thus a constant price-elasticity demand function and a one-price equilibrium, but which is otherwise identical. As is shown in appendix A.11, the gain in profits (also expressed as a fraction of steady-state total revenue) from a firm adjusting its single price in such a model is:

$$\text{Gain} = \frac{1}{2}\varepsilon(1+\varepsilon\gamma)\bar{x}\left(\mathsf{p} - \left(\mathsf{P} + \frac{1}{1+\varepsilon\gamma}\mathsf{x}\right)\right)^2 , \qquad [5.7]$$

where ε is the constant price elasticity of demand and γ is the elasticity of marginal cost with respect to quantity produced. With the production function [5.4], $\gamma = (1-\alpha)/\alpha$.

There are two key differences between the profit gains [5.5] and [5.7]. Quantitatively, the most important difference corresponds to the term $1+\varepsilon\gamma$, which appears only in [5.7]. This represents the gain from producing the optimal total quantity, which in a standard model can be achieved only through a price change. But as Proposition 4 shows, the gain from producing the optimal total quantity automatically accrues to a firm that is free to adjust its sales fraction.

The second reason for smaller gains from full price adjustment in the sales model relative to a standard model is that with a demand function consistent with sales in equilibrium, the price elasticity is decreasing in price. Thus if a price is too high then a firm's desired markup also increases, and vice versa if its price is too low. The bracketed terms in [5.5] multiplying the deviations of prices are smaller than the respective price elasticities $\bar{\zeta}_S$ and $\bar{\zeta}_N$ since the terms being subtracted from these elasticities are unambiguously positive. In contrast, in [5.7], the price deviation is multiplied simply by the price elasticity ε.

The gains from full price adjustment in the sales model are compared to those in a standard one-price model where firms are faced with the same shocks, even though a one-price model would require much larger shocks to generate the observed magnitude of price changes. The difference in the size of adjustment costs needed to rule out a flexible-price equilibrium is computed using the calibration from section 5.1 and the expressions in [5.5] and [5.7]. In the latter for the one-price model, the constant price elasticity of demand ε is chosen to generate a markup equal to the average markup found in the calibrated model of sales.[10] With an elasticity of output with respect to hours of $\alpha = 2/3$, the implied value of γ is 0.5. In the sales model, the adjustment cost needed to dissuade

[10]The bracketed terms in [5.5] are multiplied by $\bar{s}\bar{q}_S/\bar{Q}$ and $(1-\bar{s})\bar{q}_N/\bar{Q}$, which weight them according to the quantities sold at the two prices. For the baseline calibration, $\mu_S = 1.09$, $\mu_N = 1.47$, and $\bar{s}\bar{q}_S/\bar{Q} = 0.28$, which yield an average markup of 1.36. With Dixit-Stiglitz preferences, the optimal markup is $\varepsilon/(\varepsilon-1)$, so $\varepsilon = 3.77$.

a given firm from changing both its sale and normal prices is only 27% of the cost that justifies a firm not changing its single price in a standard one-price model.

A similar exercise is performed assuming that both p_S and s are optimally chosen in the model with sales, which corresponds to comparing the gains implied by [5.6] and [5.7]. This exercise reveals that the adjustment cost needed to dissuade a firm from adjusting its normal price p_N is only 14% of the cost needed to deter adjustment of a firm's single price in a standard one-price model. This constitutes approximately half of the total gains from changing both p_S and p_N, which shows that the coefficients of the deviations of p_S and p_N in [5.5] are approximately of the same magnitude.

Even though the coefficients in [5.5] are very close, at a given moment, the gains from optimally adjusting p_S are approximately 12 times larger than those from adjusting p_N. This is because the price elasticity is much higher, the profit margin is narrower, and the quantity sold significantly larger at p_S than at p_N. These factors make the profit function much more concave at p_S, so a given percentage deviation from optimality of the sale price is much more costly than a similar deviation of the normal price. The overall importance of adjusting p_S and p_N turns out to be similar because s is around 12 times smaller than $(1 - s)$. So at a given moment, if there is no intrinsic difference between the cost of optimally adjusting a normal price versus a sale price, a firm would strongly prefer to reoptimize its sale price.

A further relevant consideration is that in the model with sales, the magnitude of a firm's desired changes to its normal price in response to shocks is significantly smaller than the size of individual price changes typically observed, which correspond to shifts between the normal and sale prices. Therefore, although large price changes are often observed, full reoptimization of normal prices would require only minor adjustments, and hence only small losses if those adjustments are not made. This means that reoptimization of the normal price falls within the remit of the literature in macroeconomics which seeks to justify why firms do not always take small profit-enhancing actions, as first analysed in Mankiw (1985) and Akerlof and Yellen (1985).[11]

It may seem paradoxical that firms are able to extract most of the gains from changing price simply by varying their sales fractions, but at the same time choose to do so sparingly in response to a monetary shock. This apparent contradiction is resolved by noting the reason for the small response of the sales fraction is not its lack of efficacy for an individual firm, but that other firms also react to common shocks in the same way, and sales have been shown to be strategic substitutes.

6 Flexible wages

This section considers the model of sales with fully flexible wages. Here, households' first-order condition [3.1] for hours of labour supplied always holds. Since every household faces the same price level P and money wage W, this implies:

$$\frac{\nu_h(H)}{u_c(Y)} = \frac{W}{P} \,. \qquad [6.1]$$

[11] Direct empirical evidence on the costs of reoptimizing prices is presented in Levy, Bergen, Dutta and Venable (1997) and Zbaracki, Ritson, Levy, Dutta and Bergen (2004).

The remainder of the model is as described in section 5.

Obtaining the effects of a monetary shock now requires calibrating the utility function [2.1]. The main issue is to avoid the counterfactual prediction that real wages fluctuate by more than output. Thus a lower bound for the real effects of a monetary shock is found by choosing a utility function which implies real wages and output move one-for-one. This is done by adopting the conventional specification of log utility in consumption C and linear disutility in hours worked H:[12]

$$u(C) = \log C \ , \quad \nu(H) = aH \ .$$

As in section 5, the economy is subjected to a money-supply shock. The sales fraction s, the sale price p_S, and wage W are optimally readjusted. Only the normal price p_N is predetermined. The resulting elasticities of output and the price level with respect to the monetary surprise are:

$$\frac{d \log Y}{d \log M} = 0.685 \ , \quad \frac{d \log P}{d \log M} = 0.315 \ .$$

These results show the strength of strategic substitutability in sales decisions. Even though wages are fully flexible (and fluctuate by more than in the data), and firms face no impediments to adjusting either their sale price or sales fraction, monetary policy has large real effects.

7 Dynamics

This section extends the previous analysis to a fully dynamic environment, where firms' normal prices are readjusted, but not continuously so. There is a tractable dynamic version of the sales model and this section derives the resulting Phillips curve, which is easily embedded into any broader DSGE model. While the presence of sales in the model adds an extra feature when compared to the standard New Keynesian Phillips curve, quantitatively the difference is not large. The results are thus in line with the findings of section 5.

7.1 Staggered adjustment of normal prices

The model developed here continues to allow firms costlessly to vary their sales fractions and sale prices, but now a firm can also choose a new normal price at random times, as in the Calvo (1983) pricing model. It is important to stress that the Calvo pricing assumption is used only for changes of the normal price; a firm continues to have complete discretion to switch its individual price without cost between the normal and sale price at any given moment, and to change the sale price itself.

The assumption of Calvo adjustment times for the normal price is made for simplicity. Of course the choice of an alternative model of price adjustment, for example, state-dependent pricing, would affect the results in its own right. But there is no obvious reason to believe that the interaction of different models with firms' optimal sales decisions would significantly affect the results (unless a

[12]This reflects standard practice in the real business cycle literature following Hansen (1985), and is also a specification employed in recent theoretical work on pricing, such as Golosov and Lucas (2007) and Kehoe and Midrigan (2007).

model yielded the counterfactual prediction that normal prices are continuously adjusted, thus rendering the sales margin redundant). This is because Proposition 4 implies that the profit-maximizing normal price is a function only of the aggregate state of the economy, and furthermore, a firm's optimal sales decisions depend only on its own current normal price and the aggregate state of the economy.

In every time period, each firm has a probability $1-\phi_p$ of receiving an opportunity to adjust its normal price. Consider a firm that receives such an opportunity at time t. The new normal price it selects is referred to as its reset price, and is denoted by $R_{N,t}$. All firms that choose new normal prices at the same time choose the same reset price. In any time period, each firm's optimal sales decisions will in principle depend on its current normal price, and so on its last adjustment time. Denote by $s_{\ell,t}$ and $p_{S,\ell,t}$ the optimal sales fraction and sale price for a firm at time t that last changed its normal price ℓ periods ago. The reset price $R_{N,t}$ is chosen to maximize the present value of the firm, and asset markets are assumed to be complete, with $\mathscr{A}_{t+\ell|t}$ denoting the asset-pricing kernel for state-contingent monetary payoffs received at time $t+\ell$ (relative to the conditional probability of each state occurring as of time t):

$$\max_{R_{N,t}} \sum_{\ell=0}^{\infty} \phi_p^\ell \mathbb{E}_t \left[\mathscr{A}_{t+\ell|t} \left\{ \begin{array}{c} s_{\ell,t+\ell} p_{S,\ell,t+\ell} \mathscr{D}(p_{S,\ell,t+\ell}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) + (1-s_{\ell,t+\ell}) R_{N,t} \mathscr{D}(R_{N,t}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) \\ -\mathscr{C}\Big(s_{\ell,t+\ell} \mathscr{D}(p_{S,\ell,t+\ell}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) + (1-s_{\ell,t+\ell}) \mathscr{D}(R_{N,t}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}); W_{t+\ell} \Big) \end{array} \right\} \right] \tag{7.1}$$

The first-order condition for the profit-maximizing reset price $R_{N,t}$ is given by:

$$\sum_{\ell=0}^{\infty} \phi_p^\ell \mathbb{E}_t \left[(1-s_{\ell,t+\ell}) \mathfrak{V}_{t+\ell|t} \left\{ \frac{R_{N,t}}{P_{t+\ell}} - \mu(R_{N,t}; P_{B,t+\ell}) \frac{\mathscr{C}'(Q_{\ell,t+\ell}; W_{t+\ell})}{P_{t+\ell}} \right\} \right] = 0 , \tag{7.2}$$

$$\text{where} \quad \mathfrak{V}_{t+\ell|t} \equiv \frac{(\zeta(R_{N,t}; P_{B,t+\ell}) - 1) \mathscr{D}(R_{N,t}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) P_{t+\ell} \mathscr{A}_{t+\ell|t}}{P_t} .$$

Using [3.3], [3.4] and [3.8], an expression for the aggregate price level P_t is:

$$P_t = \left((1-\phi_p) \sum_{\ell=0}^{\infty} \phi_p^\ell \left\{ \begin{array}{c} s_{\ell,t}(\lambda + (1-\lambda) v(p_{S,\ell,t}, P_{B,t})) p_{S,\ell,t}^{1-\epsilon} \\ +(1-s_{\ell,t})(\lambda + (1-\lambda) v(R_{N,t-\ell}, P_{B,t})) R_{N,t-\ell}^{1-\epsilon} \end{array} \right\} \right)^{\frac{1}{1-\epsilon}} , \tag{7.3}$$

and the bargain hunters' price index $P_{B,t}$ from [3.4] is obtained accordingly.

The profit-maximizing sales fractions $s_{\ell,t}$ and sale prices $p_{S,\ell,t}$ are determined as in [5.1]:

$$\frac{p_{S,\ell,t} q_{S,\ell,t} - R_{N,t-\ell} q_{N,\ell,t}}{q_{S,\ell,t} - q_{N,\ell,t}} = X_{\ell,t} , \quad p_{S,\ell,t} = \mu(p_{S,\ell,t}; P_{B,t}) X_{\ell,t} , \tag{7.4}$$

where $q_{S,\ell,t}$ and $q_{N,\ell,t}$ are the quantities sold by a firm at its sale and normal prices if it changed its normal price ℓ periods ago, and $X_{\ell,t}$ is nominal marginal cost for such a firm.

7.2 A Phillips curve with sales

To study the dynamic implications of the sales model, it is helpful to derive a Phillips curve that can be compared with those from standard models with Calvo pricing. It turns out that the dynamic model with sales also yields a simple Phillips curve.

Theorem 3 *Suppose each firm determines its profit-maximizing reset price $R_{N,t}$ according to equation [7.2] and its profit-maximizing sales fraction and sale price using [7.4]. Let $\pi_t \equiv P_t/P_{t-1}$ be the inflation rate for the aggregate price level [7.3]. Log-linearizing around the flexible-price equilibrium of section 4 with zero inflation yields an equation for the reset price:*

$$\mathsf{R}_{N,t} = (1-\beta\phi_p)\sum_{\ell=0}^{\infty}(\beta\phi_p)^{\ell}\mathbb{E}_t\mathsf{X}_{t+\ell} \ ,$$

where X_t is the common level of nominal marginal cost that results from firms optimizing over their sales fractions as shown in Proposition 4, and β is the discount factor. The implied Phillips curve linking inflation π_t and real marginal cost x_t is:

$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \frac{1}{1-\psi}\left(\kappa\mathsf{x}_t + \psi\left(\Delta\mathsf{x}_t - \beta\mathbb{E}_t\Delta\mathsf{x}_{t+1}\right)\right) \ , \qquad [7.5]$$

where the parameter ψ is defined as follows:

$$\psi \equiv \left(\left(1-\frac{\partial\log P_B}{\partial\log P_S}\right)\frac{\partial\log P}{\partial s} + \frac{\partial\log P}{\partial\log P_S}\frac{\partial\log P_B}{\partial s}\right)\Big/\frac{\partial\log P_B}{\partial s} \ ,$$

and $\kappa \equiv \left((1-\phi_p)(1-\beta\phi_p)\right)/\phi_p$. By solving forwards, inflation can also be expressed as:

$$\pi_t = \frac{\kappa}{1-\psi}\sum_{\ell=0}^{\infty}\beta^{\ell}\mathbb{E}_t\mathsf{x}_{t+\ell} + \frac{\psi}{1-\psi}\Delta\mathsf{x}_t \ . \qquad [7.6]$$

Proof See appendix A.9. ■

The Phillips curve with sales [7.5] would reduce to the standard New Keynesian Phillips curve $\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \kappa\mathsf{x}_t$ were it the case that $\psi = 0$, but ψ is always positive in the model with sales. When $\psi \to 1$, the economy behaves as though all prices were fully flexible. The condition $\psi < 1$ is equivalent to:

$$-\frac{\partial\log P}{\partial s}\Big/\left(1-\frac{\partial\log P}{\partial\log P_S}\right) < -\frac{\partial\log P_B}{\partial s}\Big/\left(1-\frac{\partial\log P_B}{\partial\log P_S}\right) \ . \qquad [7.7]$$

First note that the elasticity of P_B with respect to P_S is always larger than the corresponding elasticity of P because bargain hunters buy more goods at sale prices, hence the denominator of the right-hand side is smaller than that of the left-hand side. Second, the numerator on the right-hand side is larger than on the left-hand side as long as an increase in the number of sales offered by firms benefits bargain hunters more than loyal consumers, which is intuitively plausible and true in the baseline calibration, although it cannot hold for all possible parameters. Because the first statement is always true, the second condition is sufficient but not necessary for [7.7] to hold. In the baseline calibration, ψ is 0.26.

The effect of a positive value of ψ is to increase the response of inflation to real marginal cost when compared to a standard model where the probability of changing price is the same as the probability of adjusting a normal price in the sales model. This is best seen by looking at the solved-forwards version of the Phillips curve in [7.6], where there are two distinct differences relative to the solved-forwards version of the New Keynesian Phillips curve: $\pi_t = \kappa \sum_{\ell=0}^{\infty} \beta^{\ell} \mathbb{E}_t \mathsf{X}_{t+\ell}$. The first is a scaling of the coefficient multiplying expected real marginal costs, which is isomorphic to an increase in the probability of price adjustment. The second is the presence of the term in the growth rate of real marginal cost $\Delta \mathsf{x}_t$, which is linked to the possibility of varying sales in each period. It is subsequently shown how this term affects the dynamics of output and the price level.

7.3 A DSGE model with sales

This section embeds sales into a calibrated dynamic stochastic general equilibrium model with staggered adjustment of normal prices and wages.

Household $\imath \in \mathscr{H}$'s lifetime utility function is given by:

$$U_t(\imath) = \sum_{\ell=0}^{\infty} \beta^{\ell} \mathbb{E}_t \left[\upsilon(C_{t+\ell}(\imath), m_{t+\ell}(\imath)) - \nu(H_{t+\ell}(\imath)) \right] . \quad [7.8]$$

The utility function $\upsilon(C, m)$ is differentiable, strictly increasing and strictly concave in both C and m; $\nu(H)$ is a differentiable, strictly increasing and convex function of H. Each household supplies a differentiated labour input. The parameter β is the subjective discount factor, which satisfies $0 < \beta < 1$.

Denote by $\mathcal{A}_{t+1}(\imath)$ household $\imath$'s portfolio of money-denominated Arrow-Debreu securities held between periods t and $t+1$. Household $\imath$'s period-t budget constraint is thus:

$$P_t C_t(\imath) + M_t(\imath) + \mathbb{E}_t \left[\mathscr{A}_{t+1|t} \mathcal{A}_{t+1}(\imath) \right] = W_t(\imath) H_t(\imath) + \mathfrak{D}_t + \mathfrak{T}_t + M_{t-1}(\imath) + \mathcal{A}_t(\imath) . \quad [7.9]$$

Households have equal initial financial wealth and all have the same expected lifetime income.

There are no arbitrage opportunities in financial markets, so the yield i_t on a one-period risk-free nominal bond satisfies:

$$1 + i_t = \left(\mathbb{E}_t \mathscr{A}_{t+1|t} \right)^{-1} . \quad [7.10]$$

Maximization of lifetime utility [7.8] subject to the sequence of budget constraints [7.9] implies the following first-order conditions for consumption $C_t(\imath)$ and real money balances $m_t(\imath)$:

$$\beta \frac{\upsilon_c(C_{t+1}(\imath), m_{t+1}(\imath))}{\upsilon_c(C_t(\imath), m_t(\imath))} = \mathscr{A}_{t+1|t} \frac{P_{t+1}}{P_t} , \quad [7.11a]$$

$$\frac{\upsilon_m(C_t(\imath), m_t(\imath))}{\upsilon_c(C_t(\imath), m_t(\imath))} = \frac{i_t}{1 + i_t} . \quad [7.11b]$$

Equation [7.11a] is the Euler equation for consumption across time and across states, with $\upsilon_c(C, m)$ denoting the marginal utility of consumption. The optimal tradeoff between consumption and holding money balances is given by [7.11b], with $\upsilon_m(C, m)$ denoting the marginal utility of real balances and $i_t/(1 + i_t)$ being the opportunity cost of holding money.

As in Erceg, Henderson and Levin (2000), firms hire differentiated labour inputs. So hours H in the production function [2.5] is now a composite labour input defined by the following Dixit-Stiglitz aggregator:

$$H \equiv \left(\int_{\mathscr{H}} H(\imath)^{\frac{\varsigma - 1}{\varsigma}} d\imath \right)^{\frac{\varsigma}{\varsigma - 1}} ,$$

where $H(\imath)$ is hours supplied by household $\imath \in \mathscr{H}$ to a given firm, and ς is the elasticity of substitution between labour types. It is assumed that $\varsigma > 1$, and firms are price takers in the markets for labour inputs. The money wage received by labour input $\imath$ is $W(\imath)$. The minimum cost of hiring one unit of the composite labour input H is denoted by W, and this is now the relevant wage index appearing in firms' cost function [2.6]. This wage index is given by:

$$W \equiv \left(\int_{\mathscr{H}} W(\imath)^{1-\varsigma} d\imath \right)^{\frac{1}{1-\varsigma}} , \qquad [7.12]$$

and the cost-minimizing labour demand functions are:

$$H(\imath) = \left(\frac{W(\imath)}{W} \right)^{-\varsigma} H \, . \qquad [7.13]$$

Suppose that households have a probability $1 - \phi_w$ of being able to adjust their money wage in any given time period. Since households have equal initial wealth and expected lifetime income, and as asset markets are complete, and as utility [7.8] is additively separable between hours and consumption, households are fully insured and hence have equal consumption and money balances in equilibrium. As before, consumption is the only source of expenditure, so goods market equilibrium requires $C_t = Y_t$. Thus by using [7.10], [7.11a] and [7.11b], the following intertemporal IS equation and implicit money-demand function are obtained:

$$\beta(1 + i_t)\mathbb{E}_t \left[\frac{\upsilon_c(Y_{t+1}, m_{t+1})}{\upsilon_c(Y_t, m_t)} \frac{1}{\pi_{t+1}} \right] = 1 \, , \quad \frac{\upsilon_m(Y_t, m_t)}{\upsilon_c(Y_t, m_t)} = \frac{i_t}{1 + i_t} \, . \qquad [7.14]$$

As households are selected to update their wages at random, as they enjoy the same consumption, and as they face the same demand function for their labour services, all households setting a new wage at time t choose the same wage. This common wage is referred to as the reset wage, and is denoted by $R_{W,t}$. It is chosen to maximize expected utility over the lifetime of the wage subject to the labour demand function [7.13]. As shown by Erceg, Henderson and Levin (2000), the first-order condition for this maximization problem is:

$$\sum_{\ell=0}^{\infty} (\beta\phi_w)^\ell \mathbb{E}_t \left[\frac{W_{t+\ell}^{\varsigma} H_{t+\ell} \upsilon_c(Y_{t+\ell}, m_{t+\ell})}{\upsilon_c(Y_t, m_t)} \left\{ \frac{R_{W,t}}{P_{t+\ell}} - \frac{\varsigma}{\varsigma - 1} \frac{\nu_h \left(R_{W,t}^{-\varsigma} W_{t+\ell}^{\varsigma} H_{t+\ell} \right)}{\upsilon_c(Y_{t+\ell}, m_{t+\ell})} \right\} \right] = 0 \, . \qquad [7.15]$$

The wage index W_t in [7.12] then evolves according to:

$$W_t = \left((1 - \phi_w) \sum_{\ell=0}^{\infty} \phi_w^\ell R_{W,t-\ell}^{1-\varsigma} \right)^{\frac{1}{1-\varsigma}} . \qquad [7.16]$$

7.4 Dynamic calibration

This section presents the calibration of the DSGE model described above. All the values of the calibrated parameters are listed in Table 3.

One period corresponds to one month. The discount factor β is chosen to yield a 3% annual real interest rate, the intertemporal elasticity of substitution in consumption σ_c is chosen to match a coefficient of relative risk aversion of 3, and the Frisch elasticity of labour supply σ_h is set to 0.7, which lies in the range of estimates found in the literature. The elasticity of money demand with respect to income ϑ_y, the interest semi-elasticity ϑ_i, and the real balance effect of money on consumption ϑ_m are taken from Woodford (2003), making the conversion from a quarterly to a monthly calibration.

Table 3: *Dynamic calibration*

Description	Parameter	Value
Preference parameters		
Subjective discount factor	β	0.9975
Intertemporal elasticity of substitution in consumption	σ_c	0.333
Frisch elasticity of labour supply	σ_h	0.7
Income elasticity of money demand	ϑ_y	1.0*
Interest semi-elasticity of money demand	ϑ_i	84*
Real balance effect on consumption	ϑ_m	0.0067*
Technology parameters		
Elasticity of output with respect to hours	α	0.667
Elasticity of marginal cost with respect to output	γ	0.5
Elasticity of substitution between differentiated labour inputs	ς	20†
Nominal rigidities		
Probability of stickiness of "normal" prices	ϕ_p	0.889§
Probability of wage stickiness	ϕ_w	0.889

Notes: Monthly calibration.
* *Source*: Woodford (2003)
† *Source*: Christiano, Eichenbaum and Evans (2005)
§ *Source*: Nakamura and Steinsson (2007)

The elasticity of output with respect to hours α is chosen to match a labour share of 2/3. With the specification of the production function in [5.4], this implies an elasticity of marginal cost with respect to output of $\gamma = (1 - \alpha)/\alpha$. So $\alpha = 2/3$ yields $\gamma = 0.5$. The elasticity of substitution between labour inputs ς is taken from Christiano, Eichenbaum and Evans (2005). The probability of stickiness of the normal price ϕ_p is set to match an average price-spell duration of 9 months, which is taken from Nakamura and Steinsson (2007). The same number is used for the probability of wage stickiness ϕ_w, as evidence shows that most, but not all, wages are adjusted annually.

The model is analysed under different assumptions about monetary policy. First, a first-order autoregressive process for money growth is considered:

$$\frac{M_t}{M_{t-1}} = \left(\frac{M_{t-1}}{M_{t-2}}\right)^{\varphi_m} \exp(\mathfrak{e}_t) \ , \quad \mathfrak{e}_t \sim \text{i.i.d.}(0, \mathfrak{v}^2) \ . \tag{7.17a}$$

The persistence parameter φ_m is chosen to match the empirical first-order autocorrelation coefficient of M1 growth in the U.S. from 1979:8 to 1996:12.

Second, the case of a monetary policy rule with feedback from the state of the economy is considered. A Taylor rule with interest-rate smoothing is the most popular specification for this:

$$1 + i_t = (1 + i_{t-1})^{\varphi_i} \left((1 + \bar{i}) \left(\frac{\pi_t}{\bar{\pi}}\right)^{\varphi_\pi} \left(\frac{Y_t}{\bar{Y}}\right)^{\varphi_y}\right)^{1-\varphi_i} \exp(\mathfrak{e}_t) \ , \quad \mathfrak{e}_t \sim \text{i.i.d.}(0, \mathfrak{v}^2) \ , \tag{7.17b}$$

where φ_π is the interest-rate response to inflation, φ_y is the response to output (or the output gap), and φ_i is the interest-rate smoothing parameter. The Taylor rule parameters are taken from the baseline estimates for the Volcker–Greenspan period in Clarida, Galí and Gertler (2000), which is 1979:8–1996:12 (the same sample period as was used for the money-supply growth specification).

