Sales and Monetary Policy*

Bernardo Guimaraes Kevin D. Sheedy†
London School of Economics

First draft: 8th December 2007
This draft: 8th September 2009

Abstract

A striking fact about pricing 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 customers 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.

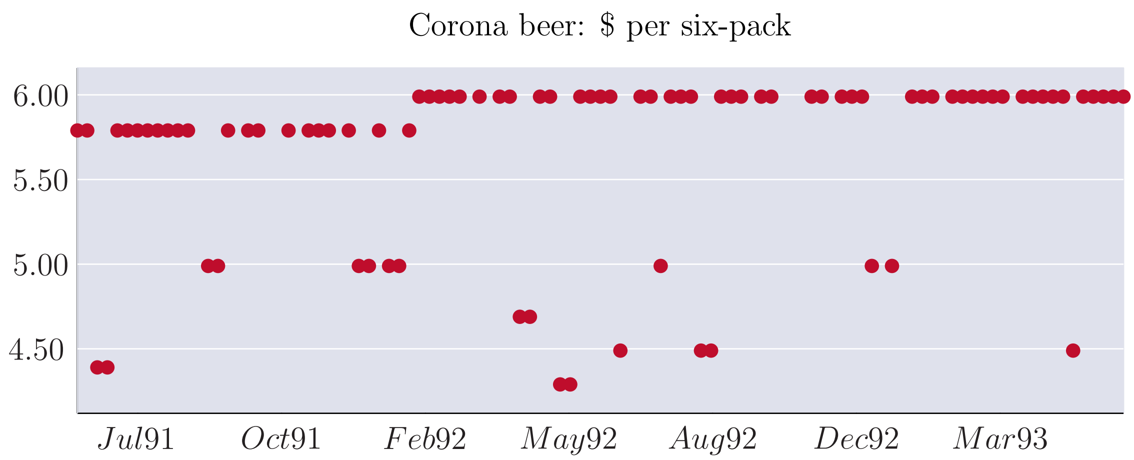
*We thank Ruediger Bachmann, Ariel Burstein, Andy Levin, Ananth Ramanarayanan, and especially Rachel Ngai for helpful discussions and suggestions. The paper has also benefited from suggestions made by three anonymous referees and the editor. We are grateful for comments from seminar and conference participants at: Amsterdam, Austrian National Bank, Bristol, Cergy-Pontoise, Essex, European Central Bank, Fundação Getúlio Vargas – Rio, Humboldt, London School of Economics, Paris School of Economics, and the Toulouse School of Economics; CEP conference 2009, CEP/ESRC Monetary Policy conference, COOL 2008, ESSIM 2009, LACEA 2008, London Business School conference, MMF research group conference 2008, RES 2009, Midwest Macro Meetings 2009, NBER Monetary Economics Fall meeting 2008, SED 2009, and the Texas Monetary Conference 2008.

†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 pricing is that many price changes are "sales": large temporary 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 (2008) 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.

In the IO and marketing literatures, the most prominent theories of sales are based on cus-

[1] See Hosken and Reiffen (2004), Klenow and Kryvtsov (2008), Nakamura and Steinsson (2008), Kehoe and Midrigan (2008), 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 stable 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).

tomer heterogeneity together with incomplete information. Leading examples include Salop and Stiglitz (1977, 1982), Varian (1980), Sobel (1984) and Narasimhan (1988). This paper builds a general-equilibrium macroeconomic model with sales that draws upon the rationale proposed in these literatures. Despite substantial heterogeneity at the microeconomic level, the model is easily aggregated to study macroeconomic questions.

The model assumes households have different preferences over goods, and for each good, some households are more price sensitive than others. There are two types: loyal customers 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 customers 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 prefer to sell their goods at high prices at some moments and at low sale prices at other moments. The choice of different prices at different moments is a profit-maximizing strategy even in an entirely deterministic environment. Firms would like to price discriminate, but as this is impossible, their best alternative strategy is holding periodic sales in order to target the two types of customers at different moments.

The existence of consumers with different price elasticities leads to sales being strategic substitutes: 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 increasing in the extent to which other firms are holding sales; in contrast, a firm can rely more on its loyal customers, whose purchases are much less sensitive to other firms' pricing decisions. Owing to sales being strategic substitutes, the resulting market equilibrium features a balance between the fractions of time firms spend targeting the two groups of consumers.