Table 4: *Parameters used for the monetary policy experiments*

Description	Parameter	Value
Exogenous path for growth of the money supply		
First-order serial correlation of the money-supply growth rate	φ_m	0.6^*
Taylor rule		
Response of interest rates to deviations of inflation from target	φ_π	$2.15^\dagger$
Response of interest rates to deviations of aggregate output from target	φ_y	$0.078^\dagger$
Degree of interest-rate smoothing	φ_i	$0.924^\dagger$

Notes: Monthly calibration.
* *Source*: Authors' calculations using data on M1 for the period 1979:8–1996:12. Series `M1SL` from Federal Reserve Economic Data (`http://research.stlouisfed.org/fred2`).
† *Source*: Clarida, Galí and Gertler (2000), converted from estimates based on quarterly data to a monthly calibration.

7.5 Dynamic simulations

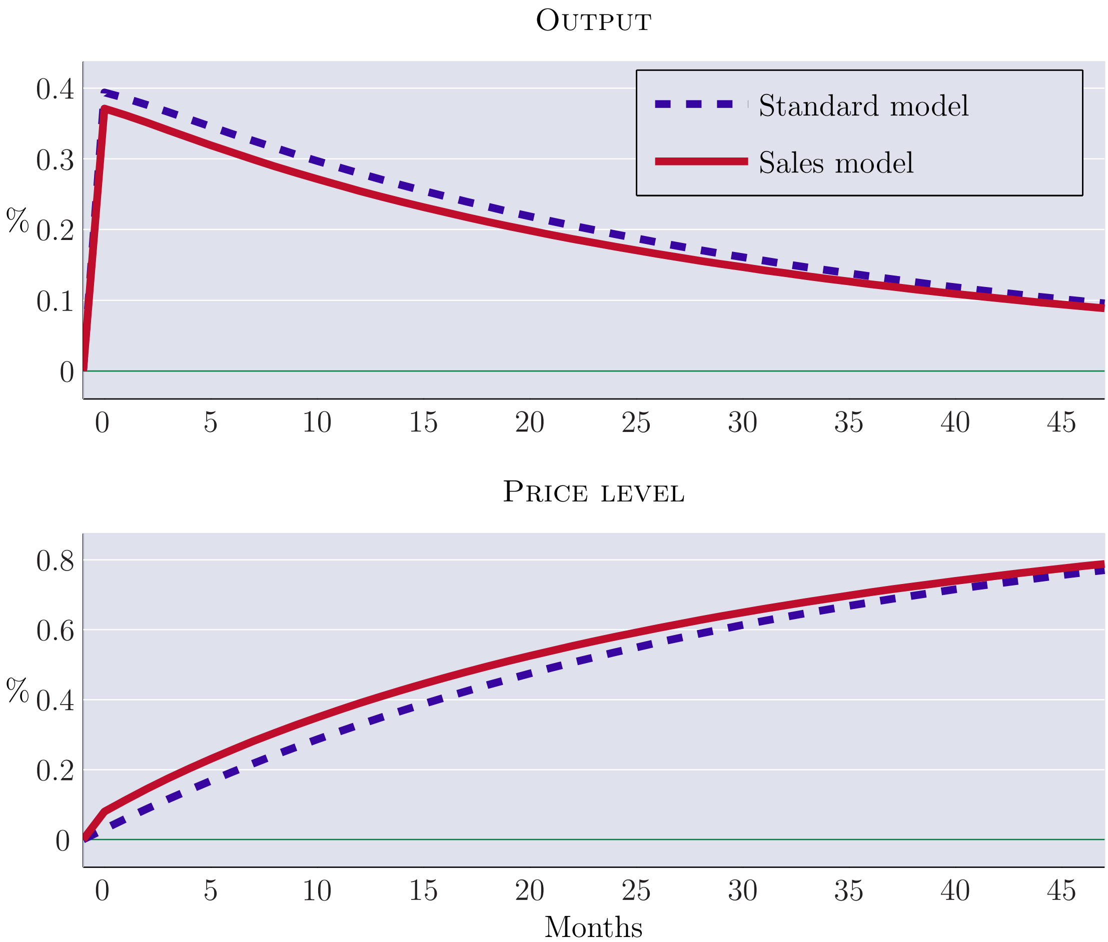
This section calculates the impulse responses of output and the price level to monetary policy shocks in the DSGE model with sales described in section 7.1 and section 7.3. These are compared to the corresponding impulse responses in a standard DSGE model, that is, one where consumers have regular Dixit-Stiglitz preferences and thus firms employ a one-price strategy, and price adjustment times are staggered according to the Calvo model. With Calvo pricing, a standard New Keynesian Phillips curve is obtained.[13] The latter model is set up so that it is otherwise identical to the DSGE model with sales.

[13]See appendix A.10 for details.

The calibrated parameters of the DSGE model with sales are given in Table 2 and Table 3. For the standard model without sales, the same parameter values from Table 3 are used, with the probability of price stickiness applying to a firm's single price, rather than to its normal price in the sales model. In place of parameters ϵ, η and λ, the standard model requires only a calibration of the constant price elasticity of demand ε. This is chosen to match the average markup found in the calibrated sales model.[14]

Impulse response functions are calculated for the two monetary policy experiments described in section 7.4 with parameters from Table 4: a persistent shock to money supply growth [7.17a]; and a shock to a Taylor rule with interest-rate smoothing [7.17b].

Figure 6: *Impulse responses to a persistent shock to money growth*



Notes: The specification of monetary policy used is equation [7.17a].

Figure 6 plots the impulse responses of aggregate output and the price level when money growth follows an AR(1) process in both the sales model and the standard model without sales. As in the analysis of section 5 with a fixed normal price, the real effects of monetary policy in the model with sales are large and very similar to those found in the standard model, in spite of firms' full freedom to adjust their sales decisions without cost. The ratio of the cumulated deviations of output in

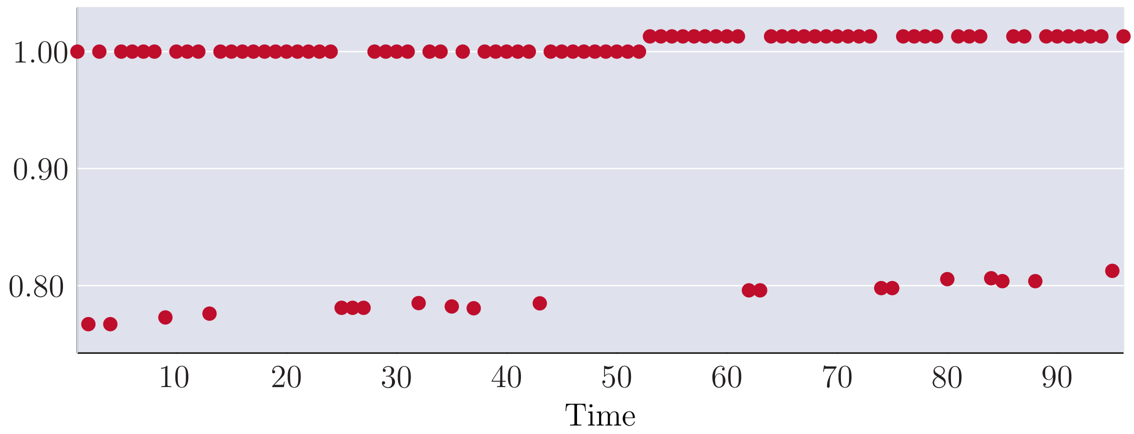
[14] See footnote 10 for details. The calculations lead to $\varepsilon = 3.77$.

the two models is 0.929. The response of the price level in the sales model shows a small jump immediately after the shock. This corresponds to the term Δx_t in the Phillips curve [7.6].

The impulse responses are not particularly sensitive to the calibrated parameters. Considering the same range of parameters as was done for the sensitivity analysis of section 5.2 leads to only small differences in these findings.

Figure 7 shows an example of an individual price path in the model with sales generated using the baseline calibration. The underlying stochastic process for the money supply is a random walk with drift. The behaviour depicted is qualitatively and quantitatively consistent with real-world examples of prices without needing to assume any idiosyncratic shocks are present.

Figure 7: *Theoretical price path implied by the model with sales*



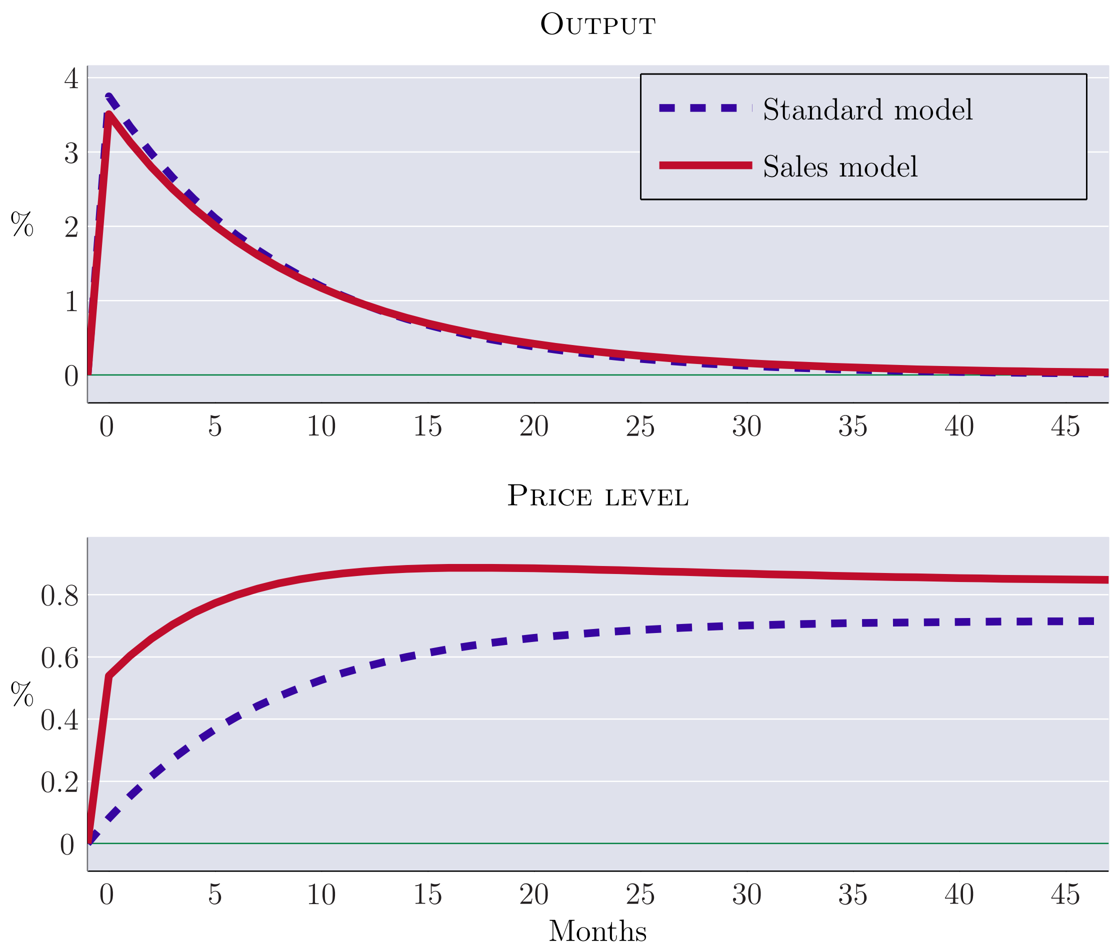
Notes: Generated using the baseline calibration of the DSGE model with sales and the money supply following a random walk with drift. The initial normal price is set to 1.

It is interesting to note from Figure 7 that the model can explain the coexistence of both small and large price changes for the same product in the presence of only macroeconomic shocks. Without any shocks at all, sales would still occur at a very similar frequency, but individual prices would switch between unchanging normal and sale prices.

When the central bank follows a Taylor rule, the reaction to monetary policy shocks is somewhat different, as is seen in Figure 8. The responses of output in the sales model and in the standard model are now virtually identical. But the responses of the price level are different. As before, the sales model features an initial jump in the price level. This is more marked than in the case of a shock to money growth. The difference in the price-level response diminishes over time, but does not vanish in the long run, converging to around 17% in the baseline calibration.

In essence, however, this finding is not in conflict with the those obtained when the money supply is exogenous. The addition of sales to the model affects the Phillips curve relationship, which determines how much inflation is generated for a given sequence of output gaps. The analysis in the case of exogenous money growth shows that sales lead to a slight reduction in the real effects of monetary policy. In the case of a Taylor rule, the effect of a monetary shock on output is approximately the same in both models, but cumulated inflation in the sales model is a little higher.

Figure 8: *Impulse responses to interest-rate shock with a Taylor rule*



Notes: The specification of monetary policy used is equation [7.17b].

8 Conclusions

For macroeconomists grappling with the welter of recent micro pricing evidence, one particularly puzzling feature is noteworthy: the large, frequent and short-lived price changes followed by prices returning exactly to their former levels. If price changes are driven purely by shocks then explaining this tendency requires a very special configuration of shocks. The model presented in this paper shows that just such pricing behaviour arises in equilibrium if firms face consumers with sufficiently different price sensitivities.

The model explains why firms choose a two-price distribution over time with a normal price and a sale price, and thus want to switch individual prices frequently between the two desired prices in their distribution. The two prices are themselves sensitive to shocks, but the magnitudes of optimizing adjustments to the normal and sale prices are dwarfed by the gap between the two. Furthermore, the impact on profits of not correcting deviations from optimality of the normal price at a particular moment is much smaller than the impact of not correcting similar percentage deviations of the sale price. So the apparent "puzzle" of why prices return to their former levels after a sale reduces to explaining why after a move from \$5.99 to \$4.49, a price returns to \$5.99

instead of \$6.02. But small costs of reoptimizing the normal price would explain firms' reluctance to make small profit-maximizing improvements in accordance with a well-established literature in macroeconomics.

One main message from the micro evidence is that the normal price is indeed considerably sticky, despite the significant flexibility of individual prices due to sales. Since the real effects of monetary policy depend on how sticky prices are, how should this evidence be interpreted? On the one hand, some would argue that temporary sales are orthogonal to monetary policy and thus such price changes should be ignored. On the other hand, others would argue that if decisions about temporary sales react to demand fluctuations, they should also react to monetary policy shocks to the extent that these shocks have an impact on aggregate demand.

The model proposed in this paper contains a rationale for sales, and therefore it can be used to understand the implications for monetary policy analysis of flexibility in sales decisions alongside stickiness in the normal price. In the model, sales are occurring for a reason, but firms potentially have an incentive to vary sales in response to shocks of all kinds, including those to monetary policy. However, it turns out that firms barely adjust sales in response to monetary policy shocks because the rationale for sales also implies that sales are strategic substitutes, that is, firms have a strong incentive to increase sales when others decrease them. While a firm may adjust sales substantially in response to shocks affecting only itself, it will not do so in the case of shocks affecting all firms.

The findings of this paper indicate that in a world with both sticky normal prices and flexible sales, it is predominantly stickiness in the normal price that matters so far as monetary policy analysis is concerned. Arriving at this conclusion requires a careful modelling of the reasons why sales occur. Thus the results highlight the importance for macroeconomics of understanding what lies behind firms' pricing decisions.

References

Akerlof, G. A. and Yellen, J. L. (1985), "Can small deviations from rationality make significant differences to economic equilibria?", *American Economic Review*, **75(4)**:708–720 (September). 26

Bils, M. and Klenow, P. J. (2004), "Some evidence on the importance of sticky prices", *Journal of Political Economy*, **112(5)**:947–985 (October). 1

Calvo, G. A. (1983), "Staggered prices in a utility-maximizing framework", *Journal of Monetary Economics*, **12(3)**:383–398 (September). 27

Chakravarthi, N., Neslin, S. and Sen, S. (1996), "Promotional elasticities and category characteristics", *Journal of Marketing*, **60(2)**:17–30 (April). 21

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (2005), "Nominal rigidities and the dynamic effects of a shock to monetary policy", *Journal of Political Economy*, **113(1)**:1–45 (February). 32

Clarida, R., Galí, J. and Gertler, M. (2000), "Monetary policy rules and macroeconomic stability: Evidence and some theory", *Quarterly Journal of Economics*, **115(1)**:147–180 (February). 33

Dhyne, E., Álvarez, L. J., Le Bihan, H., Veronese, G., Dias, D., Hoffmann, J., Jonker, N., Lünnemann, P., Rumler, F. and Vilmunen, J. (2006), "Price changes in the euro area and the United States: Some facts from individual consumer price data", *Journal of Economic Perspectives*, **20(2)**:171–192 (Spring). 1

Eichenbaum, M., Jaimovich, N. and Rebelo, S. (2008), "Reference prices and nominal rigidities", Working paper 13829, National Bureau of Economic Research. 1, 24

Erceg, C. J., Henderson, D. W. and Levin, A. T. (2000), "Optimal monetary policy with staggered wage and price contracts", *Journal of Monetary Economics*, **46(2)**:281–313 (October). 30, 31

Goldberg, P. and Hellerstein, R. (2007), "A framework for identifying the sources of local-currency price stability with an empirical application", Working paper 13183, National Bureau of Economic Research. 1

Golosov, M. and Lucas, R. E., Jr. (2007), "Menu costs and Phillips curves", *Journal of Political Economy*, **115(2)**:171–199. 27

Hansen, G. D. (1985), "Indivisible labor and the business cycle", *Journal of Monetary Economics*, **16(3)**:309–327 (November). 27

Hosken, D. and Reiffen, D. (2004), "Patterns of retail price variation", *RAND Journal of Economics*, **35(1)**:128–146 (Spring). 1

Kehoe, P. and Midrigan, V. (2007), "Sales and the real effects of monetary policy", Working paper 652, Federal Reserve Bank of Minneapolis. 1, 4, 27

Klenow, P. J. and Kryvtsov, O. (2005), "State-dependent or time-dependent pricing: Does it matter for recent U.S. inflation?", Working paper 11043, National Bureau of Economic Research. 1

Lazear, E. P. (1986), "Retail pricing and clearance sales", *American Economic Review*, **76(1)**:14–32 (March). 2

Levy, D., Bergen, M., Dutta, S. and Venable, R. (1997), "The magnitude of menu costs: Direct evidence from large U. S. supermarket chains", *Quarterly Journal of Economics*, **112(3)**:791–825 (August). 26

Mankiw, N. G. (1985), "Small menu costs and large business cycles: A macroeconomic model of monopoly", *Quarterly Journal of Economics*, **100(2)**:529–538 (May). 26

Nakamura, E. and Steinsson, J. (2007), "Five facts about prices: A reevaluation of menu cost models", Working paper, Harvard University. 1, 20, 21, 32

Salop, S. and Stiglitz, J. E. (1977), "Bargains and ripoffs: A model of monopolistically competitive price dispersion", *Review of Economic Studies*, **44(3)**:493–510 (October). 2

——— (1982), "The theory of sales: A simple model of equilibrium price dispersion with identical agents", *American Economic Review*, **72(5)**:1121–1130 (December). 2

Sobel, J. (1984), "The timing of sales", *Review of Economic Studies*, **51(3)**:353–368 (July). 2

Varian, H. (1980), "A model of sales", *American Economic Review*, **70(4)**:651–659 (September). 2

Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, New Jersey. 32, 64

Zbaracki, M. J., Ritson, M., Levy, D., Dutta, S. and Bergen, M. (2004), "Managerial and customer costs of price adjustment: Direct evidence from industrial markets", *Review of Economics and Statistics*, **86(2)**:514–533 (May). 26

A Technical appendix

A.1 Properties of the demand, total revenue and marginal revenue functions

The structure of household consumption preferences introduced in section 3.2 implies that firms face a demand curve $q = \mathscr{D}(p; P_B, \mathcal{E})$ of the form given in equation [3.8] at each shopping moment. It is easier to analyse the properties of this demand function — and the associated total and marginal revenue functions — by working with what can be thought of as the corresponding "relative" demand function $\mathcal{D}(\rho)$, defined by:

$$\mathcal{D}(\rho) \equiv \lambda\rho^{-\epsilon} + (1-\lambda)\rho^{-\eta} \,, \tag{A.1.1}$$

which satisfies $\mathcal{D}(1) = 1$ for all choices of parameters. The relative demand function $\mathfrak{q} = \mathcal{D}(\rho)$ gives the "relative" quantity sold $\mathfrak{q}$ as a function of the relative price ρ, where relative price here means money price p relative to P_B, the bargain hunters' price index from [3.4], and relative quantity means quantity q sold relative to $\mathcal{E}/P_B^\epsilon$, where $\mathcal{E} = P^\epsilon Y$ is the measure of aggregate expenditure from [3.7]:

$$\rho \equiv \frac{p}{P_B} \,, \quad \mathfrak{q} \equiv \frac{P_B^\epsilon}{\mathcal{E}} q \,. \tag{A.1.2}$$

With these definitions, the original demand function [3.8] is stated in terms of the relative demand function [A.1.1] as follows:

$$\mathscr{D}(p; P_B, \mathcal{E}) = \frac{\mathcal{E}}{P_B^\epsilon} \mathcal{D}\left(\frac{p}{P_B}\right) \,. \tag{A.1.3}$$

The relative demand function [A.1.1] is a continuously differentiable function of ρ for all $\rho > 0$, and is strictly decreasing everywhere. Notice also that $\mathcal{D}(\rho) \to \infty$ as $\rho \to 0$, and $\mathcal{D}(\rho) \to 0$ as $\rho \to \infty$. By continuity and monotonicity, this implies that for every $\mathfrak{q} > 0$ there is a unique $\rho > 0$ such that $\mathfrak{q} = \mathcal{D}(\rho)$. Thus the inverse demand function $\mathcal{D}^{-1}(\mathfrak{q})$ is well defined for all $\mathfrak{q} > 0$, and is itself strictly decreasing and continuously differentiable. The total revenue function $\mathcal{R}(\mathfrak{q})$, defined in terms of the relative demand function, is:

$$\mathcal{R}(\mathfrak{q}) \equiv \mathfrak{q}\mathcal{D}^{-1}(\mathfrak{q}) \,. \tag{A.1.4}$$

Using the inverse demand function $\rho = \mathcal{D}^{-1}(\mathfrak{q})$, an equivalent expression for total revenue is $\mathcal{R}(\mathfrak{q}) = \mathcal{D}^{-1}(\mathfrak{q})\mathcal{D}\left(\mathcal{D}^{-1}(\mathfrak{q})\right)$, and by substituting the demand function from [A.1.1]:

$$\mathcal{R}(\mathfrak{q}) = \lambda\left(\mathcal{D}^{-1}(\mathfrak{q})\right)^{1-\epsilon} + (1-\lambda)\left(\mathcal{D}^{-1}(\mathfrak{q})\right)^{1-\eta} .$$

Since $\epsilon > 1$ and $\eta > 1$, and given the limiting behaviour of the demand function established above, it follows that $\mathcal{R}(\mathfrak{q}) \to \infty$ as $\mathfrak{q} \to \infty$ and $\mathcal{R}(\mathfrak{q}) \to 0$ as $\mathfrak{q} \to 0$. Hence, $\mathcal{R}(0) = 0$, and $\mathcal{R}(\mathfrak{q})$ is continuously differentiable for all $\mathfrak{q} \geq 0$.

Differentiating the total revenue function $\mathcal{R}(\mathfrak{q})$ from [A.1.4] using the inverse function theorem, and substituting demand function [A.1.1] yields an expression for marginal revenue:

$$\mathcal{R}'\left(\mathcal{D}(\rho)\right) = \left(\frac{(\epsilon-1)\lambda + (\eta-1)(1-\lambda)\rho^{\epsilon-\eta}}{\epsilon\lambda + \eta(1-\lambda)\rho^{\epsilon-\eta}}\right)\rho . \tag{A.1.5}$$

Because $\epsilon > 1$ and $\eta > 1$, it follows that $\mathcal{R}'(\mathfrak{q}) > 0$ for all $\mathfrak{q}$, so total revenue $\mathcal{R}(\mathfrak{q})$ is strictly increasing in $\mathfrak{q}$. Furthermore, because $\rho \to \infty$ as $\mathfrak{q} \to 0$, and $\rho \to 0$ as $\mathfrak{q} \to \infty$, [A.1.5] implies $\mathcal{R}'(\mathfrak{q}) \to \infty$ as $\mathfrak{q} \to 0$ and $\mathcal{R}'(\mathfrak{q}) \to 0$ as $\mathfrak{q} \to \infty$.

Just as [A.1.3] establishes the original demand function $\mathscr{D}(p; P_B, \mathcal{E})$ in [3.8] is connected to the relative demand function $\mathcal{D}(\rho)$ in [A.1.1], there are similar relations between the original inverse demand function $\mathscr{D}^{-1}(\mathfrak{q}; P_B, \mathcal{E})$, original total revenue $\mathscr{R}(q; P_B, \mathcal{E})$ and marginal revenue $\mathscr{R}'(q; P_B, \mathcal{E})$ functions, and their equivalents defined in terms of the relative demand function. The link between the inverse demand functions follows directly from [A.1.3]:

$$\mathscr{D}^{-1}(q; P_B, \mathcal{E}) = P_B\mathcal{D}^{-1}\left(\frac{qP_B^{\epsilon}}{\mathcal{E}}\right) . \tag{A.1.6}$$

Equations [3.9], [A.1.4] and [A.1.6] justify the following connections between the total revenue functions and their derivatives:

$$\mathscr{R}(q; P_B, \mathcal{E}) = P_B^{1-\epsilon}\mathcal{E}\mathcal{R}\left(\frac{qP_B^{\epsilon}}{\mathcal{E}}\right) , \quad \mathscr{R}'(q; P_B, \mathcal{E}) = P_B\mathcal{R}'\left(\frac{qP_B^{\epsilon}}{\mathcal{E}}\right) , \quad \mathscr{R}''(q; P_B, \mathcal{E}) = \frac{P_B^{1+\epsilon}}{\mathcal{E}}\mathcal{R}''\left(\frac{qP_B^{\epsilon}}{\mathcal{E}}\right) . \tag{A.1.7}$$

The next result examines the conditions under which marginal revenue $\mathcal{R}'(\mathfrak{q})$ is non-monotonic.

Lemma 1 *Consider the marginal revenue function $\mathcal{R}'(\mathfrak{q})$ obtained from [A.1.4] using the relative demand function [A.1.1], and suppose that $\eta > \epsilon > 1$.*

(i) If $\lambda = 0$ or $\lambda = 1$ or condition [4.3] does not hold then marginal revenue $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing for all $\mathfrak{q} \geq 0$.

(ii) If $0 < \lambda < 1$ and ϵ and η satisfy condition [4.3] then there exist $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$ such that $0 < \underline{\mathfrak{q}} < \overline{\mathfrak{q}} < \infty$ and where $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing between 0 and $\underline{\mathfrak{q}}$ and above $\overline{\mathfrak{q}}$, and strictly increasing between $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$.

Proof (i) If $\lambda = 0$ then it follows from [A.1.5] that $\mathcal{R}'(\mathfrak{q}) = ((\eta-1)/\eta)\mathcal{D}^{-1}(\mathfrak{q})$, and if $\lambda = 1$ that $\mathcal{R}'(\mathfrak{q}) = ((\epsilon-1)/\epsilon)\mathcal{D}^{-1}(\mathfrak{q})$. Since the inverse demand function $\mathcal{D}^{-1}(\mathfrak{q})$ is strictly decreasing, then marginal revenue must also be so in these cases.

(ii) In what follows, assume $0 < \lambda < 1$. Differentiate [A.1.5] to obtain:

$$\mathcal{D}'(\rho)\mathcal{R}''\left(\mathcal{D}(\rho)\right) = \frac{\eta(\eta-1)\left(\frac{1-\lambda}{\lambda}\rho^{\epsilon-\eta}\right)^2 - \left((\eta-\epsilon)^2 - \eta(\epsilon-1) - \epsilon(\eta-1)\right)\left(\frac{(1-\lambda)}{\lambda}\rho^{\epsilon-\eta}\right) + \epsilon(\epsilon-1)}{\left(\epsilon + \eta\left(\frac{1-\lambda}{\lambda}\rho^{\epsilon-\eta}\right)\right)^2} , \tag{A.1.8}$$

for all $\rho > 0$, where the assumption that $\lambda \neq 0$ is used to simplify the expression by dividing through by λ^2. Define the function $\mathcal{Z}(\mathfrak{q})$ in terms of inverse demand function $\mathcal{D}^{-1}(\mathfrak{q})$:

$$\mathcal{Z}(\mathfrak{q}) \equiv \frac{1-\lambda}{\lambda}\left(\mathcal{D}^{-1}(\mathfrak{q})\right)^{\epsilon-\eta} , \tag{A.1.9}$$

and use this together with [A.1.8] to write the derivative of marginal revenue as follows:

$$\mathcal{R}''(\mathfrak{q}) = \frac{\eta(\eta-1)\left(\mathcal{Z}(\mathfrak{q})\right)^2 - \left((\eta-\epsilon)^2 - \eta(\epsilon-1) - \epsilon(\eta-1)\right)\mathcal{Z}(\mathfrak{q}) + \epsilon(\epsilon-1)}{\mathcal{D}'\left(\mathcal{D}^{-1}(\mathfrak{q})\right)\left(\epsilon + \eta\mathcal{Z}(\mathfrak{q})\right)^2} \ . \qquad \text{[A.1.10]}$$

Since $\mathcal{D}'\left(\mathcal{D}^{-1}(\mathfrak{q})\right) < 0$ for all $\mathfrak{q}$, and the remaining term in the denominator of [A.1.10] is strictly positive, the sign of $\mathcal{R}''(\mathfrak{q})$ is the opposite of that of the quadratic function:

$$\mathcal{Q}(z) \equiv \eta(\eta-1)z^2 - \left((\eta-\epsilon)^2 - \eta(\epsilon-1) - \epsilon(\eta-1)\right)z + \epsilon(\epsilon-1) \ , \qquad \text{[A.1.11]}$$

evaluated at $z = \mathcal{Z}(\mathfrak{q})$. The aim is to find a region where marginal revenue is upward sloping, which corresponds to $\mathcal{Q}(z)$ being negative, which is in turn equivalent to its having positive roots (it is U-shaped because $\eta > 1$).

Under the assumptions $\epsilon > 1$ and $\eta > 1$, the product of the roots of quadratic equation $\mathcal{Q}(z) = 0$ is positive, so it has either no real roots, two negative real roots, or two positive real roots (possibly including repetitions). In the first two cases, since $\mathcal{Q}(0) = \epsilon(\epsilon-1) > 0$ it then follows that $\mathcal{Q}(z) > 0$ for all $z > 0$. To see which combinations of elasticities ϵ and η lead to positive real roots, define the following two quadratic functions of the elasticity η (for a given value of the elasticity ϵ):

$$\mathcal{G}_p(\eta;\epsilon) \equiv \eta^2 - (4\epsilon-1)\eta + \epsilon(\epsilon+1) \ , \quad \mathcal{G}_r(\eta;\epsilon) \equiv \eta^2 - 2(3\epsilon-1)\eta + (\epsilon+1)^2 \ . \qquad \text{[A.1.12]}$$

By comparing $\mathcal{G}_p(\eta;\epsilon)$ to the coefficient of z in [A.1.11], the sum of the roots $\mathcal{Q}(z) = 0$ is positive if and only if $\mathcal{G}_p(\eta;\epsilon) > 0$ since $\eta > 1$. Then the discriminant of the quadratic $\mathcal{Q}(z)$ is factored in terms of $\mathcal{G}_r(\eta;\epsilon)$ as follows:

$$\left((\eta-\epsilon)^2 - \eta(\epsilon-1) - \epsilon(\eta-1)\right)^2 - 4\epsilon\eta(\epsilon-1)(\eta-1) = (\eta-\epsilon)^2\mathcal{G}_r(\eta;\epsilon) \ , \qquad \text{[A.1.13]}$$

and as $\eta \neq \epsilon$, the equation $\mathcal{Q}(z) = 0$ has two distinct real roots if and only if $\mathcal{G}_r(\eta;\epsilon) > 0$. To summarize, the quadratic $\mathcal{Q}(z)$ has two positive real roots if and only if $\mathcal{G}_p(\eta;\epsilon) > 0$ and $\mathcal{G}_r(\eta;\epsilon) > 0$. It turns out that in the relevant parameter region $\eta > \epsilon > 1$, the binding condition is $\mathcal{G}_r(\eta;\epsilon) > 0$.

Since $\epsilon > 1$, the quadratic equations $\mathcal{G}_p(\eta;\epsilon) = 0$ and $\mathcal{G}_r(\eta;\epsilon) = 0$ in η (for a given value of ϵ) both have two distinct positive real roots (this is confirmed by verifying that the discriminants and the sums and products of the roots are all positive). Let $\eta^*(\epsilon)$ be the larger of the two roots of the equation $\mathcal{G}_r(\eta;\epsilon) = 0$:

$$\eta^*(\epsilon) = (3\epsilon-1) + 2\sqrt{2\epsilon(\epsilon-1)} \ ,$$

and observe that $\eta^*(\epsilon) > \epsilon$ and $\eta^{*\prime}(\epsilon) > 0$ for all $\epsilon > 1$. Since both quadratics $\mathcal{G}_p(\eta;\epsilon)$ and $\mathcal{G}_r(\eta;\epsilon)$ have positive coefficients of η^2, it follows that they are negative for all η values lying strictly between their two roots.

The difference between the two quadratic functions $\mathcal{G}_p(\eta;\epsilon)$ and $\mathcal{G}_r(\eta;\epsilon)$ in [A.1.12] is:

$$\mathcal{G}_p(\eta;\epsilon) - \mathcal{G}_r(\eta;\epsilon) = (2\epsilon-1)\eta - (\epsilon+1) \ ,$$

which is a linear function of η. Thus let $\hat{\eta}(\epsilon)$ be the unique solution for η of the equation $\mathcal{G}_p(\eta;\epsilon) = \mathcal{G}_r(\eta;\epsilon)$, taking ϵ as given. As $\epsilon > 1$, such a solution exists and is unique, and $\mathcal{G}_p(\eta;\epsilon) > \mathcal{G}_r(\eta;\epsilon)$ holds if and only if $\eta > \hat{\eta}(\epsilon)$. The difference between the solution $\hat{\eta}(\epsilon)$ and ϵ is given by:

$$\hat{\eta}(\epsilon) - \epsilon = \frac{2\epsilon - (2\epsilon^2 - 1)}{2\epsilon - 1} \ . \qquad \text{[A.1.14]}$$

Consider first the case of ϵ values where $\hat{\eta}(\epsilon) \leq \epsilon$. This means that for all $\eta > \epsilon$, $\mathcal{G}_r(\eta;\epsilon) < \mathcal{G}_p(\eta;\epsilon)$. Since $\mathcal{G}_p(\epsilon;\epsilon) = -2\epsilon(\epsilon-1) < 0$ for all $\epsilon > 1$, it follows that $\mathcal{G}_r(\epsilon;\epsilon) < 0$. Therefore, the smaller root of $\mathcal{G}_r(\eta;\epsilon) = 0$ is less than ϵ. This establishes that the only η values for which all the inequalities $\eta > \epsilon$, $\mathcal{G}_r(\eta;\epsilon) > 0$ and $\mathcal{G}_p(\eta;\epsilon) > 0$ hold are those satisfying $\eta > \eta^*(\epsilon)$.

Now consider what happens in the remaining case where $\hat{\eta}(\epsilon) > \epsilon$. By rearranging the terms in [A.1.12], notice that $\mathcal{G}_p(\eta;\epsilon) = (\eta-\epsilon)^2 - 1 - ((2\epsilon-1)\eta - (\epsilon+1))$. Therefore, from the definition of $\hat{\eta}(\epsilon)$, it follows that $\mathcal{G}_p(\hat{\eta}(\epsilon);\epsilon) = \mathcal{G}_r(\hat{\eta}(\epsilon);\epsilon) = (\hat{\eta}(\epsilon)-\epsilon)^2 - 1$. As $\hat{\eta}(\epsilon) > \epsilon$ in this case, equation [A.1.14] implies that $2\epsilon - (2\epsilon^2 - 1) > 0$, and therefore $0 < \hat{\eta}(\epsilon) - \epsilon < 1$ if $2\epsilon^2 - 1 > 1$, which is equivalent to $\epsilon^2 > 1$. This must hold since $\epsilon > 1$, and hence $(\hat{\eta}(\epsilon) - \epsilon)^2 < 1$. Thus $\mathcal{G}_p(\hat{\eta}(\epsilon); \epsilon) = \mathcal{G}_r(\hat{\eta}(\epsilon); \epsilon) < 0$. As $\mathcal{G}_p(\eta; \epsilon) > \mathcal{G}_r(\eta; \epsilon)$ holds for $\eta > \hat{\eta}(\epsilon)$, the larger of the roots of $\mathcal{G}_p(\eta; \epsilon) = 0$ lies strictly between $\hat{\eta}(\epsilon)$ and $\eta^*(\epsilon)$. Therefore in this case as well, the only values of η consistent with all the inequalities $\eta > \epsilon$, $\mathcal{G}_r(\eta; \epsilon) > 0$ and $\mathcal{G}_p(\eta; \epsilon) > 0$ are those satisfying $\eta > \eta^*(\epsilon)$.

Thus for $\eta > \epsilon > 1$, if $\eta > \eta^*(\epsilon)$ then the quadratic equation $\mathcal{Q}(z) = 0$ from [A.1.11] has two distinct positive real roots $\underline{z}$ and $\overline{z}$ with $\underline{z} < \overline{z}$. $\mathcal{Q}(z) < 0$ must hold for all $z \in (\underline{z}, \overline{z})$ since the coefficient of z^2 is positive. For $z \in [0, \underline{z})$ or $z \in (\overline{z}, \infty)$, the quadratic satisfies $\mathcal{Q}(z) > 0$. If $\eta \leq \eta^*(\epsilon)$ then $\mathcal{Q}(z) > 0$ for all z (except at a single isolated point when $\eta = \eta^*(\epsilon)$ exactly). Therefore, in the case where $\eta \leq \eta^*(\epsilon)$, it follows from [A.1.10] and [A.1.11] that $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing for all $\mathfrak{q} \geq 0$.

Now restrict attention to the case where $\eta > \eta^*(\epsilon)$. Since $0 < \lambda < 1$, $\eta > \epsilon$, and the inverse demand function $\mathcal{D}^{-1}(\mathfrak{q})$ is strictly decreasing, the function $\mathcal{Z}(\mathfrak{q})$ defined in [A.1.9] is strictly increasing. Its inverse is:

$$\mathcal{Z}^{-1}(z) = \mathcal{D}\left(\left(\frac{\lambda}{1-\lambda} z\right)^{\frac{1}{\epsilon - \eta}}\right) , \qquad \text{[A.1.15]}$$

which is also a strictly increasing function. Define $\underline{\mathfrak{q}} \equiv \mathcal{Z}^{-1}(\underline{z})$ and $\overline{\mathfrak{q}} \equiv \mathcal{Z}^{-1}(\overline{z})$ using the roots $\underline{z}$ and $\overline{z}$ of the quadratic equation $\mathcal{Q}(z) = 0$. From [A.1.10] and [A.1.11] it follows that $\mathcal{R}''(\underline{\mathfrak{q}}) = 0$ and $\mathcal{R}''(\overline{\mathfrak{q}}) = 0$. As $\mathcal{Z}^{-1}(z)$ is a strictly increasing function, $\mathcal{R}'(\mathfrak{q})$ must be strictly decreasing for $0 < \mathfrak{q} < \underline{\mathfrak{q}}$ and $\mathfrak{q} > \overline{\mathfrak{q}}$, and strictly increasing for $\underline{\mathfrak{q}} < \mathfrak{q} < \overline{\mathfrak{q}}$. The condition $\eta > \eta^*(\epsilon)$ is the same as that given in [4.3], so this completes the proof. ■

When the marginal revenue function $\mathcal{R}'(\mathfrak{q})$ is non-monotonic, the following result provides the foundation for verifying the existence and uniqueness of the two-price equilibrium.

Lemma 2 *Given the total revenue function $\mathcal{R}(\mathfrak{q})$ defined in [A.1.4], suppose that $0 < \lambda < 1$, and ϵ and η are such that non-monotonicity condition [4.3] holds:*

(i) There exist unique values $\mathfrak{q}_S$ and $\mathfrak{q}_N$ such that $0 < \mathfrak{q}_N < \mathfrak{q}_S < \infty$ which satisfy the equations:

$$\mathcal{R}'(\mathfrak{q}_S) = \mathcal{R}'(\mathfrak{q}_N) = \frac{\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}(\mathfrak{q}_N)}{\mathfrak{q}_S - \mathfrak{q}_N} . \qquad \text{[A.1.16]}$$

(ii) The solutions $\mathfrak{q}_S$ and $\mathfrak{q}_N$ of the above equations must also satisfy the inequalities:

$$\mathcal{R}''(\mathfrak{q}_S) < 0 , \quad \mathcal{R}''(\mathfrak{q}_N) < 0 , \quad \mathcal{R}(\mathfrak{q}_S)/\mathfrak{q}_S > \mathcal{R}'(\mathfrak{q}_S) , \quad \mathcal{R}(\mathfrak{q}_N)/\mathfrak{q}_N > \mathcal{R}'(\mathfrak{q}_N) . \qquad \text{[A.1.17]}$$

(iii) The following inequality holds for all $\mathfrak{q} \geq 0$:

$$\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_S) + \mathcal{R}'(\mathfrak{q}_S)(\mathfrak{q} - \mathfrak{q}_S) = \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\mathfrak{q} - \mathfrak{q}_N) . \qquad \text{[A.1.18]}$$

Proof (i) When $0 < \lambda < 1$ and condition [4.3] hold then Lemma 1 demonstrates that there exist $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$ such that $0 < \underline{\mathfrak{q}} < \overline{\mathfrak{q}} < \infty$ and $\mathcal{R}''(\underline{\mathfrak{q}}) = \mathcal{R}''(\overline{\mathfrak{q}}) = 0$. Define $\underline{\mathcal{R}'} \equiv \mathcal{R}'(\underline{\mathfrak{q}})$ and $\overline{\mathcal{R}'} \equiv \mathcal{R}'(\overline{\mathfrak{q}})$. Since Lemma 1 also shows that $\mathcal{R}'(\mathfrak{q})$ is strictly increasing between $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$, it follows that $\underline{\mathcal{R}'} < \overline{\mathcal{R}'}$.

The function $\mathcal{R}'(\mathfrak{q})$ is continuously differentiable for all $\mathfrak{q} > 0$ and $\lim_{\mathfrak{q} \to 0} \mathcal{R}'(\mathfrak{q}) = \infty$. Hence there must exist a value $\underline{\mathfrak{q}}_1$ such that $\mathcal{R}'(\underline{\mathfrak{q}}_1) = \overline{\mathcal{R}'}$ and $\underline{\mathfrak{q}}_1 < \underline{\mathfrak{q}}$. Define $\overline{\mathfrak{q}}_1 \equiv \underline{\mathfrak{q}}$. According to Lemma 1, the function $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing on the interval $[\underline{\mathfrak{q}}_1, \overline{\mathfrak{q}}_1]$ and thus has range $[\underline{\mathcal{R}'}, \overline{\mathcal{R}'}]$.

Define $\underline{\mathfrak{q}}_2 \equiv \underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}_2 \equiv \overline{\mathfrak{q}}$. Given the construction of $\underline{\mathcal{R}'}$ and $\overline{\mathcal{R}'}$ and the fact that $\mathcal{R}'(\mathfrak{q})$ is strictly increasing on $[\underline{\mathfrak{q}}_2, \overline{\mathfrak{q}}_2]$, the range of $\mathcal{R}'(\mathfrak{q})$ is $[\underline{\mathcal{R}'}, \overline{\mathcal{R}'}]$ on this interval.

Now define $\underline{\mathfrak{q}}_3 \equiv \overline{\mathfrak{q}}$. Since $\lim_{\mathfrak{q} \to \infty} \mathcal{R}'(\mathfrak{q}) = 0$ and $\mathcal{R}'(\mathfrak{q})$ is continuously differentiable, there must exist a $\overline{\mathfrak{q}}_3$ such that $\mathcal{R}'(\overline{\mathfrak{q}}_3) = \underline{\mathcal{R}'}$ and $\overline{\mathfrak{q}}_3 > \underline{\mathfrak{q}}_3$. Lemma 1 shows that $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing on $[\underline{\mathfrak{q}}_3, \overline{\mathfrak{q}}_3]$ and so has range $[\underline{\mathcal{R}'}, \overline{\mathcal{R}'}]$ on this interval.

For each $\varkappa \in [0, 1]$, define the function $\mathfrak{q}_1(\varkappa)$ as follows:

$$\mathfrak{q}_1(\varkappa) \equiv (1 - \varkappa)\underline{\mathfrak{q}}_1 + \varkappa \overline{\mathfrak{q}}_1 , \qquad \text{[A.1.19]}$$

in other words, as a convex combination of $\underline{\mathfrak{q}}_1$ and $\overline{\mathfrak{q}}_1$. Note that $\mathfrak{q}_1(\varkappa)$ is strictly increasing in $\varkappa$. The construction of this function, the monotonicity of $\mathcal{R}'(\mathfrak{q})$ on $[\underline{\mathfrak{q}}_1, \overline{\mathfrak{q}}_1]$, and the definitions of $\underline{\mathcal{R}'}$ and $\overline{\mathcal{R}'}$ guarantee that $\underline{\mathcal{R}'} \leq \mathcal{R}'(\mathfrak{q}_1(\varkappa)) \leq \overline{\mathcal{R}'}$ for all $\varkappa \in [0,1]$. Given that the function $\mathcal{R}'(\mathfrak{q})$ is also strictly monotonic on each of the intervals $[\underline{\mathfrak{q}}_2, \overline{\mathfrak{q}}_2]$ and $[\underline{\mathfrak{q}}_3, \overline{\mathfrak{q}}_3]$, and has range $[\underline{\mathcal{R}'}, \overline{\mathcal{R}'}]$ on both, there must exist unique values $\mathfrak{q}_2 \in [\underline{\mathfrak{q}}_2, \overline{\mathfrak{q}}_2]$ and $\mathfrak{q}_3 \in [\underline{\mathfrak{q}}_3, \overline{\mathfrak{q}}_3]$ such that $\mathcal{R}'(\mathfrak{q}_2) = \mathcal{R}'(\mathfrak{q}_3) = \mathcal{R}'(\mathfrak{q}_1(\varkappa))$ for any particular $\varkappa$. Hence define the functions $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$ to give these values in the two intervals for each specific $\varkappa \in [0,1]$:

$$\mathcal{R}'(\mathfrak{q}_1(\varkappa)) \equiv \mathcal{R}'(\mathfrak{q}_2(\varkappa)) \equiv \mathcal{R}'(\mathfrak{q}_3(\varkappa)) \ . \qquad \text{[A.1.20]}$$

At the endpoints of the intervals (corresponding to $\varkappa = 0$ and $\varkappa = 1$) note that:

$$\mathfrak{q}_2(0) = \mathfrak{q}_3(0) = \overline{\mathfrak{q}} \ , \quad \mathfrak{q}_1(1) = \mathfrak{q}_2(1) = \underline{\mathfrak{q}} \ . \qquad \text{[A.1.21]}$$

Continuity and differentiability of $\mathcal{R}'(\mathfrak{q})$ and of $\mathfrak{q}_1(\varkappa)$ from [A.1.19] guarantee that $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$ are continuous for all $\varkappa \in [0,1]$ and differentiable for all $\varkappa \in (0,1)$. By differentiating [A.1.20] it follows that:

$$\mathfrak{q}_2'(\varkappa) = \frac{\mathcal{R}''(\mathfrak{q}_1(\varkappa))}{\mathcal{R}''(\mathfrak{q}_2(\varkappa))}\mathfrak{q}_1'(\varkappa) \ , \quad \mathfrak{q}_3'(\varkappa) = \frac{\mathcal{R}''(\mathfrak{q}_1(\varkappa))}{\mathcal{R}''(\mathfrak{q}_3(\varkappa))}\mathfrak{q}_1'(\varkappa) \ .$$

As Lemma 1 establishes $\mathcal{R}(\mathfrak{q})$ is concave on $[\underline{\mathfrak{q}}_1, \overline{\mathfrak{q}}_1]$ and $[\underline{\mathfrak{q}}_3, \overline{\mathfrak{q}}_3]$, and convex on $[\underline{\mathfrak{q}}_2, \overline{\mathfrak{q}}_2]$, the results above show that $\mathfrak{q}_2'(\varkappa) < 0$ and $\mathfrak{q}_3'(\varkappa) > 0$ for all $\varkappa \in (0,1)$.

Existence

For each $\varkappa \in [0,1]$, define the function $\mathcal{F}(\varkappa)$ in terms of the following integrals:

$$\mathcal{F}(\varkappa) \equiv \int_{\mathfrak{q}_2(\varkappa)}^{\mathfrak{q}_3(\varkappa)} \left(\mathcal{R}'(\mathfrak{q}) - \mathcal{R}'(\mathfrak{q}_2(\varkappa))\right) d\mathfrak{q} - \int_{\mathfrak{q}_1(\varkappa)}^{\mathfrak{q}_2(\varkappa)} \left(\mathcal{R}'(\mathfrak{q}_2(\varkappa)) - \mathcal{R}'(\mathfrak{q})\right) d\mathfrak{q} \ . \qquad \text{[A.1.22]}$$

From continuity and differentiability of $\mathfrak{q}_1(\varkappa)$, $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$, it follows that $\mathcal{F}(\varkappa)$ is also continuous for all $\varkappa \in [0,1]$ and differentiable for all $\varkappa \in (0,1)$. Evaluating $\mathcal{F}(\varkappa)$ at the endpoints of the interval $[0,1]$ and making use of [A.1.21] yields:

$$\mathcal{F}(0) = -\int_{\underline{\mathfrak{q}}_1}^{\overline{\mathfrak{q}}_2} \left(\overline{\mathcal{R}'} - \mathcal{R}'(\mathfrak{q})\right) d\mathfrak{q} < 0 \ , \quad \mathcal{F}(1) = \int_{\underline{\mathfrak{q}}_2}^{\overline{\mathfrak{q}}_3} \left(\mathcal{R}'(\mathfrak{q}) - \underline{\mathcal{R}'}\right) d\mathfrak{q} > 0 \ ,$$

where the first inequality follows because $\mathcal{R}'(\mathfrak{q}) < \overline{\mathcal{R}'}$ for all $\underline{\mathfrak{q}}_1 < \mathfrak{q} < \overline{\mathfrak{q}}_2$, and the second because $\mathcal{R}'(\mathfrak{q}) > \underline{\mathcal{R}'}$ for all $\underline{\mathfrak{q}}_2 < \mathfrak{q} < \overline{\mathfrak{q}}_3$. Differentiating $\mathcal{F}(\varkappa)$ in [A.1.22] using Leibniz's rule and substituting the definitions from [A.1.20] leads to the following result:

$$\mathcal{F}'(\varkappa) = -(\mathfrak{q}_3(\varkappa) - \mathfrak{q}_1(\varkappa))\mathfrak{q}_2'(\varkappa)\mathcal{R}''(\mathfrak{q}_2(\varkappa)) > 0 \ ,$$

for all $\varkappa \in (0,1)$ since $\mathfrak{q}_3(\varkappa) > \mathfrak{q}_1(\varkappa)$, $\mathfrak{q}_2'(\varkappa) < 0$, and $\mathcal{R}''(\mathfrak{q}_2(\varkappa)) > 0$ from Lemma 1. Therefore, because $\mathcal{F}(0) < 0$, $\mathcal{F}(1) > 0$, and $\mathcal{F}(\varkappa)$ is continuous and strictly increasing in $\varkappa$, there exists a unique $\varkappa^* \in (0,1)$ such that $\mathcal{F}(\varkappa^*) = 0$.