Given the pattern of price adjustment documented in Figure 1, changes in the aggregate price level can come from three sources: changes in "normal" prices, changes in the size of sale discounts, and changes in the proportion of goods on sale. Having built a model of sales, the key question to be answered is: for the purposes of monetary policy analysis, does it matter that normal prices are sticky amidst all the flexibility due to sales seen in Figure 1?

To tackle this question, the paper embeds the model of sales into a fully-fledged DSGE framework. Firms' normal prices are reoptimized at staggered intervals, but sales decisions are completely flexible and subject to no adjustment costs. Individual price paths generated by this model are similar to real-world examples such as that in Figure 1, even though no idiosyncratic shocks are assumed. This dynamic model with sticky normal prices but flexible sales is tractable, and an expression for the resulting Phillips curve is derived analytically. It is shown that flexible sales will never mimic fully flexible prices in equilibrium.

The model is then calibrated to match some simple facts about sales and hence assess quantitatively the real effects of monetary policy. The results are compared to those from the same DSGE model without sales incorporating a standard New Keynesian Phillips curve instead. The real effects of monetary policy in a model with sticky normal prices and fully flexible sales are similar to those found in a standard model with sticky prices and no sales. The cumulated response of output to a monetary policy shock in the model with fully flexible sales is 89% of the cumulated response in the standard model. The flexibility due to sales seen at the level of individual prices imparts little flexibility to the aggregate price level. These numerical results are not particularly sensitive to the calibration of the model.

The strong real effects of monetary policy follow from sales being strategic substitutes. 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, as the shock is common to all firms, 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. Thus the aggregate price level adjusts by little, so monetary policy has large real effects.

This analysis has so far assumed that sales are uniformly distributed across the whole economy. However, the evidence demonstrates this is not the case: sales are rare in some sectors and very frequent in others. A tractable two-sector version of the model is built to take account of this. Pricing behaviour in one sector features sales for the reasons described earlier. The other sector features standard pricing behaviour with no sales. Analytically, the two-sector model always implies larger real effects of monetary policy than the one-sector model of sales when the overall extent of sales is the same. Quantitatively, the model is recalibrated to account for the concentration of sales in certain sectors. The cumulated response of output to a monetary shock is now 96% of the response in a standard model without sales. Taking this as the more realistic representation of sales in the economy, it is fair to conclude that sales are essentially irrelevant for monetary policy analysis.

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 (2008). In their model, firms face different menu costs depending on whether they make permanent or temporary 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. The model of sales is introduced in section 2, and the equilibrium of the model is characterized in section 3. Section 4 embeds sales into a DSGE model and analyses the real effects of monetary policy. Section 5 presents the two-sector extension of the model. Section 6 draws some conclusions.

2 The model

2.1 Households

There is a measure-one continuum of households (indexed by $\imath$) with lifetime utility function

$$\mathscr{U}_t(\imath) = \sum_{\ell=0}^{\infty} \beta^{\ell} \mathbb{E}_t \left[\upsilon(C_{t+\ell}(\imath)) - \nu(H_{t+\ell}(\imath)) \right], \tag{2.1}$$

where $C(\imath)$ is consumption of household $\imath$'s specific basket of goods (defined below), $H(\imath)$ is hours of labour supplied, and β is the subjective discount factor ($0 < \beta < 1$). The function $\upsilon(C)$ is strictly increasing and strictly concave in C; $\nu(H)$ is strictly increasing and convex in H.

Household $\imath$'s budget constraint in time period t is

$$P_t(\imath)C_t(\imath) + M_t(\imath) + \mathbb{E}_t\left[\mathscr{A}_{t+1|t}\mathcal{A}_{t+1}(\imath)\right] = W_t H_t(\imath) + \mathfrak{D}_t + \mathfrak{T}_t + M_{t-1}(\imath) + \mathcal{A}_t(\imath), \qquad [2.2a]$$

where $P(\imath)$ is the money price of one unit of household $\imath$'s consumption basket, W is the money wage, $M(\imath)$ is household $\imath$'s end-of-period money balances, $\mathfrak{D}$ is dividends received from firms, $\mathfrak{T}$ is the net monetary transfer received by each household from the government, $\mathscr{A}$ is the asset-pricing kernel (in money terms), and $\mathcal{A}(\imath)$ is household $\imath$'s portfolio of money-denominated Arrow-Debreu securities.[4] All households have equal initial financial wealth and the same expected lifetime income. Household $\imath$ is also faces a cash-in-advance constraint on consumption purchases:

$$P_t(\imath)C_t(\imath) \leq M_{t-1}(\imath) + \mathfrak{T}_t. \qquad [2.2b]$$

Maximizing lifetime utility [2.1] subject to the sequence of budget constraints [2.2a] implies the following first-order conditions for consumption $C(\imath)$ and hours $H(\imath)$:

$$\beta\frac{\upsilon_c(C_{t+1}(\imath))}{\upsilon_c(C_t(\imath))} = \mathscr{A}_{t+1|t}\frac{P_{t+1}(\imath)}{P_t(\imath)}, \quad \text{and} \quad \frac{\nu_h(H_t(\imath))}{\upsilon_c(C_t(\imath))} = \frac{W_t}{P_t(\imath)}. \qquad [2.3]$$

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

$$1 + i_t = \left(\mathbb{E}_t\mathscr{A}_{t+1|t}\right)^{-1}. \qquad [2.4]$$

The net transfer $\mathfrak{T}_t$ is equal to the change in the money supply $\boldsymbol{\Delta}M_t \equiv M_t - M_{t-1}$. The cash-in-advance constraint [2.2b] binds when the nominal interest rate i_t is positive.

2.2 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. For each product type $\tau \in \mathscr{T}$, there is a measure-one continuum $\mathscr{B}$ of brands, with individual brands indexed by $\mathrm{b} \in \mathscr{B}$. For example, product types could include beer and dessert, and brands could be Corona beer or Ben & Jerry's ice cream.

Households have different preferences over this range 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 $\mathcal{B}(\tau)$. 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,

[4] These assumptions on asset markets are standard and play no important role in the model.

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 defined first in terms of a Dixit-Stiglitz aggregator over product types with elasticity of substitution ϵ. For a product type 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

$$C \equiv \left(\int_{\Lambda} c(\tau, \mathcal{B}(\tau))^{\frac{\epsilon-1}{\epsilon}} d\tau + \int_{\mathscr{T}\backslash\Lambda} \left(\int_{\mathscr{B}} c(\tau, \mathrm{b})^{\frac{\eta-1}{\eta}} d\mathrm{b} \right)^{\frac{\eta(\epsilon-1)}{\epsilon(\eta-1)}} d\tau \right)^{\frac{\epsilon}{\epsilon-1}}, \qquad [2.5]$$

where $c(\tau, \mathrm{b})$ is the household's consumption of brand b of product type τ.[5] It is assumed that $\eta > \epsilon$, 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 a product type for which they are loyal to a particular brand.

The elasticities ϵ and η are common to all households, as is the form of the consumption aggregator [2.5]. Furthermore, the measure of the set Λ of product types for which a household is loyal to a brand is the same across all households. This measure is denoted by λ, and it is assumed that $0 < \lambda < 1$. Hence, each household's preferences feature some mixture of loyal and bargain-hunting behaviour for different product types. The particular product types for which a household is loyal, and the particular brands receiving its loyalty, are randomly and independently assigned once and for all with equal probability. For example, one household may be loyal to Corona beer and a bargain hunter for desserts, while another may be loyal to Ben & Jerry's ice cream but a bargain hunter for beer. After aggregation, such idiosyncrasies of households' preferences are irrelevant; all that matters is households' common distribution of loyal and bargain-hunting behaviour over the whole set of goods.

Each discrete time period t contains a measure-one continuum of shopping moments when goods are purchased and consumed. A household does all its shopping at a randomly and independently chosen moment. As shown later, all households are indifferent in equilibrium between all shopping moments in the same time period.

Let $p(\tau, \mathrm{b})$ be the price of brand b of product type τ prevailing at a given household's shopping moment. The minimum expenditure required to purchase one unit of the composite good [2.5] is

$$P = \left(\int_{\Lambda} p(\tau, \mathcal{B}(\tau))^{1-\epsilon} d\tau + \int_{\mathscr{T}\backslash\Lambda} \left(\int_{\mathscr{B}} p(\tau, \mathrm{b})^{1-\eta} d\mathrm{b} \right)^{\frac{1-\epsilon}{1-\eta}} d\tau \right)^{\frac{1}{1-\epsilon}}. \qquad [2.6]$$

[5]This formulation captures the idea that different brands of a product type are not perfect substitutes even to bargain hunters. The assumption that bargain hunters have a Dixit-Stiglitz aggregator over brands, rather than making a discrete choice of brand, is inessential to the results. An earlier version of this paper experimented with a discrete choice of brand, but found qualitatively and quantitatively very similar results.