The solution of the system of equations [A.1.16] is found by setting $\mathfrak{q}_N \equiv \mathfrak{q}_1(\varkappa^*)$ and $\mathfrak{q}_S \equiv \mathfrak{q}_3(\varkappa^*)$, using the solution $\varkappa = \varkappa^*$ of the equation $\mathcal{F}(\varkappa) = 0$ obtained above. From [A.1.20], it follows immediately that $\mathcal{R}'(\mathfrak{q}_N) = \mathcal{R}'(\mathfrak{q}_S)$, establishing the first equality in [A.1.16]. Since $\mathcal{F}(\varkappa^*) = 0$, the definition of $\mathcal{F}(\varkappa)$ in equation [A.1.22] implies:

$$\int_{\mathfrak{q}_2(\varkappa^*)}^{\mathfrak{q}_S} \left(\mathcal{R}'(\mathfrak{q}) - \mathcal{R}'(\mathfrak{q}_2(\varkappa^*))\right) d\mathfrak{q} = \int_{\mathfrak{q}_N}^{\mathfrak{q}_2(\varkappa^*)} \left(\mathcal{R}'(\mathfrak{q}_2(\varkappa^*)) - \mathcal{R}'(\mathfrak{q})\right) d\mathfrak{q} \ , \qquad \text{[A.1.23]}$$

which is rearranged to deduce:

$$\int_{\mathfrak{q}_N}^{\mathfrak{q}_S} \mathcal{R}'(\mathfrak{q}) d\mathfrak{q} = (\mathfrak{q}_S - \mathfrak{q}_N)\mathcal{R}'(\mathfrak{q}_2(\varkappa^*)) \ . \qquad \text{[A.1.24]}$$

Equation [A.1.20] implies $\mathcal{R}'(\mathfrak{q}_2(\varkappa^*)) = \mathcal{R}'(\mathfrak{q}_N) = \mathcal{R}'(\mathfrak{q}_S)$ which together with the above establishes that:

$$\mathcal{R}'(\mathfrak{q}_S) = \mathcal{R}'(\mathfrak{q}_N) = \frac{\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}(\mathfrak{q}_N)}{\mathfrak{q}_S - \mathfrak{q}_N} , \quad \text{[A.1.25]}$$

that is, the values of $\mathfrak{q}_N$ and $\mathfrak{q}_S$ are indeed a solution of the system of equations in [A.1.16].

Uniqueness

First note that given the monotonicity of $\mathcal{R}'(\mathfrak{q})$ on the intervals $[0, \underline{\mathfrak{q}}]$ and $[\overline{\mathfrak{q}}, \infty)$, and using the fact that the range of $\mathcal{R}'(\mathfrak{q})$ is $[\underline{\mathcal{R}'}, \overline{\mathcal{R}'}]$ on $[\underline{\mathfrak{q}}_1, \overline{\mathfrak{q}}_1]$, $[\underline{\mathfrak{q}}_2, \overline{\mathfrak{q}}_2]$ and $[\underline{\mathfrak{q}}_3, \overline{\mathfrak{q}}_3]$, it follows that no solution of [A.1.16] can be found in either $[0, \underline{\mathfrak{q}}_1)$ or $(\overline{\mathfrak{q}}_3, \infty)$ since marginal revenue needs to be equalized at two quantities. Furthermore, as the definitions of the functions $\mathfrak{q}_1(\varkappa)$, $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$ in [A.1.20] make clear, it is necessary that those two quantities correspond to two out of the three of $\mathfrak{q}_1(\varkappa)$, $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$ for some particular $\varkappa \in [0, 1]$ if marginal revenue is to be equalized at two distinct points.

In addition to equalizing marginal revenue, the solution $\mathfrak{q}_S$ and $\mathfrak{q}_N$ must satisfy the second equality in [A.1.16]. If $\mathfrak{q}_N$ is set equal to $\mathfrak{q}_1(\varkappa)$ and $\mathfrak{q}_S$ equal to $\mathfrak{q}_3(\varkappa)$ for the same value of $\varkappa \in [0, 1]$ then equations [A.1.23] and [A.1.24] show that the second equality in [A.1.16] requires $\digamma(\varkappa) = 0$. But it has already been demonstrated that there is one and only one solution of this equation.

Now consider the alternative choices of setting $\mathfrak{q}_N$ to $\mathfrak{q}_1(\varkappa)$ and $\mathfrak{q}_S$ to $\mathfrak{q}_2(\varkappa)$ for some common $\varkappa \in [0, 1]$, or to $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$ respectively, again for some common value of $\varkappa$. Since [A.1.20] holds by construction, and the function $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing on the intervals $[\underline{\mathfrak{q}}_1, \overline{\mathfrak{q}}_1]$ and $[\underline{\mathfrak{q}}_3, \overline{\mathfrak{q}}_3]$, and strictly increasing on $[\underline{\mathfrak{q}}_2, \overline{\mathfrak{q}}_2]$, it follows that:

$$\int_{\mathfrak{q}_1(\varkappa)}^{\mathfrak{q}_2(\varkappa)} \mathcal{R}'(\mathfrak{q}) d\mathfrak{q} < (\mathfrak{q}_2(\varkappa) - \mathfrak{q}_1(\varkappa))\mathcal{R}'(\mathfrak{q}_2(\varkappa)) , \quad \int_{\mathfrak{q}_2(\varkappa)}^{\mathfrak{q}_3(\varkappa)} \mathcal{R}'(\mathfrak{q}) d\mathfrak{q} > (\mathfrak{q}_3(\varkappa) - \mathfrak{q}_2(\varkappa))\mathcal{R}'(\mathfrak{q}_2(\varkappa)) ,$$

and hence both inequalities $\mathcal{R}(\mathfrak{q}_2(\varkappa)) - \mathcal{R}(\mathfrak{q}_1(\varkappa)) < (\mathfrak{q}_2(\varkappa) - \mathfrak{q}_1(\varkappa))\mathcal{R}'(\mathfrak{q}_2(\varkappa))$ and $\mathcal{R}(\mathfrak{q}_3(\varkappa)) - \mathcal{R}(\mathfrak{q}_2(\varkappa)) > (\mathfrak{q}_3(\varkappa) - \mathfrak{q}_2(\varkappa))\mathcal{R}'(\mathfrak{q}_2(\varkappa))$ must hold for every $\varkappa \in [0, 1]$. Consequently, there is no way that all three equations in [A.1.25] can hold except by setting $\mathfrak{q}_N = \mathfrak{q}_1(\varkappa^*)$ and $\mathfrak{q}_S = \mathfrak{q}_3(\varkappa^*)$. Therefore the solution of [A.1.16] constructed above is unique.

(ii) Lemma 1 shows that $\mathcal{R}(\mathfrak{q})$ is a strictly concave function on the intervals $[0, \underline{\mathfrak{q}}]$ and $[\overline{\mathfrak{q}}, \infty)$. The argument above demonstrating the existence and uniqueness of the solution establishes that $\mathfrak{q}_N$ and $\mathfrak{q}_S$ must lie respectively in the intervals $(\underline{\mathfrak{q}}_1, \overline{\mathfrak{q}}_1)$ and $(\underline{\mathfrak{q}}_3, \overline{\mathfrak{q}}_3)$, which are themselves contained in $[0, \underline{\mathfrak{q}}]$ and $[\overline{\mathfrak{q}}, \infty)$ respectively. Together these findings imply $\mathcal{R}''(\mathfrak{q}_N) < 0$ and $\mathcal{R}''(\mathfrak{q}_S) < 0$, and that the following inequalities must hold:

$$\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\mathfrak{q} - \mathfrak{q}_N) \quad \forall \mathfrak{q} \in [0, \underline{\mathfrak{q}}] , \quad \text{and } \mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_S) + \mathcal{R}'(\mathfrak{q}_S)(\mathfrak{q} - \mathfrak{q}_S) \quad \forall \mathfrak{q} \in [\overline{\mathfrak{q}}, \infty) , \quad \text{[A.1.26]}$$

where the inequalities are strict for $\mathfrak{q} \neq \mathfrak{q}_N$ and $\mathfrak{q} \neq \mathfrak{q}_S$ respectively. Note that an implication of the equations characterizing $\mathfrak{q}_S$ and $\mathfrak{q}_N$ in [A.1.16] is:

$$\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}'(\mathfrak{q}_S)\mathfrak{q}_S = \mathcal{R}(\mathfrak{q}_N) - \mathcal{R}'(\mathfrak{q}_N)\mathfrak{q}_N . \quad \text{[A.1.27]}$$

By evaluating the first inequality in [A.1.26] at $\mathfrak{q} = 0$, where $\mathcal{R}(0) = 0$, and making use of the equation above it is deduced that:

$$\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}'(\mathfrak{q}_S)\mathfrak{q}_S > 0 , \quad \mathcal{R}(\mathfrak{q}_N) - \mathcal{R}'(\mathfrak{q}_N)\mathfrak{q}_N > 0 ,$$

and thus $\mathcal{R}(\mathfrak{q}_S)/\mathfrak{q}_S > \mathcal{R}'(\mathfrak{q}_S)$ and $\mathcal{R}(\mathfrak{q}_N)/\mathfrak{q}_N > \mathcal{R}'(\mathfrak{q}_N)$. This confirms all the inequalities given in [A.1.17].

(iii) By applying the inequalities in [A.1.26] at the endpoints $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$ of the intervals $[0, \underline{\mathfrak{q}}]$ and $[\overline{\mathfrak{q}}, \infty)$ it follows that:

$$\mathcal{R}(\underline{\mathfrak{q}}) \leq \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\underline{\mathfrak{q}} - \mathfrak{q}_N) , \quad \text{and } \mathcal{R}(\overline{\mathfrak{q}}) \leq \mathcal{R}(q_N) + \mathcal{R}'(\mathfrak{q}_N)(\overline{\mathfrak{q}} - \mathfrak{q}_N) . \quad \text{[A.1.28]}$$

Now take any $\mathfrak{q} \in (\underline{\mathfrak{q}}, \overline{\mathfrak{q}})$ and note that because Lemma 1 demonstrates $\mathcal{R}(\mathfrak{q})$ is a convex function on this interval:

$$\mathcal{R}(\mathfrak{q}) \equiv \mathcal{R}\left(\left(\frac{\overline{\mathfrak{q}} - \mathfrak{q}}{\overline{\mathfrak{q}} - \underline{\mathfrak{q}}}\right)\underline{\mathfrak{q}} + \left(\frac{\mathfrak{q} - \underline{\mathfrak{q}}}{\overline{\mathfrak{q}} - \underline{\mathfrak{q}}}\right)\overline{\mathfrak{q}}\right) \leq \left(\frac{\overline{\mathfrak{q}} - \mathfrak{q}}{\overline{\mathfrak{q}} - \underline{\mathfrak{q}}}\right)\mathcal{R}(\underline{\mathfrak{q}}) + \left(\frac{\mathfrak{q} - \underline{\mathfrak{q}}}{\overline{\mathfrak{q}} - \underline{\mathfrak{q}}}\right)\mathcal{R}(\overline{\mathfrak{q}}) , \qquad \text{[A.1.29]}$$

using the fact that the coefficients of $\mathcal{R}(\underline{\mathfrak{q}})$ and $\mathcal{R}(\overline{\mathfrak{q}})$ in the above are positive and sum to one. A weighted average of the two inequalities in [A.1.28] using as weights the coefficients from [A.1.29] yields $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\mathfrak{q} - \mathfrak{q}_N)$ for all $\mathfrak{q} \in (\underline{\mathfrak{q}}, \overline{\mathfrak{q}})$. This finding, together with the inequalities in [A.1.26] and the equations [A.1.25] and [A.1.27], implies:

$$\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_S) + \mathcal{R}'(\mathfrak{q}_S)(\mathfrak{q} - \mathfrak{q}_S) = \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\mathfrak{q} - \mathfrak{q}_N)$$

for all $\mathfrak{q} \geq 0$. Thus [A.1.18] is established, which completes the proof. ■

The existence and uniqueness of the solution of equations [A.1.16] has been demonstrated given condition [4.3] for the non-monotonicity of the marginal revenue function $\mathcal{R}'(\mathfrak{q})$. A method for computing this solution and a characterization of which parameters it depends upon is provided in the following result.

Lemma 3 *Let $\mathfrak{q}_S$ and $\mathfrak{q}_N$ be the solution of equations* [A.1.16] *(under the conditions assumed in Lemma 2), and let $\rho_N \equiv \mathcal{D}^{-1}(\mathfrak{q}_N)$ and $\rho_S \equiv \mathcal{D}^{-1}(\mathfrak{q}_S)$ be the corresponding relative prices consistent with the demand function* [A.1.1]. *In addition, define the markup ratio $\mu \equiv \mu_S/\mu_N = \rho_S/\rho_N$ and the quantity ratio $\chi \equiv \mathfrak{q}_S/\mathfrak{q}_N$.*

Consider the functions:

$$\mathfrak{a}_0(\mu;\epsilon,\eta) \equiv \epsilon(\epsilon-1)\mu^{\eta-\epsilon} , \qquad \text{[A.1.30a]}$$

$$\mathfrak{a}_1(\mu;\epsilon,\eta) \equiv \eta(\epsilon-1)\left(\frac{1-\mu^{\eta-\epsilon+1}}{1-\mu}\right) + \epsilon(\eta-1)\left(\frac{\mu^{\eta-\epsilon}-\mu}{1-\mu}\right) , \qquad \text{[A.1.30b]}$$

$$\mathfrak{a}_2(\eta) \equiv \eta(\eta-1) , \qquad \text{[A.1.30c]}$$

$$\mathfrak{b}_0(\mu;\epsilon,\eta) \equiv (\epsilon-1)\left(\frac{\mu^{2(\eta-\epsilon)}-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right) , \qquad \text{[A.1.30d]}$$

$$\mathfrak{b}_1(\mu;\epsilon,\eta) \equiv (\eta-1)\left(\frac{\mu^{2(\eta-\epsilon)}-\mu^{\eta}}{1-\mu^{\eta}}\right) + 2(\epsilon-1)\left(\frac{\mu^{\eta-\epsilon}-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right) , \qquad \text{[A.1.30e]}$$

$$\mathfrak{b}_2(\mu;\epsilon,\eta) \equiv (\epsilon-1)\left(\frac{1-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right) + 2(\eta-1)\left(\frac{\mu^{\eta-\epsilon}-\mu^{\eta}}{1-\mu^{\eta}}\right) , \qquad \text{[A.1.30f]}$$

$$\mathfrak{b}_3(\eta) \equiv (\eta-1) , \qquad \text{[A.1.30g]}$$

and the resultant $\mathfrak{R}(\mu;\epsilon,\eta)$, defined in terms of the following determinant:

$$\mathfrak{R}(\mu;\epsilon,\eta) \equiv \begin{vmatrix} \mathfrak{a}_0(\mu;\epsilon,\eta) & \mathfrak{a}_1(\mu;\epsilon,\eta) & \mathfrak{a}_2(\eta) & 0 & 0 \\ 0 & \mathfrak{a}_0(\mu;\epsilon,\eta) & \mathfrak{a}_1(\mu;\epsilon,\eta) & \mathfrak{a}_2(\eta) & 0 \\ 0 & 0 & \mathfrak{a}_0(\mu;\epsilon,\eta) & \mathfrak{a}_1(\mu;\epsilon,\eta) & \mathfrak{a}_2(\eta) \\ \mathfrak{b}_0(\mu;\epsilon,\eta) & \mathfrak{b}_1(\mu;\epsilon,\eta) & \mathfrak{b}_2(\mu;\epsilon,\eta) & \mathfrak{b}_3(\eta) & 0 \\ 0 & \mathfrak{b}_0(\mu;\epsilon,\eta) & \mathfrak{b}_1(\mu;\epsilon,\eta) & \mathfrak{b}_2(\mu;\epsilon,\eta) & \mathfrak{b}_3(\eta) \end{vmatrix} . \qquad \text{[A.1.31]}$$

Define also the function $\mathfrak{z}(\mu;\epsilon,\eta)$:

$$\mathfrak{z}(\mu;\epsilon,\eta) \equiv \frac{-\mathfrak{a}_1(\mu;\epsilon,\eta) - \sqrt{\mathfrak{a}_1(\mu;\epsilon,\eta)^2 - 4\mathfrak{a}_2(\eta)\mathfrak{a}_0(\mu;\epsilon,\eta)}}{2\mathfrak{a}_2(\eta)} . \qquad \text{[A.1.32]}$$

(i) *The markup ratio $\mu \equiv \rho_S/\rho_N$ is the only solution of $\mathfrak{R}(\mu;\epsilon,\eta) = 0$ for $0 < \mu < 1$ where $\mathfrak{z}(\mu;\epsilon,\eta)$ is a positive real number. Thus μ depends only on parameters ϵ and η.*

(ii) *Given the value of μ satisfying $\mathfrak{R}(\mu;\epsilon,\eta) = 0$, the quantity ratio $\chi \equiv \mathfrak{q}_S/\mathfrak{q}_N$ is:*

$$\chi = \mu^{-\epsilon}\left(\frac{1+\mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu;\epsilon,\eta)}{1+\mathfrak{z}(\mu;\epsilon,\eta)}\right) , \qquad \text{[A.1.33]}$$

which depends only on parameters ϵ and η.

(iii) *The equilibrium markups μ_S and μ_N from [4.6] depend only on ϵ and η and are given by:*

$$\mu_S = \frac{\epsilon + \eta\mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu;\epsilon,\eta)}{(\epsilon-1) + (\eta-1)\mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu;\epsilon,\eta)}\,, \quad \mu_N = \frac{\epsilon + \eta\mathfrak{z}(\mu;\epsilon,\eta)}{(\epsilon-1)+(\eta-1)\mathfrak{z}(\mu;\epsilon,\eta)}\,. \qquad \text{[A.1.34]}$$

(iv) *The equilibrium values of ρ_N, ρ_S, $\mathfrak{q}_N$ and $\mathfrak{q}_S$ depend on parameters ϵ, η and λ and are obtained as follows:*

$$\rho_N = \left(\frac{\lambda}{1-\lambda}\mathfrak{z}(\mu;\epsilon,\eta)\right)^{-\frac{1}{\eta-\epsilon}}\,, \quad \rho_S = \left(\frac{\lambda}{1-\lambda}\mathfrak{z}(\mu;\epsilon,\eta)\right)^{-\frac{1}{\eta-\epsilon}}\mu\,, \qquad \text{[A.1.35]}$$

with $\mathfrak{q}_N = \mathcal{D}(\rho_N)$ and $\mathfrak{q}_S = \mathcal{D}(\rho_S)$ using the relative demand function $\mathcal{D}(\rho)$ from [A.1.1].

Proof (i) Using the expression for marginal revenue from [A.1.5], the first equality in [A.1.16] is equivalent to the requirement that:

$$\left(\frac{\lambda(\epsilon-1)+(1-\lambda)(\eta-1)\rho_N^{\epsilon-\eta}}{\lambda\epsilon+(1-\lambda)\eta\rho_N^{\epsilon-\eta}}\right)\rho_N = \left(\frac{\lambda(\epsilon-1)+(1-\lambda)(\eta-1)\rho_S^{\epsilon-\eta}}{\lambda\epsilon+(1-\lambda)\eta\rho_S^{\epsilon-\eta}}\right)\rho_S\,.$$

By dividing numerator and denominator of the above by λ, defining $z \equiv ((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, and restating the resulting equation in terms of $\mu = \rho_S/\rho_N$ and z it follows that:

$$\mu = \left(\frac{\epsilon+\eta\mu^{-(\eta-\epsilon)}z}{\epsilon+\eta z}\right)\left(\frac{(\epsilon-1)+(\eta-1)z}{(\epsilon-1)+(\eta-1)\mu^{-(\eta-\epsilon)}z}\right)\,. \qquad \text{[A.1.36]}$$

Since $\rho_S < \rho_N$ the markup ratio satisfies $0<\mu<1$, and thus neither of the denominators of the fractions above can be zero. Hence for a given value of μ, equation [A.1.36] is rearranged to obtain a quadratic equation in z:

$$\eta(\eta-1)\mu^{-(\eta-\epsilon)}(1-\mu)z^2 + \left(\epsilon(\eta-1)\left(1-\mu^{1-(\eta-\epsilon)}\right)+\eta(\epsilon-1)\left(\mu^{-(\eta-\epsilon)}-\mu\right)\right)z + \epsilon(\epsilon-1)(1-\mu) = 0\,,$$

which as $0<\mu<1$ is in turn multiplied on both sides by $\mu^{\eta-\epsilon}/(1-\mu)$ to obtain an equivalent quadratic:

$$\eta(\eta-1)z^2 + \left(\eta(\epsilon-1)\left(\frac{1-\mu^{\eta-\epsilon+1}}{1-\mu}\right)+\epsilon(\eta-1)\left(\frac{\mu^{\eta-\epsilon}-\mu}{1-\mu}\right)\right)z + \epsilon(\epsilon-1)\mu^{\eta-\epsilon} = 0\,. \qquad \text{[A.1.37]}$$

This quadratic is denoted by $\mathfrak{Q}(z;\mu,\epsilon,\eta) \equiv \mathfrak{a}_0(\mu;\epsilon,\eta)+\mathfrak{a}_1(\mu;\epsilon,\eta)z+\mathfrak{a}_2(\eta)z^2$, where the coefficient functions $\mathfrak{a}_0(\mu;\epsilon,\eta)$, $\mathfrak{a}_1(\mu;\epsilon,\eta)$ and $\mathfrak{a}_2(\eta)$ listed in [A.1.30] are obtained directly from [A.1.37].

Now note that $\mathcal{R}(\mathfrak{q}_N) - \mathfrak{q}_N\mathcal{R}'(\mathfrak{q}_N) = \mathcal{R}(\mathfrak{q}_S) - \mathfrak{q}_S\mathcal{R}'(\mathfrak{q}_S)$ is deduced by rearranging the equations in [A.1.16]. The definition of the total revenue function $\mathcal{R}(\mathfrak{q})$ in [A.1.4] shows that $\mathcal{R}(\mathcal{D}(\rho)) = \rho\mathcal{D}(\rho)$ is a valid alternative expression for all $\rho > 0$. By combining these two observations and substituting $\mathfrak{q}_S = \mathcal{D}(\rho_S)$ and $\mathfrak{q}_N = \mathcal{D}(\rho_N)$, the relative prices and quantities must satisfy:

$$\mathfrak{q}_S\left(\rho_S - \mathcal{R}'(\mathfrak{q}_S)\right) = \mathfrak{q}_N\left(\rho_N - \mathcal{R}'(\mathfrak{q}_N)\right)\,. \qquad \text{[A.1.38]}$$

After expressing this in terms of the quantity ratio $\chi \equiv \mathfrak{q}_S/\mathfrak{q}_N$ and dividing both sides by $\mathcal{R}'(\mathfrak{q}_S) = \mathcal{R}'(\mathfrak{q}_N)$ (justified by [A.1.16]), equation [A.1.38] becomes:

$$\chi = \left(\frac{\rho_N}{\mathcal{R}'(\mathcal{D}(\rho_N))} - 1\right)\Big/\left(\frac{\rho_S}{\mathcal{R}'(\mathcal{D}(\rho_S))} - 1\right)\,. \qquad \text{[A.1.39]}$$

The formula for marginal revenue $\mathcal{R}'(\mathcal{D}(\rho))$ in [A.1.5] is then rearranged to show:

$$\frac{\rho}{\mathcal{R}'(\mathcal{D}(\rho))} - 1 = \frac{\lambda+(1-\lambda)\rho^{\epsilon-\eta}}{\lambda(\epsilon-1)+(\eta-1)(1-\lambda)\rho^{\epsilon-\eta}}\,,$$

which is substituted into [A.1.39] to obtain:

$$\chi = \left(\frac{\lambda + (1-\lambda)\rho_N^{\epsilon-\eta}}{\lambda + (1-\lambda)\rho_S^{\epsilon-\eta}}\right)\left(\frac{(\epsilon-1)\lambda + (\eta-1)(1-\lambda)\rho_S^{\epsilon-\eta}}{(\epsilon-1)\lambda + (\eta-1)(1-\lambda)\rho_N^{\epsilon-\eta}}\right) .$$

By dividing numerator and denominator of both fractions by λ and recalling $\mu = \rho_S/\rho_N$ and the definition $z \equiv ((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, this equation is equivalent to:

$$\chi = \left(\frac{1+z}{1+\mu^{-(\eta-\epsilon)}z}\right)\left(\frac{(\epsilon-1) + (\eta-1)\mu^{-(\eta-\epsilon)}z}{(\epsilon-1)+(\eta-1)z}\right) . \qquad \text{[A.1.40]}$$

The quantity ratio is then written as $\chi = \mathcal{D}(\rho_S)/\mathcal{D}(\rho_N)$ using the relative demand function $\mathfrak{q} = \mathcal{R}(\rho)$ from equation [A.1.1], and thus:

$$\chi = \frac{\lambda\rho_S^{-\epsilon} + (1-\lambda)\rho_S^{-\eta}}{\lambda\rho_N^{-\epsilon} + (1-\lambda)\rho_N^{-\eta}} .$$

By factorizing $\lambda\rho_S^{-\epsilon}$ and $\lambda\rho_N^{-\epsilon}$ from the numerator and denominator respectively, and using $\mu = \rho_S/\rho_N$ and the definition $z \equiv ((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, the above expression for χ becomes:

$$\chi = \mu^{-\epsilon}\left(\frac{1+\mu^{-(\eta-\epsilon)}z}{1+z}\right) . \qquad \text{[A.1.41]}$$

Putting together the two expressions for the quantity ratio χ in [A.1.40] and [A.1.41], μ and z must satisfy the equation:

$$\left(\frac{1+z}{1+\mu^{-(\eta-\epsilon)}z}\right)\left(\frac{(\epsilon-1)+(\eta-1)\mu^{-(\eta-\epsilon)}z}{(\epsilon-1)+(\eta-1)z}\right) = \mu^{-\epsilon}\left(\frac{1+\mu^{-(\eta-\epsilon)}z}{1+z}\right) . \qquad \text{[A.1.42]}$$

Since the quantity ratio χ is finite, none of the terms in the denominators of [A.1.40] or [A.1.41] can be zero, so [A.1.42] is rearranged as follows to obtain a cubic equation in z for a given value of μ:

$$\begin{aligned}(\eta-1)\mu^{-(2\eta-\epsilon)}\left(1-\mu^{\eta}\right)z^3 + \mu^{-(2\eta-\epsilon)}\left((\epsilon-1)\left(1-\mu^{2\eta-\epsilon}\right) + 2(\eta-1) + \left(\mu^{\eta-\epsilon}-\mu^{\eta}\right)\right)z^2 \\ + \mu^{-(2\eta-\epsilon)}\left((\eta-1)\left(\mu^{2(\eta-\epsilon)}-\mu^{\eta}\right) + 2(\epsilon-1)\left(\mu^{\eta-\epsilon}-\mu^{2\eta-\epsilon}\right)\right)z \\ + (\epsilon-1)\mu^{-(2\eta-\epsilon)}\left(\mu^{2(\eta-\epsilon)}-\mu^{2\eta-\epsilon}\right) = 0 .\end{aligned}$$

Because $0 < \mu < 1$, both sides of the above are multiplied by $\mu^{2\eta-\epsilon}/(1-\mu^{\eta})$ to obtain an equivalent cubic equation:

$$\begin{aligned}(\eta-1)z^3 + \left((\epsilon-1)\left(\frac{1-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right) + 2(\eta-1)\left(\frac{\mu^{\eta-\epsilon}-\mu^{\eta}}{1-\mu^{\eta}}\right)\right)z^2 \\ + \left((\eta-1)\left(\frac{\mu^{2(\eta-\epsilon)}-\mu^{\eta}}{1-\mu^{\eta}}\right) + 2(\epsilon-1)\left(\frac{\mu^{\eta-\epsilon}-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right)\right)z \\ + (\epsilon-1)\left(\frac{\mu^{2(\eta-\epsilon)}-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right) = 0 .\end{aligned} \qquad \text{[A.1.43]}$$

This cubic is denoted by $\mathfrak{C}(z;\mu,\epsilon,\eta) \equiv \mathfrak{b}_0(\mu;\epsilon,\eta) + \mathfrak{b}_1(\mu;\epsilon,\eta)z + \mathfrak{b}_2(\mu;\epsilon,\eta)z^2 + \mathfrak{b}_3(\eta)z^3$, where the coefficient functions $\mathfrak{b}_0(\mu;\epsilon,\eta)$, $\mathfrak{b}_1(\mu;\epsilon,\eta)$, $\mathfrak{b}_2(\mu;\epsilon,\eta)$ and $\mathfrak{b}_3(\eta)$ listed in [A.1.30] are obtained directly from [A.1.43].

These steps demonstrate that starting from a solution $\mathfrak{q}_S$ and $\mathfrak{q}_N$ of [A.1.16], the quadratic and the cubic equations [A.1.37] and [A.1.43] in z must simultaneously hold for $z = ((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, with $\rho_N \equiv \mathcal{D}^{-1}(\mathfrak{q}_N)$, and where the coefficient functions [A.1.30] are evaluated at $\mu = \rho_S/\rho_N$, with $\rho_S \equiv \mathcal{D}^{-1}(\mathfrak{q}_S)$. If the quadratic equation $\mathfrak{Q}(z;\mu,\epsilon,\eta) = 0$ and cubic equation $\mathfrak{C}(z;\mu,\epsilon,\eta) = 0$ share a root then it is a standard result from the theory of polynomials that the resultant $\mathfrak{R}(\mu;\epsilon,\eta)$, as defined in [A.1.31], is zero. Since the coefficients of the polynomials $\mathfrak{Q}(z;\mu,\epsilon,\eta)$ and $\mathfrak{C}(z;\mu,\epsilon,\eta)$ are functions only of the markup ratio μ and the parameters ϵ and η, solving the equation $\mathfrak{R}(\mu;\epsilon,\eta)=0$ provides a straightforward procedure for finding the equilibrium markup ratio μ. Furthermore, the only parameters appearing in the equation $\mathfrak{R}(\mu;\epsilon,\eta)=0$ are ϵ and η, so the equilibrium markup ratio μ depends only on these parameters.

Lemma 2 shows that the solution of [A.1.16] for $\mathfrak{q}_S$ and $\mathfrak{q}_N$ is unique, and therefore the solution of $\mathfrak{R}(\mu;\epsilon,\eta)=0$ for μ must also be unique, given the added condition that the shared root z of the quadratic $\mathfrak{Q}(z;\mu,\epsilon,\eta)=0$ and cubic $\mathfrak{C}(z;\mu,\epsilon,\eta)=0$ is a positive real number. This restriction is required because $z=((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, and ρ_N must of course be a positive real number. Since $\eta>\epsilon>1$, the product of the roots of the quadratic $\mathfrak{Q}(z;\mu,\epsilon,\eta)=0$ is positive, so the shared root z is positive and real if and only if either branch of the quadratic root function is positive and real. Hence this condition is verified by checking whether $\mathfrak{z}(\mu;\epsilon,\eta)$ is positive and real.

Note that the resultant $\mathfrak{R}(\mu;\epsilon,\eta)$ is always zero at $\mu=0$ and $\mu=1$ for all values of ϵ and η. This is seen by taking limits of the coefficients in [A.1.30] as $\mu\to 0$ and $\mu\to 1$ and applying L'Hôpital's rule, which yields:

$$\mathfrak{C}(z;0,\epsilon,\eta)=z\mathfrak{Q}(z;0,\epsilon,\eta)\ ,\quad \mathfrak{C}(z;1,\epsilon,\eta)=(1+z)\mathfrak{Q}(z;1,\epsilon,\eta)\ .$$

As the polynomials $\mathfrak{Q}(z;\mu,\epsilon,\eta)$ and $\mathfrak{C}(z;\mu,\epsilon,\eta)$ clearly share roots when $\mu=0$ or $\mu=1$, it follows that $\mathfrak{R}(0;\epsilon,\eta)=\mathfrak{R}(1;\epsilon,\eta)=0$. Thus these zeros of the equation $\mathfrak{R}(\mu;\epsilon,\eta)=0$ must be ignored when solving for μ.

(ii) The quadratic equation $\mathfrak{Q}(z;\mu,\epsilon,\eta)=0$ with $z=((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$ determines a relative price ρ_N such that with $\rho_S=\mu\rho_N$, marginal revenue is equalized at both ρ_S and ρ_N. Lemma 1 demonstrates that there are two candidate solutions for ρ_N that meet this criterion under the conditions shown by Lemma 2 to be necessary for a solution $\mathfrak{q}_S$ and $\mathfrak{q}_N$ of [A.1.16] to exist. However, Lemma 2 shows that both ρ_N and ρ_S are on the downward-sloping sections of the marginal revenue function. To rule out a solution in the middle upward-sloping section of marginal revenue, the smaller of the two ρ_N candidate values must be discarded to select the correct solution. Since z is decreasing in ρ_N, this is equivalent to discarding the larger of the two roots of the quadratic. Given that $\mathfrak{a}_2(\eta)$ in [A.1.30] is positive, the smaller of the two roots of quadratic $\mathfrak{Q}(z;\mu,\epsilon,\eta)=0$ is found using the expression for $\mathfrak{z}(\mu;\epsilon,\eta)$ in [A.1.32].

The equilibrium quantity ratio χ is obtained by substituting $z=\mathfrak{z}(\mu;\epsilon,\eta)$ into [A.1.41]. This construction demonstrates that χ depends only on ϵ and η.

(iii) Since $\rho_S\equiv P_S/P_B$ and $\rho_N\equiv P_N/P_B$ according to [A.1.2], the formula for the purchase multipliers in [3.8] implies $v_N=\rho_N^{\epsilon-\eta}$ and $v_S=\mu^{\epsilon-\eta}v_N$. Using the fact that $z\equiv((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, and dividing numerator and denominator of the expression in [4.5] by λ yields [A.1.34].

(iv) The expressions for the relative prices ρ_S and ρ_N in [A.1.35] are obtained by rearranging the definition of $z\equiv((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$ and using $\rho_S=\mu\rho_N$. This completes the proof. ■

A.2 Proof of Proposition 2

Using the relationship between the total revenue function $\mathscr{R}(q;P_B,\mathcal{E})$ and its equivalent $\mathcal{R}(\mathfrak{q})$ defined in [A.1.4] using the relative demand function $\mathcal{D}(\rho)$ from [A.1.1], the corresponding marginal revenue functions $\mathscr{R}'(q;P_B,\mathcal{E})$ and $\mathcal{R}'(\mathfrak{q})$ are proportional according to [A.1.7]. Lemma 1 demonstrates that $\mathcal{R}'(\mathfrak{q})$ has the described pattern of non-monotonicity under the conditions $0<\lambda<1$ and [4.3], and is otherwise a decreasing function of q. This completes the proof.

A.3 Proof of Theorem 1

Existence of a two-price equilibrium

For a two-price equilibrium to exist, first-order conditions [4.4] for profit-maximization must be satisfied at two prices p_S and p_N, with associated quantities $q_S=\mathscr{D}(p_S;P_B,\mathcal{E})$ and $q_N=\mathscr{D}(p_N;P_B,\mathcal{E})$, where P_B is the bargain hunters' price index from [3.4], and $\mathcal{E}=P^\epsilon Y$ is the measure of aggregate expenditure from [3.7].

The necessary conditions for the two-price equilibrium are now restated in terms of the relative demand function $\mathcal{D}(\rho)$ defined in [A.1.1], and its associated total and marginal revenue functions $\mathcal{R}(\mathfrak{q})$ and $\mathcal{R}'(\mathfrak{q})$, as defined in [A.1.4] and analysed in section A.1. The relative demand function $\mathfrak{q} = \mathcal{D}(\rho)$ is specified in terms of the relative price $\rho \equiv p/P_B$ and relative quantity $\mathfrak{q} \equiv q/(\mathcal{E}/P_B^\epsilon)$, in accordance with [A.1.2]. Using the relationships in [A.1.3] and [A.1.7], the first two optimality conditions in [4.4] are equivalent to:

$$\mathcal{R}'\left(\frac{q_S P_B^\epsilon}{\mathcal{E}}\right) = \mathcal{R}'\left(\frac{q_N P_B^\epsilon}{\mathcal{E}}\right) = \frac{\mathcal{R}\left(\frac{q_S P_B^\epsilon}{\mathcal{E}}\right) - \mathcal{R}\left(\frac{q_N P_B^\epsilon}{\mathcal{E}}\right)}{\frac{q_S P_B^\epsilon}{\mathcal{E}} - \frac{q_N P_B^\epsilon}{\mathcal{E}}} \ . \qquad \text{[A.3.1]}$$

With $\mathfrak{q}_S \equiv q_S/(\mathcal{E}/P_B^\epsilon)$ and $\mathfrak{q}_N \equiv q_N/(\mathcal{E}/P_B^\epsilon)$, the first-order conditions in [A.3.1] are identical to the equations in [A.1.16] studied in Lemma 2. These clearly require the equalization of marginal revenue $\mathcal{R}'(\mathfrak{q})$ at two different quantities, which means that the marginal revenue function must be non-monotonic. Lemma 1 then shows that $0 < \lambda < 1$ and parameters ϵ and η satisfying the inequality [4.3] are necessary and sufficient for this. If these conditions are met then Lemma 2 demonstrates the existence of a unique solution $\mathfrak{q}_S$ and $\mathfrak{q}_N$ of the equations [A.1.16].

The relative quantities $\mathfrak{q}_S$ and $\mathfrak{q}_N$ must also be well defined if the solution is to be economically meaningful. This means that if $\rho_S = \mathcal{D}^{-1}(\mathfrak{q}_S)$ and $\rho_N = \mathcal{D}^{-1}(\mathfrak{q}_N)$ are the corresponding prices p_S and p_N relative to P_B then $\rho_S < 1 < \rho_N$. This is a necessary requirement because the expression [4.8] for the bargain hunters' price index P_B implies:

$$s\rho_S^{1-\eta} + (1-s)\rho_N^{1-\eta} = 1 \ , \qquad \text{[A.3.2]}$$

and the equilibrium sales fraction s must satisfy $s \in (0, 1)$.

Assume the parameters are such that ϵ and η satisfy [4.3], and consider a given value of $\lambda \in (0, 1)$. Lemma 3 shows that the markup ratio (or price ratio) $\mu \equiv \mu_S/\mu_N = \rho_S/\rho_N$ consistent with the unique solution of [A.1.16] is a function only of the elasticities ϵ and η. The equilibrium relative prices ρ_S and ρ_N are functions of all three parameters ϵ, η and λ, and are obtained from equation [A.1.35] by substituting the equilibrium value of μ into the function $\mathfrak{z}(\mu; \epsilon, \eta)$ defined in [A.1.32]. Since $\rho_S = \mu\rho_N$ and $\mu < 1$, the requirement $\rho_S < 1 < \rho_N$ implies $\mu < \rho_S < 1$. By substituting for ρ_S from [A.1.35], this condition is equivalent to:

$$\mathfrak{z}(\mu; \epsilon, \eta) < \frac{1-\lambda}{\lambda} < \mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu; \epsilon, \eta) \ . \qquad \text{[A.3.3]}$$

Define lower and upper bounds for λ conditional on ϵ and η using the function $\mathfrak{z}(\mu; \epsilon, \eta)$ together with the equilibrium value of μ (which is a function only of ϵ and η):

$$\underline{\lambda}(\epsilon, \eta) \equiv \frac{1}{1 + \mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu; \epsilon, \eta)} \ , \quad \text{and} \quad \overline{\lambda}(\epsilon, \eta) \equiv \frac{1}{1 + \mathfrak{z}(\mu; \epsilon, \eta)} \ . \qquad \text{[A.3.4]}$$

Note that if $\mathfrak{z}(\mu; \epsilon, \eta) > 0$ and $0 < \mu < 1$ then $0 < \underline{\lambda}(\epsilon, \eta) < \overline{\lambda}(\epsilon, \eta) < 1$. By rearranging the inequality [A.3.3] and using the above definitions of the bounds on λ, the inequality is equivalent to λ lying in the interval:

$$\underline{\lambda}(\epsilon, \eta) < \lambda < \overline{\lambda}(\epsilon, \eta) \ . \qquad \text{[A.3.5]}$$

This restriction on λ is necessary and sufficient for the existence of an equilibrium sales fraction $s \in (0, 1)$ satisfying [A.3.2]. The equivalence is demonstrated by substituting the expressions for ρ_S and ρ_N from [A.1.35] into [A.3.2]:

$$\left(1 + s\left(\mu^{-(\eta-1)} - 1\right)\right)\left(\frac{\lambda}{1-\lambda}\mathfrak{z}(\mu; \epsilon, \eta)\right)^{\frac{\eta-1}{\eta-\epsilon}} = 1 \ .$$

This is a linear equation in s, and has a unique solution because $\eta > 1$ and $0 < \mu < 1$. Solving explicitly for s yields:

$$s = \frac{\left(\frac{\lambda}{1-\lambda}\mathfrak{z}(\mu; \epsilon, \eta)\right)^{-\left(\frac{\eta-1}{\eta-\epsilon}\right)} - 1}{\mu^{-(\eta-1)} - 1} \ . \qquad \text{[A.3.6]}$$

Recalling the equivalence of inequalities [A.3.3] and [A.3.5], it follows that $s \in (0, 1)$ if and only if $\lambda \in$

$(\underline{\lambda}(\epsilon,\eta),\overline{\lambda}(\epsilon,\eta))$. So for $\lambda \in [0,\underline{\lambda}(\epsilon,\eta)]$ or $\lambda \in [\overline{\lambda}(\epsilon,\eta),1]$ there is no two-price equilibrium. But given elasticities ϵ and η satisfying the non-monotonicity condition [4.3] and a loyal fraction $\lambda \in (\underline{\lambda}(\epsilon,\eta),\overline{\lambda}(\epsilon,\eta))$, by using the arguments above there exist two distinct relative prices $\rho_S \equiv p_S/P_B$ and $\rho_N \equiv p_N/P_B$ and a sales fraction $s \in (0,1)$ consistent with the first two equalities in [4.4]. Lemma 3 then demonstrates that the two purchase multipliers v_S and v_N and the two optimal markups μ_S and μ_N are determined. Equations [4.1], [4.2] and [4.5] show that using the optimal markups in [4.6] is equivalent to satisfying the remaining first-order condition involving marginal cost in [4.4]. The other variables relevant to the macroeconomic equilibrium are then determined as discussed in section 4.

Confirming that the two-price equilibrium exists then requires checking that the remaining first-order condition [3.11c] is satisfied and that the first-order conditions are sufficient as well as necessary to characterize the maximum of the profit function. Using the relationships in [A.1.7] and the results of Lemma 2 in [A.1.17] the following inequalities are deduced:

$$\mathscr{R}(q_S;P_B,\mathcal{E})-\mathscr{R}'(q_S;P_B,\mathcal{E})q_S>0\ , \quad \text{and } \mathscr{R}(q_N;P_B,\mathcal{E})-\mathscr{R}'(q_N;P_B,\mathcal{E})q_N>0\ . \tag{A.3.7}$$

Since $s \in (0,1)$, the Lagrangian multiplier $\aleph$ from first-order conditions [3.11b]–[3.11c] is determined as follows:

$$\aleph=\mathscr{R}(q_S;P_B,\mathcal{E})-\mathscr{R}'(q_S;P_B,\mathcal{E})q_S=\mathscr{R}(q_N;P_B,\mathcal{E})-\mathscr{R}'(q_N;P_B,\mathcal{E})q_N\ ,$$

and hence $\aleph>0$ because of [A.3.7]. By combining this expression for the Lagrangian multiplier with the first-order condition [3.11c]:

$$\mathscr{R}(q;P_B,\mathcal{E})\leq\mathscr{R}(q_N;P_B,\mathcal{E})+\mathscr{R}'(q_N;P_B,\mathcal{E})(q-q_N)=\mathscr{R}(q_S;P_B,\mathcal{E})+\mathscr{R}'(q_S;P_B,\mathcal{E})(q-q_S)\ , \tag{A.3.8}$$

which is required to hold for all $q \geq 0$. Appealing to the result of Lemma 2 in [A.1.18] and again using [A.1.7] verifies the inequality.

The assumptions about the production function [2.5] ensure that the total cost function $\mathscr{C}(Q;W)$ in [2.6] is continuously differentiable and convex, so for all $q \geq 0$:

$$\mathscr{C}(q;W)\geq\mathscr{C}(Q;W)+\mathscr{C}'(Q;W)(q-Q)\ , \tag{A.3.9}$$

where $Q \equiv sq_S+(1-s)q_N$ is the specific total physical quantity sold using the two-price strategy constructed earlier. Now consider a general alternative pricing strategy for a given firm, assuming that all other firms continue to use the same two-price strategy. The new strategy is specified in terms of a distribution function $F(p)$ for prices. Let $G(q) \equiv 1-F\left(\mathscr{D}(p;P_B,\mathcal{E})\right)$ be the implied distribution function for quantities sold. The level of profits $\mathscr{P}$ from the new strategy is obtained by making a change of variable from prices to quantities in the integrals of [3.10]:

$$\mathscr{P}=\int_q\mathscr{R}(q;P_B,\mathcal{E})dG(q)-\mathscr{C}\left(\int_q qdG(q);W\right)\ .$$

Applying the inequalities involving the total revenue and total cost functions from [A.3.8] and [A.3.9] to the expression for profits yields:

$$\mathscr{P}\leq\left(\mathscr{R}(q_N;P_B,\mathcal{E})-\mathscr{R}'(q_N;P_B,\mathcal{E})q_N\right)-\left(\mathscr{C}(Q;W)-\mathscr{C}'(Q;W)Q\right) \\ +\left(\mathscr{R}'(q_N;P_B,\mathcal{E})-\mathscr{C}'(Q;W)\right)\left(\int_q qdG(q)\right)\ .$$

The first-order conditions [4.4] imply that the coefficient of the integral in the above expression is zero, and that $\mathscr{R}(q_N;P_B,\mathcal{E})-\mathscr{R}'(q_N;P_B,\mathcal{E})q_N=\mathscr{R}(q_S;P_B,\mathcal{E})-\mathscr{R}'(q_S;P_B,\mathcal{E})q_S$. Recalling $Q=sq_S+(1-s)q_N$, it follows that:

$$\mathscr{P}\leq s\mathscr{R}(q_S;P_B,\mathcal{E})+(1-s)\mathscr{R}(q_N;P_B,\mathcal{E})-\mathscr{C}\left(sq_S+(1-s)q_N;W\right)\ ,$$

for all alternative pricing strategies. Hence there is no profit-improving deviation from the two-price strategy. This establishes that a two-price equilibrium exists when [4.3] and $\lambda \in (\underline{\lambda}(\epsilon,\eta),\overline{\lambda}(\epsilon,\eta))$ hold, and that it is unique within the class of two-price equilibria.

Uniqueness of the two-price equilibrium

Suppose the parameters ϵ, η and λ are such that a two-price equilibrium exists. Now consider the possibility that a one-price equilibrium also exists for the same parameters. Since all firms are symmetric, the relative price found in this one-price equilibrium is necessarily equal to one. The relative prices ρ_S and ρ_N in the two-price equilibrium cannot be on the same side of one, implying $\mu < \rho_S < 1$ and thus $\rho_S < 1 < \rho_N$, where $\rho_S = \mathcal{D}^{-1}(\mathfrak{q}_S)$ and $\rho_N = \mathcal{D}^{-1}(\mathfrak{q}_N)$ using the relative quantities $\mathfrak{q}_S$ and $\mathfrak{q}_N$. Since [A.1.1] implies $\mathcal{D}(1) = 1$ and because the relative demand function $\mathcal{D}(\rho)$ is strictly decreasing in ρ, it follows that $\mathfrak{q}_N < 1 < \mathfrak{q}_S$.

Given that marginal revenue must be non-monotonic if a two-price equilibrium is to exist, it follows from Lemma 1 that $\mathcal{R}(\mathfrak{q})$ is strictly concave on the intervals $(0, \underline{\mathfrak{q}})$ and $(\bar{\mathfrak{q}}, \infty)$, strictly convex on $(\underline{\mathfrak{q}}, \bar{\mathfrak{q}})$, and from Lemma 2 that $\mathfrak{q}_N < \underline{\mathfrak{q}} < \bar{\mathfrak{q}} < \mathfrak{q}_S$.

Consider first the case where $\underline{\mathfrak{q}} < 1 < \bar{\mathfrak{q}}$. Since $\mathfrak{q}_1 = 1$ for all firms in the one-price equilibrium, the actual common quantity produced is $q_1 = \mathcal{E}/P_B^\epsilon$ using [A.1.2], where P_B and $\mathcal{E}$ are the values of these variables associated with the putative one-price equilibrium. Since $\mathcal{R}''(1) > 0$, equation [A.1.7] implies $\mathscr{R}''(q_1; P_B, \mathcal{E}) > 0$. Therefore, for sufficiently small $\xi > 0$, the profits $\mathscr{P}$ from selling quantity $q_1 - \xi$ at one half of shopping moments and $q_1 + \xi$ at the other half exceed the profits from offering one price and hence one quantity at all shopping moments:

$$\frac{1}{2}\mathscr{R}(q_1 - \xi; P_B, \mathcal{E}) + \frac{1}{2}\mathscr{R}(q_1 + \xi; P_B, \mathcal{E}) - \mathscr{C}\left(\frac{1}{2}(q_1 - \xi) + \frac{1}{2}(q_1 + \xi); W\right) > \mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{C}(q_1; W)\,.$$

Therefore a one-price equilibrium cannot exist in this case.

Next consider the case where $\mathfrak{q}_N < 1 < \underline{\mathfrak{q}}$. Let $p_1 = P_B$ denote the price it is claimed all firms charge in a one-price equilibrium, and $q_1 = \mathcal{E}/P_B^\epsilon$ the associated quantity sold. Now let $q_S = \mathscr{D}(\rho_S p_1; P_B, \mathcal{E})$ be quantity sold if the sale *relative* price $\rho_S = \mathcal{D}^{-1}(\mathfrak{q}_S)$ is used when other firms are following the one-price strategy of charging p_1 at all shopping moments. Consider an alternative strategy where price $\rho_S p_1$ is offered at a fraction ξ of moments and price p_1 at the remaining fraction $1-\xi$ of moments. Profits $\mathscr{P}$ from the hybrid strategy are given by:

$$\mathscr{P} = (1-\xi)\mathscr{R}(q_1; P_B, \mathcal{E}) + \xi\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{C}\left((1-\xi)q_1 + \xi q_S; W\right)\,. \qquad \text{[A.3.10]}$$

As the cost function $\mathscr{C}(q; W)$ is differentiable in q, the above equation implies:

$$\mathscr{P} = (\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{C}(q_1; W)) + \xi(q_S - q_1)\left(\frac{\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}(q_1; P_B, \mathcal{E})}{q_S - q_1} - \mathscr{C}'(q_1; W)\right) + \mathscr{O}\left(\xi^2\right)\,,$$

where $\mathscr{O}\left(\xi^2\right)$ denotes second- and higher-order terms in ξ. A necessary condition for a one-price equilibrium to exist is that the single price p_1 is chosen optimally, in which case first-order conditions [3.11] reduce to the usual marginal revenue equals marginal cost condition $\mathscr{R}'(q_1; P_B, \mathcal{E}) = \mathscr{C}'(q_1; W)$. Hence the above expression for $\mathscr{P}$ becomes:

$$\mathscr{P} = (\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{C}(q_1; W)) + \xi(q_S - q)\left(\frac{\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}(q_1; P_B, \mathcal{E})}{q_S - q_1} - \mathscr{R}'(q_1; P_B, \mathcal{E})\right) + \mathscr{O}\left(\xi^2\right)\,. \qquad \text{[A.3.11]}$$

Since $\mathfrak{q}_N < 1 < \mathfrak{q}_S$ in the case under consideration and $\mathfrak{q}_1 = 1$, the results from Lemma 2 in [A.1.16] can be expressed as follows:

$$\int_{\mathfrak{q}_N}^{1} \mathcal{R}'(\mathfrak{q})d\mathfrak{q} + \mathcal{R}(\mathfrak{q}_S) - \mathcal{R}(\mathfrak{q}_1) = (\mathfrak{q}_S - \mathfrak{q}_N)\mathcal{R}'(\mathfrak{q}_N)\,. \qquad \text{[A.3.12]}$$

As $\mathfrak{q}_N < 1 < \underline{\mathfrak{q}}$ and $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing for $\mathfrak{q} < \underline{\mathfrak{q}}$, the integral above satisfies:

$$\int_{\mathfrak{q}_N}^{1} \mathcal{R}'(\mathfrak{q})d\mathfrak{q} < (1 - \mathfrak{q}_N)\mathcal{R}'(\mathfrak{q}_N)\,. \qquad \text{[A.3.13]}$$

Noting that $\mathcal{R}'(\mathfrak{q}_N) > \mathcal{R}'(1)$ because of $\mathfrak{q}_N < 1 < \underline{\mathfrak{q}}$, and substituting [A.3.13] into [A.3.12] and rearranging yields:

$$\frac{\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}(1)}{\mathfrak{q}_S - 1} > \mathcal{R}'(\mathfrak{q}_N) > \mathcal{R}'(1) \ , \tag{A.3.14}$$

where $\mathfrak{q}_S > 1$ ensures that the direction of the inequality is preserved. Now given the fact that $q_1 = (\mathcal{E}/P_B^\epsilon)$ and $q_S = (\mathcal{E}/P_B^\epsilon)\mathfrak{q}_S$ from [A.1.2], and the links between the functions $\mathcal{R}(\mathfrak{q})$ and $\mathscr{R}(q; P_B, \mathcal{E})$ as set out in [A.1.7]:

$$\frac{\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}(q_1; P_B, \mathcal{E})}{q_S - q_1} > \mathscr{R}'(q_1; P_B, \mathcal{E}) \ . \tag{A.3.15}$$

Therefore, by comparing this inequality with [A.3.11] and noting $q_S > q_1$, it follows for sufficiently small $\xi > 0$ that $\mathscr{P} > \mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{C}(q_1; W)$, so profits from a hybrid strategy exceed those from following the strategy required for the one-price equilibrium to exist.

The remaining case to consider is $\bar{\mathfrak{q}} < 1 < \mathfrak{q}_S$. The argument here is analogous to that given above. The alternative strategy considered is offering price $p_N = \rho_N p_1$ (where $\rho_N = \mathcal{D}^{-1}(\mathfrak{q}_N)$) at a fraction ξ of shopping moments and price $p_1 = P_B$ at the remaining fraction $1 - \xi$, with quantities sold respectively at those moments of $q_N = \mathscr{D}(\rho_N p_1; P_B, \mathcal{E})$ and q_1. Following the steps in [A.3.10]–[A.3.11] leads to an expression for profits $\mathscr{P}$ resulting from this hybrid strategy:

$$\mathscr{P} = (\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{C}(q_1; W)) + \xi(q_1 - q_N)\left(\mathscr{R}'(q_1; P_B, \mathcal{E}) - \frac{\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{R}(q_N; P_B, \mathcal{E})}{q_1 - q_N}\right) + \mathscr{O}\left(\xi^2\right) \ . \tag{A.3.16}$$

Appealing to the properties of $\mathcal{R}(\mathfrak{q})$ for $\mathfrak{q} > \bar{\mathfrak{q}}$ and following similar steps to those in [A.3.12]–[A.3.14] implies $\mathcal{R}'(1) > \mathcal{R}'(\mathfrak{q}_S) > (\mathcal{R}(1) - \mathcal{R}(\mathfrak{q}_N))/(1 - \mathfrak{q}_N)$, and hence an equivalent of [A.3.15]:

$$\mathscr{R}'(q_1; P_B, \mathcal{E}) > \frac{\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{R}(q_N; P_B, \mathcal{E})}{q_1 - q_N} \ . \tag{A.3.17}$$

Since $q_1 > q_N$, for sufficiently small $\xi > 0$, [A.3.16] and [A.3.17] demonstrate that there is a hybrid strategy which delivers higher profits than the one-price strategy used by all other firms. This proves that for all parameters where the two-price equilibrium exists, a one-price equilibrium cannot exist for any of these same parameter values.

One-price equilibrium

The first point to note is that when a two-price equilibrium fails to exist owing to a violation of the non-monotonicity condition [4.3], Lemma 1 implies that marginal revenue $\mathscr{R}'(q; P_B, \mathcal{E})$ is strictly decreasing for all q. This is equivalent to total revenue $\mathscr{R}(q; P_B, \mathcal{E})$ being a strictly concave function of quantity q. Since total cost $\mathscr{C}(q; W)$ is a convex function of the quantity produced, it follows immediately that the profit function is globally concave, and thus a one-price equilibrium always exists, and is the only possible equilibrium in the parameter range where ϵ or η fail to satisfy [4.3], or where $\lambda = 0$ or $\lambda = 1$.

Now suppose the parameters are such that the marginal revenue function is non-monotonic, but a two-price equilibrium fails to exist owing to λ not lying between $\underline{\lambda}(\epsilon, \eta)$ and $\overline{\lambda}(\epsilon, \eta)$. Note that [A.3.3] and [A.3.4] imply $\lambda \in [0, \underline{\lambda}(\epsilon, \eta)]$ and $\lambda \in [\overline{\lambda}(\epsilon, \eta), 1]$ are equivalent to $1 > \mathfrak{q}_S$ and $1 < \mathfrak{q}_N$ respectively.

Taking the first of these cases, Lemma 1 demonstrates the concavity of $\mathcal{R}(\mathfrak{q})$ on $[\bar{\mathfrak{q}}, \infty)$ (containing $\mathfrak{q}_S$), which establishes that $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(1) + \mathcal{R}'(1)(\mathfrak{q} - 1)$ for all $\mathfrak{q} \in [\bar{\mathfrak{q}}, \infty)$. Lemma 2 shows that $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_S) + \mathcal{R}'(\mathfrak{q}_S)(\mathfrak{q} - \mathfrak{q}_S)$ for all $\mathfrak{q} \geq 0$. Note that the concavity of $\mathcal{R}(\mathfrak{q})$ in the relevant range implies $\mathcal{R}'(\mathfrak{q}_S) > \mathcal{R}'(1)$, which together with the second of the previous inequalities yields $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_S) + \mathcal{R}'(1)(\mathfrak{q} - \mathfrak{q}_S)$ for all $\mathfrak{q} \in [0, \mathfrak{q}_S]$. Applying the first inequality at $\mathfrak{q} = \mathfrak{q}_S$ establishes that $\mathcal{R}(\mathfrak{q}_S) \leq \mathcal{R}(1) + \mathcal{R}'(1)(\mathfrak{q}_S - 1)$. By combining these results it follows that $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(1) + \mathcal{R}'(1)(\mathfrak{q} - 1)$ for all $\mathfrak{q} \geq 0$. Translating this into a property of the original total revenue function $\mathscr{R}(q; P_B, \mathcal{E})$ using [A.1.2] and [A.1.7] yields the following for all q:

$$\mathscr{R}(q; P_B, \mathcal{E}) \leq \mathscr{R}(q_1; P_B, \mathcal{E}) + \mathscr{R}'(q_1; P_B, \mathcal{E})(q - q_1) \ . \tag{A.3.18}$$

When $\lambda \in [\overline{\lambda}(\epsilon, \eta), 1]$ the other case to consider is $1 < \mathfrak{q}_N$. Using an exactly analogous argument to that given above, it is deduced that $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(1) + \mathcal{R}'(1)(\mathfrak{q} - 1)$ for all $\mathfrak{q} \geq 0$ in this case as well. Hence [A.3.18] holds in both cases. The convexity of the total cost function $\mathscr{C}(q; W)$ together with [A.3.18] proves that no pricing strategy can improve on that used in the one-price equilibrium.

Non-existence of equilibria with more than two prices

Take any two prices p_1 and p_2 offered by a firm at a positive fraction of shopping moments, and define $\rho_1 \equiv p_1/P_B$ and $\rho_2 \equiv p_2/P_B$ in accordance with [A.1.2]. Denote the quantities sold by q_1 and q_2 and define $\mathfrak{q}_1 \equiv (P_B^\epsilon/\mathcal{E})q_1$ and $\mathfrak{q}_2 \equiv (P_B^\epsilon/\mathcal{E})q_2$ also in accordance with [A.1.2]. Using the first-order conditions [3.11] together with [A.1.2] and [A.1.7], it follows that $\mathfrak{q}_1$ and $\mathfrak{q}_2$ must satisfy the system of equations [A.1.16] in place of $\mathfrak{q}_S$ and $\mathfrak{q}_N$. But as Lemma 2 demonstrates that the solution to this system of equations is unique, there is a maximum of two distinct prices in any firm's profit-maximizing strategy. This completes the proof.

A.4 Proof of Proposition 3

(i) Lemma 3 shows that μ and χ are uniquely determined as functions of ϵ and η when the inequality [4.3] is satisfied, as is necessary for the two-price equilibrium to exist. Lemma 3 also gives solutions for μ_S and μ_N, and implicitly determines the purchase multipliers v_S and v_N using the expressions for ρ_S and ρ_N in [A.1.35] and the fact that $v_S = (p_S/P_B)^{-(\eta-\epsilon)}$ and $v_N = (p_N/P_B)^{-(\eta-\epsilon)}$ from [3.8]. Hence Lemma 3 shows that these variables depend only on ϵ, η and λ. In conjunction with equation [4.8], knowledge of ρ_S and ρ_N from [A.1.35] yields a linear equation for s after dividing both sides of the equation by P_B.

(ii) Lemma 3 shows that μ, μ_S, μ_N and χ are independent of λ, hence verifying the claim.

(iii) Substituting the bounds for λ from [A.3.4] into equation [A.3.6] proves the first two results. Differentiating [A.3.6] with respect to λ yields the third result.

(iv) The markup ratio μ is characterized implicitly as a root of the function $\mathfrak{R}(\mu;\epsilon,\eta) = 0$ from [A.1.31]. This is a determinant of a matrix whose elements are continuous functions of μ, ϵ and η. Therefore, μ is a continuous function of ϵ and η.

Let $\underline{z}$ and $\overline{z}$ be the roots of the quadratic $\mathcal{Q}(z) = 0$ from [A.1.11] of Lemma 1. Taking the limit as $\epsilon \to 1^+$ yields $\underline{z} \to 0$ and $\overline{z} \to (\eta-2)/\eta$. Note that $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$ from Lemma 1 are related to $\underline{z}$ and $\overline{z}$ through the transformation $\mathcal{Z}^{-1}(z)$ from [A.1.15], which is strictly increasing. Now define z_S and z_N as follows in terms of relative prices ρ_S and ρ_N:

$$z_S \equiv \frac{1-\lambda}{\lambda}\rho_S^{\epsilon-\eta}, \quad z_N \equiv \frac{1-\lambda}{\lambda}\rho_N^{\epsilon-\eta}. \qquad \text{[A.4.1]}$$

Lemma 2 shows that $\mathfrak{q}_N < \underline{\mathfrak{q}} < \overline{\mathfrak{q}} < \mathfrak{q}_S$, and hence $z_N < \underline{z} < \overline{z} < z_S$ using the monotonicity of the $\mathcal{Z}^{-1}(z)$ transformation from [A.1.15]. It follows from these inequalities and the definitions in [A.4.1] that $\mu = \rho_S/\rho_N$ must satisfy:

$$\mu = \left(\frac{z_N}{z_S}\right)^{\frac{1}{\eta-\epsilon}} < \left(\underline{z}/\overline{z}\right)^{\frac{1}{\eta-\epsilon}}.$$

So as $\epsilon \to 1^+$, μ converges to zero. Then note that χ is determined by [A.1.33] with $\mathfrak{z}(\mu;\epsilon,\eta) = z_N$, and hence $\chi \to \infty$ as $\epsilon \to 1^+$.

The proof of Lemma 1 shows that the value of the function $\mathcal{G}_r(\eta;\epsilon)$ from [A.1.12] tends to zero as $\eta \to \eta^*(\epsilon)$. This implies the discriminant of the quadratic $\mathcal{Q}(z)$ in [A.1.13] tends to zero. Therefore, the roots $\underline{z}$ and $\overline{z}$ of $\mathcal{Q}(z) = 0$ converge to some common point. Given the continuity of the transformation $\mathcal{Z}^{-1}(z)$, it follows that $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$ must also converge to a common point $\mathfrak{q}_0$. Thus in the limit, $\mathcal{R}''(\mathfrak{q}) < 0$ everywhere except at $\mathfrak{q} = \mathfrak{q}_0$. At each stage in approaching this limit, $\mathcal{R}'(\mathfrak{q}_S) = \mathcal{R}'(\mathfrak{q}_N)$ holds, and therefore it follows that $\mathfrak{q}_S \to \mathfrak{q}_N$, and consequently χ converges to one. The continuity of the demand function $\mathcal{D}(\rho)$ implies that $\rho_S \to \rho_N$ and so μ also converges to one. This completes the proof.

A.5 Log linearizations

A.5.1 Sales model from section 5

The notational convention adopted here is that a bar above a variable denotes its flexible-price steady-state value as determined in section 4, and the corresponding sans serif letter denotes the log deviation of the variable from its steady-state value (except for the sales fraction s, where it denotes just the deviation from steady state).

Consider first the demand function faced by firms. In the following, P_S and P_N denote the common sale and normal prices p_S and p_N chosen by all firms, or in subsequent extensions of the model, the averages of the normal prices and the sale prices chosen by firms. The levels of demand q_S and q_N at the sale and normal prices are obtained from [4.10], which has the following log-linearized form:

$$\mathsf{q}_S = \left(\frac{(1-\lambda)\bar{v}_S}{\lambda + (1-\lambda)\bar{v}_S}\right)\mathsf{v}_S - \epsilon(\mathsf{P}_S - \mathsf{P}) + \mathsf{Y} \,, \tag{A.5.1a}$$

$$\mathsf{q}_N = \left(\frac{(1-\lambda)\bar{v}_N}{\lambda + (1-\lambda)\bar{v}_N}\right)\mathsf{v}_N - \epsilon(\mathsf{P}_N - \mathsf{P}) + \mathsf{Y} \,, \tag{A.5.1b}$$

and where the expressions are given in terms of log deviations of the purchase multipliers v_S and v_N from [3.8]:

$$\mathsf{v}_S = -(\eta - \epsilon)\left(\mathsf{P}_S - \mathsf{P}_B\right) \,, \quad \mathsf{v}_N = -(\eta - \epsilon)\left(\mathsf{P}_N - \mathsf{P}_B\right) \,. \tag{A.5.2}$$

By substituting the purchase multipliers into the demand functions [A.5.1]:

$$\mathsf{q}_S = -\left(\frac{\lambda\epsilon + (1-\lambda)\eta\bar{v}_S}{\lambda + (1-\lambda)\bar{v}_S}\right)\mathsf{P}_S + (\eta - \epsilon)\left(\frac{(1-\lambda)\bar{v}_S}{\lambda + (1-\lambda)\bar{v}_S}\right)\mathsf{P}_B + \epsilon\mathsf{P} + \mathsf{Y} \,, \tag{A.5.3a}$$

$$\mathsf{q}_N = -\left(\frac{\lambda\epsilon + (1-\lambda)\eta\bar{v}_N}{\lambda + (1-\lambda)\bar{v}_N}\right)\mathsf{P}_N + (\eta - \epsilon)\left(\frac{(1-\lambda)\bar{v}_N}{\lambda + (1-\lambda)\bar{v}_N}\right)\mathsf{P}_B + \epsilon\mathsf{P} + \mathsf{Y} \,. \tag{A.5.3b}$$

From equation [4.5], the log-linearized optimal markups at given sale and normal prices are:

$$\mu_S = -\mathfrak{c}_S\mathsf{v}_S \,, \quad \text{with} \;\; \mathfrak{c}_S \equiv \frac{\lambda(1-\lambda)(\eta-\epsilon)\bar{v}_S}{\left(\lambda\epsilon + (1-\lambda)\eta\bar{v}_S\right)\left(\lambda(\epsilon-1) + (1-\lambda)(\eta-1)\bar{v}_S\right)} \,, \tag{A.5.4a}$$

$$\mu_N = -\mathfrak{c}_N\mathsf{v}_N \,, \quad \text{with} \;\; \mathfrak{c}_N \equiv \frac{\lambda(1-\lambda)(\eta-\epsilon)\bar{v}_N}{\left(\lambda\epsilon + (1-\lambda)\eta\bar{v}_N\right)\left(\lambda(\epsilon-1) + (1-\lambda)(\eta-1)\bar{v}_N\right)} \,, \tag{A.5.4b}$$

which are given in terms of the purchase multipliers from [A.5.2]. Overall demand $Q = sq_S + (1-s)q_N$ is log-linearized as follows:

$$\mathsf{Q} = \left(\frac{\bar{q}_S - \bar{q}_N}{\bar{s}\bar{q}_S + (1-\bar{s})\bar{q}_N}\right)\mathsf{s} + \left(\frac{\bar{s}\bar{q}_S}{\bar{s}\bar{q}_S + (1-\bar{s})\bar{q}_N}\right)\mathsf{q}_S + \left(\frac{(1-\bar{s})\bar{q}_N}{\bar{s}\bar{q}_S + (1-\bar{s})\bar{q}_N}\right)\mathsf{q}_N \,. \tag{A.5.5}$$

The bargain hunters' price index P_B as given in [4.8] (and its later generalizations) is log-linearized as follows:

$$\mathsf{P}_B = \theta_B\mathsf{P}_S + (1-\theta_B)\mathsf{P}_N - \psi_B\mathsf{s} \,, \quad \text{where} \tag{A.5.6}$$

$$\theta_B \equiv \left(\frac{\bar{s}}{\bar{s} + (1-\bar{s})\bar{\mu}^{\eta-1}}\right) \,, \quad \text{and} \;\; \psi_B \equiv \frac{1}{\eta-1}\left(\frac{1-\bar{\mu}^{\eta-1}}{\bar{s} + (1-\bar{s})\bar{\mu}^{\eta-1}}\right) \,,$$

where s is the average deviation of firms' sales fractions. Similarly, the log-linearized aggregate price level P from [3.3] (and its later generalizations) is:

$$\begin{aligned}\mathsf{P} = {} & \bar{s}(\lambda + (1-\lambda)\bar{v}_S)\bar{\varrho}_S^{1-\epsilon}\mathsf{P}_S + (1-\bar{s})(\lambda + (1-\lambda)\bar{v}_N)\bar{\varrho}_N^{1-\epsilon}\mathsf{P}_N - \left(\frac{1-\lambda}{\epsilon-1}\right)\bar{s}\bar{v}_S\bar{\varrho}_S^{1-\epsilon}\mathsf{v}_S \\ & - \left(\frac{1-\lambda}{\epsilon-1}\right)(1-\bar{s})\bar{v}_N\bar{\varrho}_N^{1-\epsilon}\mathsf{v}_N - \frac{1}{\epsilon-1}\left((\lambda + (1-\lambda)\bar{v}_S)\bar{\varrho}_S^{1-\epsilon} - (\lambda + (1-\lambda)\bar{v}_N)\bar{\varrho}_N^{1-\epsilon}\right)\mathsf{s} \,.\end{aligned}$$

Using the expressions for the purchase multipliers and relative prices in the flexible-price equilibrium together with the log deviations of v_S and v_N from [A.5.2], and the expression for P_B in [A.5.6], the log-linearized aggregate price level is written as follows:

$$\mathsf{P} = \theta_P \mathsf{P}_S + (1-\theta_P)\mathsf{P}_N - \psi_P \mathsf{s}\,, \quad \text{where} \qquad \text{[A.5.7]}$$

$$\theta_P \equiv \bar{s}(\lambda + (1-\lambda)\bar{v}_S)\bar{\varrho}_S^{1-\epsilon}\,, \quad \text{and} \;\; \psi_P \equiv \frac{\lambda}{\epsilon - 1}\left(\bar{\varrho}_S^{1-\epsilon} - \bar{\varrho}_N^{1-\epsilon}\right) + \frac{1-\lambda}{\eta - 1}\left(\bar{v}_S \bar{\varrho}_S^{1-\epsilon} - \bar{v}_N \bar{\varrho}_N^{1-\epsilon}\right)\,.$$

The log linearization of the production function [2.5] is:

$$\mathsf{Q} = \alpha \mathsf{H}\,, \quad \text{where} \;\; \alpha \equiv \frac{\mathcal{F}^{-1}(\bar{Q})\mathcal{F}'(\mathcal{F}^{-1}(\bar{Q}))}{\mathcal{F}(\mathcal{F}^{-1}(\bar{Q}))}\,. \qquad \text{[A.5.8]}$$

The nominal marginal cost function [2.6] has the following log-linear form:

$$\mathsf{X} = \gamma \mathsf{Q} + \mathsf{W}\,, \quad \text{where} \;\; \gamma \equiv \frac{\bar{Q}\mathscr{C}''(\bar{Q};\bar{W})}{\mathscr{C}'(\bar{Q};\bar{W})} = \left(-\frac{\mathcal{F}^{-1}(\bar{Q})\mathcal{F}''(\mathcal{F}^{-1}(\bar{Q}))}{\mathcal{F}'(\mathcal{F}^{-1}(\bar{Q}))}\right)\left(\frac{\bar{Q}}{\mathcal{F}^{-1}(\bar{Q})\mathcal{F}'(\mathcal{F}^{-1}(\bar{Q}))}\right)\,. \qquad \text{[A.5.9]}$$

The final relationship to derive is that linking Y and Q. The log-deviation of the ratio Y/Q is denoted by $\delta = \mathsf{Y} - \mathsf{Q}$. To find its determinants, begin by substituting [A.5.3] into [A.5.5], and using $\mathsf{P}_S = \mathsf{P}_B = \mathsf{X}$:

$$\mathsf{Q} = \mathsf{Y} + \epsilon \mathsf{P} + \left(\frac{\bar{q}_S - \bar{q}_N}{\bar{Q}}\right)\mathsf{s} - \left(\frac{(1-\bar{s})\bar{\zeta}_N \bar{q}_N}{\bar{Q}}\right)\mathsf{P}_N$$
$$+ \left(\delta(\eta - \epsilon)(1-\lambda)\left(\bar{s}\bar{v}_S\bar{\varrho}_S^{-\epsilon} + (1-\bar{s})\bar{v}_N\bar{\varrho}_N^{-\epsilon}\right) - \frac{\bar{s}\bar{\zeta}_S\bar{q}_S}{\bar{Q}}\right)\mathsf{X}\,. \qquad \text{[A.5.10]}$$

Substituting $\mathsf{P}_S = \mathsf{P}_B = \mathsf{X}$ into the expression for P_B in [A.5.6] and rearranging terms yields:

$$\mathsf{s} = \frac{(1-\theta_B)}{\psi_B}(\mathsf{P}_N - \mathsf{X})\,. \qquad \text{[A.5.11]}$$

Using the above equation and making the same substitutions in the expression for P from [A.5.7]:

$$\mathsf{P}_N = \left(\frac{\psi_B}{(1-\theta_P)\psi_B - (1-\theta_B)\psi_P}\right)\mathsf{P} - \left(\frac{(1-\theta_B)\psi_P\theta_P\psi_B}{(1-\theta_P)\psi_B - (1-\theta_B)\psi_P}\right)\mathsf{X}\,. \qquad \text{[A.5.12]}$$

Substituting equations [A.5.11] and [A.5.12] into [A.5.10] yields the following formula for $\delta = \mathsf{Y} - \mathsf{Q}$:

$$\delta = \left(\epsilon + \frac{(\bar{q}_S - \bar{q}_N)(1-\theta_B) - \psi_B(1-\bar{s})\bar{\zeta}_N\bar{q}_N}{((1-\theta_P)\psi_B - (1-\theta_B)\psi_P)\,\bar{Q}}\right)(\mathsf{X} - \mathsf{P})\,, \qquad \text{[A.5.13]}$$

which has been simplified by noting that all the constituent equations are homogeneous of degree zero in nominal variables, so the resulting expression for δ must be expressible in terms of real marginal cost $\mathsf{X} - \mathsf{P}$. Writing this as $\delta = \delta_x \mathsf{x}$, where $\mathsf{x} = \mathsf{X} - \mathsf{P}$, the coefficient δ_x is given by:

$$\delta_x \equiv \epsilon - \delta\frac{\psi_B(1-\bar{s})\left(\lambda\epsilon + (1-\lambda)\eta\bar{v}_N\right)\bar{\varrho}_N^{-\epsilon} - (1-\theta_B)\left((\lambda + (1-\lambda)\bar{v}_S)\bar{\varrho}_S^{-\epsilon} - (\lambda + (1-\lambda)\bar{v}_N)\bar{\varrho}_N^{-\epsilon}\right)}{(1-\theta_P)\psi_B - (1-\theta_B)\psi_P}\,, \qquad \text{[A.5.14]}$$

where the expressions for the flexible-price equilibrium values of $\bar{\zeta}_N$, $\bar{q}_S$, $\bar{q}_N$ and $\bar{Q}$ are used.

A.5.2 Sales model with flexible wages from section 6

The log-linearized labour supply equation [6.1] is:

$$\mathsf{w} = \frac{\sigma_h^{-1}}{\alpha}\mathsf{Q} + \sigma_c^{-1}\mathsf{Y}\,, \quad \text{where} \;\; \sigma_c \equiv -\left(\frac{\bar{Y}u_{cc}(\bar{Y})}{u_c(\bar{Y})}\right)^{-1}\,, \quad \text{and} \;\; \sigma_h \equiv \left(\frac{\mathcal{F}^{-1}(\bar{Y}/\delta)\nu_{hh}(\mathcal{F}^{-1}(\bar{Y}/\delta))}{\nu_h(\mathcal{F}^{-1}(\bar{Y}/\delta))}\right)^{-1}\,, \qquad \text{[A.5.15]}$$

and where the log deviation of the real wage is $\mathsf{w} = \mathsf{W} - \mathsf{P}$.

A.5.3 Aggregation in the dynamic sales model from section 7.1

The equivalent of the expression for the aggregate price level P_t in [7.3] for $P_{B,t}$, the bargain hunters' price index, is obtained from the definition in [3.4]:

$$P_{B,t} = \left((1-\phi_p) \sum_{\ell=0}^{\infty} \phi_p^{\ell} \left\{ s_{\ell,t} p_{S,\ell,t}^{1-\eta} + (1-s_{\ell,t}) R_{N,t-\ell}^{1-\eta} \right\} \right)^{\frac{1}{1-\eta}} .$$

Using the demand function [3.8], the total quantity sold by a vintage-ℓ firm is:

$$Q_{\ell,t} \equiv s_{\ell,t} q_{S,\ell,t} + (1-s_{\ell,t}) q_{N,\ell,t} \ , \quad \text{where } q_{S,\ell,t} = \mathscr{D}(p_{S,\ell,t}; P_{B,t}, \mathcal{E}_t) \ , \quad \text{and } q_{N,\ell,t} = \mathscr{D}(R_{N,t-\ell}; P_{B,t}, \mathcal{E}_t) \ .$$

where $q_{S,\ell,t}$ and $q_{N,\ell,t}$ are the respective quantities sold at prices $p_{S,\ell,t}$ and $R_{N,t-\ell}$. The corresponding purchase multipliers are $v_{S,\ell,t} = v(p_{S,\ell,t}; P_{B,t})$ and $v_{N,\ell,t} = v(R_{N,t-\ell}; P_{B,t})$. Given the total quantity produced $Q_{\ell,t}$, the vintage-specific hours of labour hired $H_{\ell,t}$ and nominal marginal cost $X_{\ell,t}$ are:

$$H_{\ell,t} = \mathcal{F}^{-1}(Q_{\ell,t}) \ , \quad X_{\ell,t} \equiv \mathscr{C}'(Q_{\ell,t}; W_t) \ .$$

Proposition 4 shows that $\mathsf{X}_{\ell,t} = \mathsf{X}_t$, $\mathsf{Q}_{\ell,t} = \mathsf{Q}_t$ and $\mathsf{p}_{S,\ell,t} = \mathsf{P}_{S,t}$. It follows immediately that $\mathsf{H}_{\ell,t} = \mathsf{H}_t$, $\mathsf{q}_{S,\ell,t} = \mathsf{q}_{S,t}$ and $\mathsf{v}_{S,\ell,t} = \mathsf{v}_{S,t}$.

The log linearizations derived in section A.5.1 continue to hold in the fully dynamic version of the model if certain variables are reinterpreted as weighted averages over normal-price vintages. These weighted averages are:

$$\mathsf{s}_t \equiv (1-\phi_p) \sum_{\ell=0}^{\infty} \phi_p^{\ell} \mathsf{s}_{\ell,t} \ , \quad \mathsf{q}_{N,t} \equiv (1-\phi_p) \sum_{\ell=0}^{\infty} \phi_p^{\ell} \mathsf{q}_{N,\ell,t} \ , \quad \mathsf{v}_{N,t} \equiv (1-\phi_p) \sum_{\ell=0}^{\infty} \phi_p^{\ell} \mathsf{v}_{N,\ell,t} \ ,$$

and also:

$$\mathsf{P}_{N,t} \equiv (1-\phi_p) \sum_{\ell=0}^{\infty} \phi_p^{\ell} \mathsf{R}_{N,t-\ell} \ . \qquad \text{[A.5.16]}$$

A.5.4 DSGE model from section 7.3

The log linearization of the intertemporal IS equation in [7.14] is:

$$\mathsf{Y}_t = \mathbb{E}_t \mathsf{Y}_{t+1} + \vartheta_m \left(\mathsf{m}_t - \mathbb{E}_t \mathsf{m}_{t+1} \right) - \sigma_c \left(\mathsf{i}_t - \mathbb{E}_t \pi_{t+1} \right) \ , \qquad \text{[A.5.17]}$$

where $\mathsf{i}_t \equiv \log(1+i_t) - \log(1+\bar{i})$ is the log deviation of the gross nominal interest rate, $\pi_t \equiv \log \pi_t - \log \bar{\pi}$ is the log deviation of the gross inflation rate, and the elasticities σ_c and ϑ_m are given by:

$$\sigma_c \equiv -\left(\frac{\bar{Y} \upsilon_{cc}(\bar{Y}, \bar{m})}{\upsilon_c(\bar{Y}, \bar{m})} \right)^{-1} \ , \quad \vartheta_m \equiv -\frac{\bar{m} \upsilon_{mc}(\bar{Y}, \bar{m})}{\bar{Y} \upsilon_{cc}(\bar{Y}, \bar{m})} \ .$$

Money demand from [7.14] is log linearized as follows:

$$\mathsf{m}_t = \vartheta_y \mathsf{Y}_t - \vartheta_i \mathsf{i}_t \ , \qquad \text{[A.5.18]}$$

where the income elasticity ϑ_y and interest semi-elasticity ϑ_i are:

$$\vartheta_y \equiv \frac{\frac{\bar{Y} \upsilon_{mc}(\bar{Y},\bar{m})}{\upsilon_m(\bar{Y},\bar{m})} - \frac{\bar{Y} \upsilon_{cc}(\bar{Y},\bar{m})}{\upsilon_c(\bar{Y},\bar{m})}}{\frac{\bar{m} \upsilon_{mc}(\bar{Y},\bar{m})}{\upsilon_c(\bar{Y},\bar{m})} - \frac{\bar{m} \upsilon_{mm}(\bar{Y},\bar{m})}{\upsilon_m(\bar{Y},\bar{m})}} \ , \quad \vartheta_i \equiv \frac{\beta}{(1-\beta) \left(\frac{\bar{m} \upsilon_{mc}(\bar{Y},\bar{m})}{\upsilon_c(\bar{Y},\bar{m})} - \frac{\bar{m} \upsilon_{mm}(\bar{Y},\bar{m})}{\upsilon_m(\bar{Y},\bar{m})} \right)} \ .$$

Note that after specifying σ_c, ϑ_y and ϑ_i, the steady-state ratio of real money balances to income (the reciprocal of money velocity) is determined as follows:

$$\bar{m} = \left(\frac{(1-\beta)\vartheta_m}{\frac{\beta\sigma_c\vartheta_y}{(1-\beta)\vartheta_i} - 1} \right) \bar{Y} \,. \qquad \text{[A.5.19]}$$

The log-linearized version of equation [7.15] for the utility-maximizing reset wage is:

$$\mathsf{R}_{W,t} = \frac{(1-\beta\phi_w)}{(1+\varsigma\sigma_h^{-1})} \sum_{\ell=0}^{\infty} (\beta\phi_w)^{\ell} \mathbb{E}_t \left[\mathsf{P}_{t+\ell} + \varsigma\sigma_h^{-1}\mathsf{W}_{t+\ell} + \sigma_c^{-1}\left(\mathsf{Y}_{t+\ell} - \vartheta_m \mathsf{m}_{t+\ell}\right) + \sigma_h^{-1}\mathsf{H}_{t+\ell} \right] \,,$$

which has the following recursive form:

$$\mathsf{R}_{W,t} = \beta\phi_w \mathbb{E}_t \mathsf{R}_{W,t+1} + \frac{(1-\beta\phi_w)}{(1+\varsigma\sigma_h^{-1})} \left(\mathsf{P}_t + \varsigma\sigma_h^{-1}\mathsf{W}_t + \sigma_c^{-1}\left(\mathsf{Y}_t - \vartheta_m \mathsf{m}_t\right) + \sigma_h^{-1}\mathsf{H}_t \right) \,. \qquad \text{[A.5.20]}$$

The log-linearized wage index [7.16] is:

$$\mathsf{W}_t = \sum_{\ell=0}^{\infty} (1-\phi_w)\phi_w^{\ell} \mathsf{R}_{W,t-\ell} \,,$$

which also has a recursive form:

$$\mathsf{W}_t = \phi_w \mathsf{W}_{t-1} + (1-\phi_w)\mathsf{R}_{W,t} \,. \qquad \text{[A.5.21]}$$

Combining the reset wage equation [A.5.20] with the wage index equation [A.5.21] yields an expression for wage inflation $\pi_{W,t} \equiv \mathsf{W}_t - \mathsf{W}_{t-1}$:

$$\pi_{W,t} = \beta\mathbb{E}_t \pi_{W,t+1} + \frac{(1-\phi_w)(1-\beta\phi_w)}{\phi_w} \frac{1}{1+\varsigma\sigma_h^{-1}} \left(\frac{\sigma_h^{-1}}{\alpha}\mathsf{Q}_t + \sigma_c^{-1}\left(\mathsf{Y}_t - \vartheta_m \mathsf{m}_t\right) - \mathsf{w}_t \right) \,, \qquad \text{[A.5.22]}$$

where the link between hours H_t and quantity Q_t in [A.5.8] is used.

A.6 Proof of Theorem 2

(i) Suppose that all firms share the same fixed normal price p_N consistent with the flexible-price equilibrium of section 4. The first-order condition for the optimal choice of the sales fraction s, which is the first equation in [5.1], is log-linearized as follows:

$$(\bar{q}_S - \bar{q}_N)\mathsf{X} = \bar{\mu}_S \bar{q}_S \mathsf{p}_S + (\bar{\mu}_S - 1)\bar{q}_S(\mathsf{q}_S - \mathsf{q}_N) \,,$$

where the fact that $(\bar{\mu}_S - 1)\bar{q}_S = (\bar{\mu}_N - 1)\bar{q}_N$ is used to simplify the expression. By using the log-linearized demand functions [A.5.3] and setting $\mathsf{p}_N = 0$:

$$(\bar{q}_S - \bar{q}_N)\mathsf{X} = \left(\bar{\mu}_S - (\bar{\mu}_S - 1)\left(\frac{\lambda\epsilon + (1-\lambda)\eta\bar{\upsilon}_S}{\lambda + (1-\lambda)\bar{\upsilon}_S} \right) \right) \bar{q}_S \mathsf{p}_S$$
$$+ (\eta - \epsilon)\left(\frac{(1-\lambda)\bar{\upsilon}_S}{\lambda + (1-\lambda)\bar{\upsilon}_S} - \frac{(1-\lambda)\bar{\upsilon}_N}{\lambda + (1-\lambda)\bar{\upsilon}_N} \right)(\bar{\mu}_S - 1)\bar{q}_S \mathsf{P}_B \,. \qquad \text{[A.6.1]}$$

Given the expression for $\bar{\mu}_S$ in [4.6], the coefficient of p_S in the above is zero. Since $\bar{q}_S > \bar{q}_N$, this equation implies X is independent of p_S. By using $(\bar{\mu}_S - 1)\bar{q}_S = (\bar{\mu}_N - 1)\bar{q}_N$ again, [A.6.1] implies:

$$(\bar{q}_S - \bar{q}_N)\mathsf{X} = (\bar{q}_S - \bar{q}_N)\mathsf{P}_B - \left(1 - (\eta - \epsilon)\left(\frac{(1-\lambda)\bar{\upsilon}_S}{\lambda + (1-\lambda)\bar{\upsilon}_S} \right)(\bar{\mu}_S - 1) \right) \bar{q}_S \mathsf{P}_B$$
$$+ \left(1 - (\eta - \epsilon)\left(\frac{(1-\lambda)\bar{\upsilon}_N}{\lambda + (1-\lambda)\bar{\upsilon}_N} \right)(\bar{\mu}_N - 1) \right) \bar{q}_N \mathsf{P}_B \,. \qquad \text{[A.6.2]}$$

After substituting the expressions for $\bar{\mu}_S$ and $\bar{\mu}_N$ from [4.6], the above equation reduces to:

$$(\bar{q}_S - \bar{q}_N)\mathsf{X} = (\bar{q}_S - \bar{q}_N)\mathsf{P}_B + (\epsilon - 1)\left((\bar{\mu}_S - 1)\bar{q}_S - (\bar{\mu}_N - 1)\bar{q}_N\right)\mathsf{P}_B \ ,$$

and noting that the coefficient on the final term is zero, it follows that $\mathsf{X} = \mathsf{P}_B$ for all p_S.

(ii) The optimal p_S is characterized by the second equation in [5.1]. In log-linear terms it is:

$$\mathsf{p}_S = \mu_S + \mathsf{X} \ .$$

By substituting the expression for the log-linearized optimal sale markup from [A.5.4] and the sale purchase multiplier from [A.5.2], and then rearranging terms yields:

$$(1 - (\eta - \epsilon)\mathfrak{c}_S)(\mathsf{p}_S - \mathsf{X}) = 0 \ , \tag{A.6.3}$$

so $\mathsf{p}_S = \mathsf{X}$ if the coefficient in the above is different from zero. The expressions for $\mathfrak{c}_S$ from [A.5.4] and $\bar{\mu}_S$ from [4.6] imply:

$$\frac{(1 - (\eta - \epsilon)\mathfrak{c}_S)}{\bar{\mu}_S} = \frac{(\lambda(\epsilon - 1) + (1 - \lambda)(\eta - 1)\bar{v}_S)(\lambda\epsilon + (1 - \lambda)\eta\bar{v}_S) - (\eta - \epsilon)^2\lambda(1 - \lambda)\bar{v}_S}{(\lambda\epsilon + (1 - \lambda)\eta\bar{v}_S)^2} \ .$$

Using [A.1.8] and noting that $v_S = \rho_S^{\epsilon - \eta}$ it follows that $1 - (\eta - \epsilon)\mathfrak{c}_S = \mu_S \mathcal{D}'(\rho_S)\mathcal{R}''(\mathcal{D}(\rho_S))$, where the functions $\mathcal{D}(\rho)$ and $\mathcal{R}(\mathfrak{q})$ are defined in [A.1.1] and [A.1.4]. The coefficient in [A.6.3] is strictly positive because $\mathcal{D}'(\rho_S) < 0$ and Lemma 2 shows that $\mathcal{R}''(\mathcal{D}(\rho_S)) < 0$, and therefore $\mathsf{p}_S = \mathsf{X}$. This completes the proof.

A.7 Solving the log-linearized model

A.7.1 Sales model from section 5

The model is log-linearized around the flexible-price and flexible-wage equilibrium characterized in section 4. The system of log-linearized equations is:

$$\mathsf{P} = \theta_P \mathsf{p}_S - \psi_P \mathsf{s} \ , \tag{A.7.1a}$$

$$\mathsf{P}_B = \theta_B \mathsf{p}_S - \psi_B \mathsf{s} \ , \tag{A.7.1b}$$

$$\mathsf{p}_S = \mathsf{X} \ , \tag{A.7.1c}$$

$$\mathsf{P}_B = \mathsf{X} \ , \tag{A.7.1d}$$

$$\mathsf{Y} = \mathsf{Q} + \delta_x(\mathsf{X} - \mathsf{P}) \ , \tag{A.7.1e}$$

$$\mathsf{X} = \gamma \mathsf{Q} \ , \tag{A.7.1f}$$

$$\mathsf{Y} = \mathsf{M} - \mathsf{P} \ . \tag{A.7.1g}$$

Equations [A.7.1a] and [A.7.1b] are [A.5.7] and [A.5.6] with $\mathsf{P}_N = 0$. Equations [A.7.1c] and [A.7.1d] are the results of Theorem 2. Equation [A.7.1e] is taken from [A.5.13] and [A.5.14] with $\delta = \mathsf{Y} - \mathsf{Q}$. Equation [A.7.1f] follows from [A.5.9] with $\mathsf{W} = 0$. Finally, equation [A.7.1g] is the log linearization of [2.4]. The money supply M is exogenous.

A.7.2 Sales model with flexible wages from section 6

The system of equations is the same as [A.7.1] except that [A.7.1f] is dropped and replaced by [A.5.9], and an additional equation for the wage W is taken from [A.5.15]:

$$\mathsf{W} = \mathsf{P} + \frac{\sigma_h^{-1}}{\alpha}\mathsf{Q} + \sigma_c^{-1}\mathsf{Y} \ , \tag{A.7.2a}$$

$$\mathsf{X} = \mathsf{W} + \gamma \mathsf{Q} \ . \tag{A.7.2b}$$

A.8 Proof of Proposition 4

(i) Consider a firm with arbitrary deviations p_S and p_N of its sale and normal prices from the flexible-price equilibrium. The log-linearized first-order condition for the sales fraction (the first equation in [5.1]) is:

$$(\bar{q}_S - \bar{q}_N)\mathsf{X} = \bar{\mu}_S\bar{q}_S\mathsf{p}_S - \bar{\mu}_N\bar{q}_N\mathsf{p}_N + (\bar{\mu}_S - 1)\bar{q}_S(\mathsf{q}_S - \mathsf{q}_N) \, , \quad \text{[A.8.1]}$$

where the fact that $(\bar{\mu}_S - 1)\bar{q}_S = (\bar{\mu}_N - 1)\bar{q}_N$ is used to simplify the expression. By using [A.5.3]:

$$\begin{aligned}(\bar{q}_S - \bar{q}_N)\mathsf{X} = &\left(\bar{\mu}_S - (\bar{\mu}_S - 1)\left(\frac{\lambda\epsilon + (1-\lambda)\eta\bar{v}_S}{\lambda + (1-\lambda)\bar{v}_S}\right)\right)\bar{q}_S\mathsf{p}_S \\ &- \left(\bar{\mu}_N - (\bar{\mu}_N - 1)\left(\frac{\lambda\epsilon + (1-\lambda)\eta\bar{v}_N}{\lambda + (1-\lambda)\bar{v}_N}\right)\right)\bar{q}_N\mathsf{p}_N \\ &+ (\eta - \epsilon)\left(\frac{(1-\lambda)\bar{v}_S}{\lambda + (1-\lambda)\bar{v}_S} - \frac{(1-\lambda)\bar{v}_N}{\lambda + (1-\lambda)\bar{v}_N}\right)(\bar{\mu}_S - 1)\bar{q}_S\mathsf{P}_B \, .\end{aligned}$$

Given the expressions for $\bar{\mu}_S$ and $\bar{\mu}_N$ in [4.5], the coefficients of both p_S and p_N in the above are zero. Since $\bar{q}_S > \bar{q}_N$, this equation implies X is independent of p_S and p_N. Using $(\bar{\mu}_S - 1)\bar{q}_S = (\bar{\mu}_N - 1)\bar{q}_N$ again in equation [A.8.1] yields the same expression involving X and P_B as found in [A.6.2]. Following the same steps subsequent to [A.6.2] establishes that $\mathsf{X} = \mathsf{P}_B$.

(ii) Since all firms face the same wage W, and as part (i) shows that all have the same nominal marginal cost X, the log linearization of nominal marginal cost in [A.5.9] shows that all must produce the same total quantity Q.

(iii) The profit-maximizing sale and normal prices from [4.5] are $p_S^* = \mu_S X$ and $p_N^* = \mu_N X$. In log-linear terms these equations are:

$$\mathsf{p}_S^* = \upmu_S + \mathsf{X} \, , \quad \mathsf{p}_N^* = \upmu_N + \mathsf{X} \, .$$

By following the same steps as in the proof of part (ii) of Theorem 2, it follows that $\mathsf{p}_S^* = \mathsf{X}$ and $\mathsf{p}_N^* = \mathsf{X}$.

(iv) Let p_S and p_N be given sale and normal prices for a particular firm, and let s be the optimal sales fraction implied by the first equation in [5.1]. The resulting profits $\mathscr{P}$ from [3.10] are:

$$\mathscr{P} = sp_Sq_S + (1-s)p_Nq_N - \mathscr{C}(Q;W) \, ,$$

where $\mathscr{C}(Q;W)$ is the total cost function [2.6]. Taking a second-order Taylor expansion of profits around the flexible-price equilibrium yields:

$$\begin{aligned}\mathscr{P} = &\bar{s}\bar{p}_S\bar{q}_S(\mathsf{p}_S + \mathsf{q}_S) + \bar{p}_S\bar{q}_S\mathsf{s} + (1-\bar{s})\bar{p}_N\bar{q}_N(\mathsf{p}_N + \mathsf{q}_N) - \bar{p}_N\bar{q}_N\mathsf{s} + \bar{p}_S\bar{q}_S\mathsf{s}(\mathsf{p}_S + \mathsf{q}_S) \\ &- \bar{p}_N\bar{q}_N\mathsf{s}(\mathsf{p}_N + \mathsf{q}_N) + \frac{1}{2}\bar{p}_S\bar{q}_S\bar{s}(\mathsf{p}_S + \mathsf{q}_S)^2 + \frac{1}{2}\bar{p}_N\bar{q}_N(1-\bar{s})(\mathsf{p}_N + \mathsf{q}_N)^2 \\ &- \bar{Q}\bar{X}\mathsf{Q} - \frac{1}{2}\bar{Q}\bar{X}(1+\gamma)\mathsf{Q}^2 - \bar{Q}\bar{X}\mathsf{Q}\mathsf{W} + \text{t.i.p.} + \mathscr{O}(3) \, , \quad \text{[A.8.2]}\end{aligned}$$

where "t.i.p." denotes terms independent of an individual firm's normal and sale prices, and $\mathscr{O}(3)$ represents third- and higher-order terms in the log deviations of variables from their steady-state values. The first-order approximation [A.5.3] of the demand functions [4.10] is extended to include second-order terms as follows:

$$\mathsf{q}_S = -\bar{\zeta}_S\mathsf{p}_S + \mathsf{d}_S + \frac{(\eta-\epsilon)^2\lambda(1-\lambda)\bar{v}_S}{2(\lambda + (1-\lambda)\bar{v}_S)^2}(\mathsf{p}_S - \mathsf{P}_B)^2 + \mathscr{O}(3) \, , \quad \text{[A.8.3a]}$$

$$\mathsf{q}_N = -\bar{\zeta}_N\mathsf{p}_N + \mathsf{d}_N + \frac{(\eta-\epsilon)^2\lambda(1-\lambda)\bar{v}_N}{2(\lambda + (1-\lambda)\bar{v}_N)^2}(\mathsf{p}_N - \mathsf{P}_B)^2 + \mathscr{O}(3) \, , \quad \text{[A.8.3b]}$$

where expressions for the price elasticities $\bar{\zeta}_S$ and $\bar{\zeta}_N$ are obtained from [4.1], and the following terms are defined:

$$\mathsf{d}_S = \frac{(\eta-\epsilon)(1-\lambda)\bar{v}_S}{\lambda+(1-\lambda)\bar{v}_S}\mathsf{P}_B + \epsilon\mathsf{P} + \mathsf{Y}\,, \quad \mathsf{d}_N = \frac{(\eta-\epsilon)(1-\lambda)\bar{v}_N}{\lambda+(1-\lambda)\bar{v}_N}\mathsf{P}_B + \epsilon\mathsf{P} + \mathsf{Y}\,.$$

Then by using the second-order Taylor expansion of total quantity sold $Q = sq_S + (1-s)q_N$:

$$\bar{Q}\left(\mathsf{Q} + \frac{\mathsf{Q}^2}{2}\right) = \bar{s}\bar{q}_S\mathsf{q}_S + (1-\bar{s})\bar{q}_N\mathsf{q}_N + (\bar{q}_S - \bar{q}_N)\mathsf{s} + \frac{\bar{s}\bar{q}_S}{2}\mathsf{q}_S^2 + \frac{(1-\bar{s})\bar{q}_N}{2}\mathsf{q}_N^2 + \bar{q}_S\mathsf{s}\mathsf{q}_S - \bar{q}_N\mathsf{s}\mathsf{q}_N + \mathscr{O}(3)\,,$$

the level of profits $\mathscr{P}$ from [A.8.2] is broken down into four components:

$$\mathscr{P} = \mathfrak{P}_1 + \mathfrak{P}_2 + \mathfrak{P}_3 + \mathfrak{P}_4 + \text{t.i.p.} + \mathscr{O}(3)\,,$$

with:

$$\mathfrak{P}_1 \equiv \bar{s}\bar{p}_S\bar{q}_S\left(\mathsf{p}_S + \mathsf{q}_S\right) + (1-\bar{s})\bar{p}_N\bar{q}_N\left(\mathsf{p}_N + \mathsf{q}_N\right) + (\bar{p}_S\bar{q}_S - \bar{p}_N\bar{q}_N)\mathsf{s}$$
$$- \bar{X}(\bar{s}\bar{q}_S\mathsf{q}_S + (1-\bar{s})\bar{q}_N\mathsf{q}_N + (\bar{q}_S - \bar{q}_N)\mathsf{s})\,, \qquad \text{[A.8.4a]}$$

$$\mathfrak{P}_2 \equiv \frac{1}{2}\bar{p}_S\bar{q}_S\bar{s}\left(\mathsf{p}_S + \mathsf{q}_S\right)^2 + \frac{1}{2}\bar{p}_N\bar{q}_N(1-\bar{s})\left(\mathsf{p}_N + \mathsf{q}_N\right)^2 - \bar{X}\left(\frac{\bar{s}\bar{q}_S}{2}\mathsf{q}_S^2 + \frac{(1-\bar{s})\bar{q}_N}{2}\mathsf{q}_N^2\right)\,, \qquad \text{[A.8.4b]}$$

$$\mathfrak{P}_3 \equiv \bar{p}_S\bar{q}_S\mathsf{s}\left(\mathsf{p}_S + \mathsf{q}_S\right) - \bar{p}_N\bar{q}_N\mathsf{s}\left(\mathsf{p}_N + \mathsf{q}_N\right) - \bar{X}\left(\bar{q}_S\mathsf{s}\mathsf{q}_S - \bar{q}_N\mathsf{s}\mathsf{q}_N\right)\,, \qquad \text{[A.8.4c]}$$

$$\mathfrak{P}_4 \equiv -\frac{\gamma\bar{X}\bar{Q}}{2}\mathsf{Q}^2 - \bar{X}\bar{Q}\mathsf{W}\mathsf{Q}\,. \qquad \text{[A.8.4d]}$$

By using the equations $\bar{p}_S = \bar{\mu}_S\bar{X}$ and $\bar{p}_N = \bar{\mu}_N\bar{X}$ and simplifying, an equivalent expression for $\mathfrak{P}_1$ in [A.8.4a] is:

$$\mathfrak{P}_1 = \bar{s}\bar{q}_S\bar{X}(\bar{\mu}_S\mathsf{p}_S + (\bar{\mu}_S - 1)\mathsf{q}_S) + (1-\bar{s})\bar{q}_N\bar{X}(\bar{\mu}_N\mathsf{p}_N + (\bar{\mu}_N - 1)\mathsf{q}_N) + \bar{X}(\bar{q}_S(\bar{\mu}_S - 1) - \bar{q}_N(\bar{\mu}_N - 1))\mathsf{s}\,.$$

The first-order terms in the above are shown to have zero coefficients by substituting the second-order expansions of demand from [A.8.3] and using the expressions for $\bar{\mu}_S$ and $\bar{\mu}_N$ from [4.6], and $\bar{q}_S(\bar{\mu}_S - 1) = \bar{q}_N(\bar{\mu}_N - 1)$. Thus:

$$\mathfrak{P}_1 = \frac{\bar{s}\bar{q}_S\bar{X}(\bar{\mu}_S - 1)(\eta-\epsilon)^2\lambda(1-\lambda)\bar{v}_S}{2(\lambda+(1-\lambda)\bar{v}_S)^2}(\mathsf{p}_S - \mathsf{P}_B)^2$$
$$+ \frac{(1-\bar{s})\bar{q}_N\bar{X}(\bar{\mu}_N - 1)(\eta-\epsilon)^2\lambda(1-\lambda)\bar{v}_N}{2(\lambda+(1-\lambda)\bar{v}_N)^2}(\mathsf{p}_N - \mathsf{P}_B)^2 + \text{t.i.p.} + \mathscr{O}(3)\,. \qquad \text{[A.8.5]}$$

The expression for $\mathfrak{P}_2$ is simplified by noting that [A.8.3] implies $\mathsf{q}_S = -\bar{\zeta}_S\mathsf{p}_S + \mathsf{d}_S + \mathscr{O}(2)$, and by substituting this into the following equation:

$$\bar{p}_S\left(\mathsf{p}_S + \mathsf{q}_S\right)^2 - \bar{X}\mathsf{q}_S^2 = \bar{X}\left(\frac{\bar{\mu}_S}{(\bar{\mu}_S - 1)^2}\mathsf{p}_S^2 - 2\frac{\bar{\mu}_S}{\bar{\mu}_S - 1}\mathsf{p}_S\mathsf{d}_S + \bar{\mu}_S\mathsf{d}_S^2\right)$$
$$- \bar{X}\left(\frac{\bar{\mu}_S^2}{(\bar{\mu}_S - 1)^2}\mathsf{p}_S^2 - 2\frac{\bar{\mu}_S}{\bar{\mu}_S - 1}\mathsf{p}_S\mathsf{d}_S + \mathsf{d}_S^2\right) + \mathscr{O}(3)$$
$$= -\bar{X}\frac{\bar{\mu}_S}{\bar{\mu}_S - 1}\mathsf{p}_S^2 + \text{t.i.p.} + \mathscr{O}(3)\,,$$

where $\bar{\mu}_S - 1 = 1/(\bar{\zeta}_S - 1)$ is used. An analogous expression holds for p_N and q_N. Substituting these results into [A.8.4b] yields:

$$\mathfrak{P}_2 = -\frac{\bar{X}}{2}\left(\bar{s}\bar{q}_S\bar{\zeta}_S\mathsf{p}_S^2 + (1-\bar{s})\bar{q}_N\bar{\zeta}_N\mathsf{p}_N^2\right) + \text{t.i.p.} + \mathscr{O}(3)\,. \qquad \text{[A.8.6]}$$

By taking out the term s as a common factor from $\mathfrak{P}_3$ in [A.8.4c] and noting that $\bar{p}_S = \bar{\mu}_S\bar{X}$ and $\bar{p}_N = \bar{\mu}_N\bar{X}$:

$$\mathfrak{P}_3 = \bar{X}\left(\bar{q}_S\left(\bar{\mu}_S\mathsf{p}_S + (\bar{\mu}_S - 1)\mathsf{q}_S\right) - \bar{q}_N\left(\bar{\mu}_N\mathsf{p}_N + (\bar{\mu}_N - 1)\mathsf{q}_N\right)\right)\mathsf{s}\,. \qquad \text{[A.8.7]}$$

Equation [A.8.3] implies $\mathsf{q}_S = -\bar{\zeta}_S\mathsf{p}_S + \mathsf{d}_S + \mathscr{O}(2)$ and $\mathsf{q}_N = -\bar{\zeta}_N\mathsf{p}_N + \mathsf{d}_N + \mathscr{O}(2)$, and by substituting these expressions into [A.8.7] and noting that $\bar{\mu}_S - 1 = 1/(\bar{\zeta}_S - 1)$ and $(\bar{\mu}_S - 1)\bar{q}_S = (\bar{\mu}_N - 1)\bar{q}_N$:

$$\mathfrak{P}_3 = \bar{X}\bar{q}_S(\bar{\mu}_S - 1)\mathsf{s}(\mathsf{d}_S - \mathsf{d}_N) + \mathscr{O}(3) \; . \tag{A.8.8}$$

The expression for $\mathfrak{P}_3$ is simplified by noting that:

$$\begin{aligned}
\bar{q}_S(\bar{\mu}_S - 1)(\mathsf{d}_S - \mathsf{d}_N) &= \bar{q}_S(\bar{\mu}_S - 1)\left(\frac{(\eta-\epsilon)(1-\lambda)\bar{v}_S}{\lambda + (1-\lambda)\bar{v}_S} - \frac{(\eta-\epsilon)(1-\lambda)\bar{v}_N}{\lambda + (1-\lambda)\bar{v}_N}\right)\mathsf{P}_B \\
&= \left(\bar{q}_S\frac{(\eta-\epsilon)(1-\lambda)\bar{v}_S(\bar{\mu}_S - 1)}{\lambda + (1-\lambda)\bar{v}_S} - \bar{q}_N\frac{(\eta-\epsilon)(1-\lambda)\bar{v}_N(\bar{\mu}_N - 1)}{\lambda + (1-\lambda)\bar{v}_N}\right)\mathsf{P}_B \\
&= \left(\bar{q}_S\left(1 - (\epsilon - 1)(\bar{\mu}_S - 1)\right) - \bar{q}_N\left(1 - (\epsilon-1)(\bar{\mu}_N - 1)\right)\right)\mathsf{P}_B = (\bar{q}_S - \bar{q}_N)\mathsf{P}_B \; ,
\end{aligned}$$

using $(\bar{\mu}_S - 1)\bar{q}_S = (\bar{\mu}_N - 1)\bar{q}_N$ repeatedly and the expressions for $\bar{\mu}_S$ and $\bar{\mu}_N$ from [4.6]. The following expression is obtained by substituting the above result into [A.8.8]:

$$\mathfrak{P}_3 = \bar{X}\mathsf{P}_B(\bar{q}_S - \bar{q}_N)\mathsf{s} + \mathscr{O}(3) = -\bar{X}(\bar{s}\bar{q}_S\mathsf{q}_S + (1-\bar{s})\bar{q}_N\mathsf{q}_N - \bar{Q}\mathsf{Q})\mathsf{P}_B + \mathscr{O}(3) \; ,$$

where the second equality makes use of the first-order expansion of total quantity Q from [A.5.5].

Appealing to Proposition 4, the log deviations of nominal marginal cost X and total quantity sold Q are independent of an individual firm's p_S and p_N. Therefore all the terms affecting $\mathfrak{P}_4$ in [A.8.4d] are independent of an individual firm's sale and normal prices. Furthermore, in the expression for $\mathfrak{P}_3$, the product of Q and P_B is also independent of an individual firm's two prices. Thus:

$$\mathfrak{P}_3 = -\bar{X}(\bar{s}\bar{q}_S\mathsf{q}_S + (1-\bar{s})\bar{q}_N\mathsf{q}_N)\mathsf{P}_B + \text{t.i.p.} + \mathscr{O}(3) \; , \quad \mathfrak{P}_4 = \text{t.i.p.} \tag{A.8.9}$$

The following expression for the sum of $\mathfrak{P}_2$ and $\mathfrak{P}_3$ from [A.8.6] and [A.8.9] results after substituting the first-order expansions of the levels of demand q_S and q_N into $\mathfrak{P}_3$:

$$\mathfrak{P}_2 + \mathfrak{P}_3 = -\frac{\bar{X}}{2}\left(\bar{s}\bar{q}_S\bar{\zeta}_S\mathsf{p}_S^2 + (1-\bar{s})\bar{q}_N\bar{\zeta}_N\mathsf{p}_N^2\right) + \bar{X}(\bar{s}\bar{q}_S\bar{\zeta}_S\mathsf{p}_S + (1-\bar{s})\bar{q}_N\bar{\zeta}_N\mathsf{p}_N)\mathsf{P}_B + \text{t.i.p.} + \mathscr{O}(3) \; .$$

By completing the square of the above and noting that the remainder is independent of prices:

$$\mathfrak{P}_2 + \mathfrak{P}_3 = -\frac{1}{2}\bar{s}\bar{q}_S\bar{\zeta}_S\bar{X}\left(\mathsf{p}_S - \mathsf{P}_B\right)^2 - \frac{1}{2}(1-\bar{s})\bar{q}_N\bar{\zeta}_N\bar{X}\left(\mathsf{p}_N - \mathsf{P}_B\right)^2 + \text{t.i.p.} + \mathscr{O}(3) \; .$$

Proposition 4 shows that $\mathsf{P}_B = \mathsf{X} + \mathscr{O}(2)$, and hence by combining the above equation with the expression for $\mathfrak{P}_1$ from [A.8.5]:

$$\begin{aligned}
\mathscr{P} = -\frac{1}{2}\bar{s}\bar{q}_S\bar{X}&\left(\bar{\zeta}_S - \frac{(\eta-\epsilon)^2\lambda(1-\lambda)\bar{v}_S(\bar{\mu}_S - 1)}{(\lambda + (1-\lambda)\bar{v}_S)^2}\right)(\mathsf{p}_S - \mathsf{X})^2 \\
&- \frac{1}{2}(1-\bar{s})\bar{q}_N\bar{X}\left(\bar{\zeta}_N - \frac{(\eta-\epsilon)^2\lambda(1-\lambda)\bar{v}_N(\bar{\mu}_N - 1)}{(\lambda + (1-\lambda)\bar{v}_N)^2}\right)(\mathsf{p}_N - \mathsf{X})^2 + \text{t.i.p.} + \mathscr{O}(3) \; ,
\end{aligned}$$

which completes the proof.

A.9 Proof of Theorem 3

The first step is to log-linearize equation [7.2] for the optimal reset price $R_{N,t}$ at time t. Since $\bar{R}_N = \bar{p}_N = \bar{\mu}_N\bar{X}$, it follows that this equation simplifies to:

$$\sum_{\ell=0}^{\infty}(\beta\phi_p)^{\ell}\mathbb{E}_t\left[\mathsf{R}_{N,t} - \mathsf{\mu}_{N,\ell,t+\ell} - \mathsf{X}_{\ell,t+\ell}\right] = 0 \; , \tag{A.9.1}$$

where $\mathsf{\mu}_{N,\ell,t}$ is the log-deviation of the optimal markup $\mu_{N,\ell,t} \equiv \mu(R_{N,t-\ell}; P_{B,t})$. The optimal markup function is log-linearized in [A.5.4] and is given in terms of the corresponding purchase multiplier, itself log-linearized in [A.5.2]. Putting together those results, it follows that $\mathsf{\mu}_{N,\ell,t+\ell} = (\eta - \epsilon)\mathfrak{c}_N\left(\mathsf{R}_{N,t} - \mathsf{P}_{B,t+\ell}\right)$.

Proposition 4 shows that marginal cost is equalized across all price vintages and thus $\mathsf{X}_{\ell,t} = \mathsf{X}_t$. Furthermore, the proposition establishes that $\mathsf{X}_t = \mathsf{P}_{B,t}$. Substituting all these findings into [A.9.1] yields:

$$(1-(\eta-\epsilon)\mathfrak{c}_N)\sum_{\ell=0}^{\infty}(\beta\phi_p)^{\ell}\mathbb{E}_t\left[R_{N,t}-\mathsf{X}_{t+\ell}\right]=0\ .$$

The proof of part (iii) of Proposition 4 demonstrates that $1-(\eta-\epsilon)\mathfrak{c}_N > 0$, and hence:

$$\mathsf{R}_{N,t}=(1-\beta\phi_p)\sum_{\ell=0}^{\infty}(\beta\phi_p)^{\ell}\mathbb{E}_t\mathsf{X}_{t+\ell}\ ,$$

which is expressed in an equivalent recursive form:

$$\mathsf{R}_{N,t}=\beta\phi_p\mathbb{E}_t\mathsf{R}_{t+1}+(1-\beta\phi_p)\mathsf{X}_t\ . \tag{A.9.2}$$

Using the log-linearizations [A.5.7] and [A.5.6] and the definition of the price index $\mathsf{P}_{N,t}$ in [A.5.16], the expressions for P_t and $\mathsf{P}_{B,t}$ are:

$$\mathsf{P}_t=\theta_P\mathsf{P}_{S,t}+(1-\theta_P)\mathsf{P}_{N,t}-\psi_P\mathsf{s}_t\ ,\quad \mathsf{P}_{B,t}=\theta_B\mathsf{P}_{S,t}+(1-\theta_B)\mathsf{P}_{N,t}-\psi_B\mathsf{s}_t\ , \tag{A.9.3}$$

where the fact that $\mathsf{p}_{S,\ell,t}=\mathsf{P}_{S,t}$ is used in accordance with Proposition 4. The recursive form of the expression for $\mathsf{P}_{N,t}$ in [A.5.16] is:

$$\mathsf{P}_{N,t}=\phi_p\mathsf{P}_{N,t-1}+(1-\phi_p)\mathsf{R}_{N,t}\ . \tag{A.9.4}$$

Proposition 4 establishes that $\mathsf{P}_{S,t}=\mathsf{X}_t$ and hence by substituting this into [A.9.3]:

$$\psi_P\mathsf{s}_t=\theta_P(\mathsf{X}_t-\mathsf{P}_t)+(1-\theta_P)(\mathsf{P}_{N,t}-\mathsf{P}_t)\ . \tag{A.9.5}$$

Likewise, by using $\mathsf{P}_{B,t}=\mathsf{X}_t$ and performing similar substitutions in the second part of [A.9.3]:

$$\psi_B\mathsf{s}_t=(1-\theta_B)(\mathsf{P}_{N,t}-\mathsf{X}_t)\ . \tag{A.9.6}$$

Equation [A.9.5] is written as:

$$\psi_P\mathsf{s}_t=\theta_P(\mathsf{X}_t-\mathsf{P}_t)+(1-\theta_P)\left((\mathsf{P}_{N,t}-\mathsf{X}_t)-(\mathsf{X}_t-\mathsf{P}_t)\right)\ ,$$

and s_t is eliminated using [A.9.6]. After some rearrangement this leads to:

$$\mathsf{X}_t-\mathsf{P}_{N,t}=\frac{1}{1-\psi}\mathsf{x}_t\ , \tag{A.9.7}$$

where ψ is as defined in the statement of the theorem and $\mathsf{x}_t=\mathsf{X}_t-\mathsf{P}_t$ is real marginal cost.

Multiplying both sides of [A.9.2] by $(1-\phi_p)$ and substituting the recursive equation [A.9.4] for $\mathsf{P}_{N,t}$ yields:

$$\mathsf{P}_{N,t}-\phi_p\mathsf{P}_{N,t-1}=\beta\phi_p\mathbb{E}_t\left[\mathsf{P}_{N,t+1}-\phi_p\mathsf{P}_{N,t}\right]+(1-\phi_p)(1-\beta\phi_p)\mathsf{X}_t\ ,$$

which is expressed in terms of normal-price inflation $\pi_{N,t}\equiv\mathsf{P}_{N,t}-\mathsf{P}_{N,t-1}$:

$$\pi_{N,t}=\beta\mathbb{E}_t\pi_{N,t+1}+\kappa(\mathsf{X}_t-\mathsf{P}_{N,t})\ , \tag{A.9.8}$$

and where κ is as defined in the statement of the theorem.

Taking the first difference of [A.9.6] yields:

$$\Delta\mathsf{s}_t=-\frac{(1-\theta_B)}{\psi_B}\left(\Delta\mathsf{X}_t-\pi_{N,t}\right)\ . \tag{A.9.9}$$

Now use the first part of [A.9.3] and make the substitution $\mathsf{P}_{S,t}=\mathsf{X}_t$ as before, and then take first differences and rearrange:

$$\pi_t = \pi_{N,t} + \theta_P \left(\Delta \mathsf{X}_t - \pi_{N,t}\right) - \psi_P \Delta \mathsf{s}_t \ .$$

By eliminating $\Delta \mathsf{s}_t$ from this equation using [A.9.9]:

$$\pi_t = \pi_{N,t} + \psi \left(\Delta \mathsf{X}_t - \pi_{N,t}\right) \ .$$

Substituting the first difference of equation [A.9.7] into the above yields:

$$\pi_{N,t} = \pi_t - \frac{\psi}{1-\psi} \Delta \mathsf{x}_t \ .$$

Combining this equation with [A.9.7] and [A.9.8] implies:

$$\left(\pi_t - \frac{\psi}{1-\psi}\Delta \mathsf{x}_t\right) = \beta \mathbb{E}_t \left[\pi_{t+1} - \frac{\psi}{1-\psi}\Delta \mathsf{x}_{t+1}\right] + \frac{\kappa}{1-\psi}\mathsf{x}_t \ ,$$

which is rearranged to yield the result [7.5]. Recursive forward substitution of equation [7.5] leads to:

$$\pi_t = \frac{1}{1-\psi}\sum_{\ell=0}^{\infty} \beta^{\ell} \mathbb{E}_t \left[\kappa \mathsf{x}_{t+\ell} + \psi \left(\Delta \mathsf{x}_{t+\ell} - \beta \Delta \mathsf{x}_{t+1+\ell}\right)\right] \ .$$

Notice that all $\Delta \mathsf{x}_{t+\ell}$ terms apart from $\Delta \mathsf{x}_t$ cancel out because each occurs twice with opposite signs. Hence equation [7.6] is obtained, which completes the proof.

A.10 Solving the log-linearized DSGE model

The fully dynamic model is log-linearized around the flexible-price and flexible-wage equilibrium characterized in section 4, with [4.11] replaced by:

$$\bar{x} = \frac{\varsigma}{\varsigma - 1} \frac{\nu_h \left(\mathcal{F}^{-1}(\bar{Y}/\delta)\right)}{\upsilon_c(\bar{Y}, \bar{m}) \mathcal{F}' \left(\mathcal{F}^{-1}(\bar{Y}/\delta)\right)} \ ,$$

and where the link between $\bar{m}$ and $\bar{Y}$ is given in [A.5.19]. The system of log-linearized equations is:

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \frac{1}{1-\psi}\left(\kappa \mathsf{x}_t + \psi \left(\Delta \mathsf{x}_t - \beta \mathbb{E}_t \Delta \mathsf{x}_{t+1}\right)\right) \ , \qquad \text{[A.10.1a]}$$

$$\pi_{W,t} = \beta \mathbb{E}_t \pi_{W,t+1} + \frac{(1-\phi_w)(1-\beta\phi_w)}{\phi_w} \frac{1}{1+\varsigma \sigma_h^{-1}} \left(\frac{\sigma_h^{-1}}{\alpha} \mathsf{Q}_t + \sigma_c^{-1}\left(\mathsf{Y}_t - \vartheta_m \mathsf{m}_t\right) - \mathsf{w}_t\right) \ , \qquad \text{[A.10.1b]}$$

$$\Delta \mathsf{w}_t = \pi_{W,t} - \pi_t \ , \qquad \text{[A.10.1c]}$$

$$\mathsf{Y}_t = \mathsf{Q}_t + \delta_x \mathsf{x}_t \ , \qquad \text{[A.10.1d]}$$

$$\mathsf{x}_t = \mathsf{w}_t + \gamma \mathsf{Q}_t \ , \qquad \text{[A.10.1e]}$$

$$\mathsf{Y}_t = \mathbb{E}_t \mathsf{Y}_{t+1} + \vartheta_m \left(\mathsf{m}_t - \mathbb{E}_t \mathsf{m}_{t+1}\right) - \sigma_c \left(\mathsf{i}_t - \mathbb{E}_t \pi_{t+1}\right) \ , \qquad \text{[A.10.1f]}$$

$$\mathsf{m}_t = \vartheta_y \mathsf{Y}_t - \vartheta_i \mathsf{i}_t \ . \qquad \text{[A.10.1g]}$$

Equation [A.10.1a] is the Phillips curve with sales derived in Theorem 3. Equation [A.10.1b] is the Phillips curve for wage inflation from [A.5.22], and equation [A.10.1c] follows from the definition of the real wage. Equations [A.10.1d] and [A.10.1e] are taken from [A.7.1e] and [A.7.2b], which continue to hold in the dynamic model. The IS equation [A.10.1f] and money demand [A.10.1g] come from [A.5.17] and [A.5.18].

There are two specifications of monetary policy considered: exogenous money growth [7.17a]:

$$\Delta \mathsf{M}_t = \varphi_m \Delta \mathsf{M}_{t-1} + \mathsf{e}_t \ , \qquad \text{[A.10.1h]}$$

and the Taylor rule with interest-rate smoothing [7.17b]:

$$\mathsf{i}_t = \varphi_i \mathsf{i}_{t-1} + (1-\varphi_i)\left(\varphi_\pi \pi_t + \varphi_y \mathsf{Y}_t\right) + \mathsf{e}_t \ . \qquad \text{[A.10.1i]}$$

The standard model with Dixit-Stiglitz preferences, a one-price equilibrium, and Calvo staggered price-adjustment times features the following New Keynesian Phillips curve:

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \frac{\kappa}{1+\varepsilon\gamma} \mathsf{x}_t \ ,$$

in place of [A.10.1a].[15] Equation [A.10.1d] is replaced by $\mathsf{Q}_t = \mathsf{Y}_t$.

A.11 Second-order approximation of profits with standard consumer preferences

Suppose a given firm charges price p and the aggregate price level is P and output is Y. Standard Dixit-Stiglitz preferences imply the following demand function with constant price elasticity ε:

$$q = \left(\frac{p}{P}\right)^{-\varepsilon} Y \ .$$

Assume the total cost function is $\mathscr{C}(q; W)$. Profits $\mathscr{P}$ are then given by:

$$\mathscr{P} = \frac{p^{1-\varepsilon}}{P^{-\varepsilon}} Y - \mathscr{C}\left(\left(\frac{p}{P}\right)^{-\varepsilon} Y; W\right) \ .$$

Taking a second-order approximation of total revenue yields:

$$\frac{p^{1-\varepsilon}}{P^{-\varepsilon}} Y = \bar{Y}\left(1 + (1-\varepsilon)\mathsf{p} - \varepsilon\mathsf{P} + \mathsf{Y} + \frac{1}{2}\left((1-\varepsilon)\mathsf{p} - \varepsilon\mathsf{P} + \mathsf{Y}\right)^2\right) + \mathscr{O}(3) \ ,$$

and of total cost:

$$\mathscr{C}(q; W) = \mathscr{C}(\bar{Y}; \bar{W}) + \left(\frac{\varepsilon - 1}{\varepsilon}\right) \bar{Y}\left(-\varepsilon(\mathsf{p} - \mathsf{P}) + \mathsf{Y} + \frac{1}{2}(1+\gamma)\left(-\varepsilon(\mathsf{p} - \mathsf{P}) + \mathsf{Y}\right)^2\right) + \mathscr{O}(3) \ ,$$

where $\gamma \equiv \bar{Y}\mathscr{C}''(\bar{Y}; \bar{W})/\mathscr{C}'(\bar{Y}; \bar{W})$, and $\mathscr{C}'(\bar{Y}; \bar{W}) = (\varepsilon - 1)/\varepsilon$ and $\mathsf{q} = -\varepsilon(\mathsf{p} - \mathsf{P}) + \mathsf{Y}$ are used. Combining these equations and rearranging terms leads to the following expression for profits:

$$\mathscr{P} = -\frac{1}{2}\varepsilon(1+\varepsilon\gamma)\bar{x}\bar{P}\bar{Y}\left(\mathsf{p} - \left(\mathsf{P} + \frac{1}{1+\varepsilon\gamma}\mathsf{x}\right)\right)^2 + \text{t.i.p.} + \mathscr{O}(3) \ ,$$

where $\mathsf{x} = \gamma\mathsf{Y}$ is real marginal cost averaged over all firms.

[15] See Woodford (2003) for a derivation of this equation.