The expenditure-minimizing demand functions are

$$c(\tau,\mathrm{b}) = \begin{cases} \left(\frac{p(\tau,\mathrm{b})}{p_B(\tau)}\right)^{-\eta}\left(\frac{p_B(\tau)}{P}\right)^{-\epsilon} C & \text{if } \tau \in \mathscr{T}\backslash\Lambda, \quad \text{where } p_B(\tau) \equiv \left(\int_{\mathscr{B}} p(\tau,\mathrm{b})^{1-\eta} d\mathrm{b}\right)^{\frac{1}{1-\eta}}, \\ \left(\frac{p(\tau,\mathrm{b})}{P}\right)^{-\epsilon} C & \text{if } \tau \in \Lambda \text{ and } \mathrm{b} = \mathcal{B}(\tau), \\ 0 & \text{if } \tau \in \Lambda \text{ and } \mathrm{b} \neq \mathcal{B}(\tau), \end{cases} \quad [2.7]$$

where C is the amount of the composite good consumed, and P is the price level given in [2.6].[6] The term $p_B(\tau)$ is an index of prices for all brands of product type τ, as is relevant to those households who are bargain hunters for that product type. Total expenditure on all goods is PC. It is assumed that $\epsilon > 1$ to ensure the demand functions faced by firms are always price elastic.

As shown later, a firm will not charge the same price for its good at all shopping moments in a given time period. At each moment, it will randomly draw a price from some desired price distribution. When this distribution is common to all firms, the price index for bargain hunters is the same for all product types and at all shopping moments, that is, $P_B = p_B(\tau)$, and all households' price levels are the same and equal across all shopping moments, that is, $P(\imath) = P$. Thus there is a general price level P in spite of households' individual consumption baskets all differing.[7]

Given that households share a common price level, have the same preferences [2.1] over their composite goods and hours, and have the same initial wealth and expected lifetime income, all households choose the same levels of composite consumption and hours, hence $C(\imath) = C$ and $H(\imath) = H$ for all $\imath$. Since consumption is the only source of demand in the economy, goods market equilibrium requires $C = Y$, where Y is aggregate output.

2.3 Firms

Each brand b of each product type τ is produced by a single firm. All firms have the same production function

$$Q = \mathcal{F}(H), \quad [2.8]$$

where $\mathcal{F}(\cdot)$ is a strictly increasing function with $\mathcal{F}(0) = 0$. Generally, $\mathcal{F}(\cdot)$ is assumed to be strictly concave, though the milder assumption of weak concavity is used at some points in the paper. The minimum total money cost $\mathscr{C}(Q;W)$ of producing output Q for a given money wage W is

$$\mathscr{C}(Q;W) = W\mathcal{F}^{-1}(Q). \quad [2.9]$$

The cost function $\mathscr{C}(Q;W)$ is strictly increasing and generally strictly convex in Q, and satisfies $\mathscr{C}(0;W) = 0$.

[6] In the case where the household is loyal, the demand function should be interpreted as a density over a one-dimensional set, as with standard Dixit-Stiglitz preferences. When the household is a bargain hunter, the demand function should be interpreted as a density over a two-dimensional set.

[7] The price indices are the same across product types, shopping moments and households under the much weaker condition that the *distribution* of firms' price distributions is the same across product types and shopping moments. This condition is satisfied at all points in the paper.

Production takes place at the beginning of each discrete time period. Firms hold inventories during the period and sell some output at every shopping moment, but not necessarily at the same price at all moments. This captures the fact that firms can sell a batch of output at multiple prices.[8]

At a particular shopping moment, the quantity sold by the producer of good (τ, b) at price p is obtained by aggregating customers' demand functions from [2.7]:

$$\int c(\tau, \mathrm{b}) d\imath = \left(\lambda p^{-\epsilon} + (1-\lambda) P_B^{\eta-\epsilon} p^{-\eta}\right) P^{\epsilon} C,$$

where $P_B = p_B(\tau)$ is the common value of the bargain hunters' price index. The first term corresponds to demand from loyal customers and the second term to demand from bargain hunters for the same product type as the firm's own brand.[9]

It is helpful to state a firm's demand function at a shopping moment $\mathscr{D}(p; P_B, \mathcal{E})$ in terms of factors that shift it proportionately and factors that have differential effects depending on the price being charged by the firm at that particular moment:

$$\mathscr{D}(p; P_B, \mathcal{E}) = (\lambda + (1-\lambda)\mathrm{v}(p; P_B))\, p^{-\epsilon} \mathcal{E}, \quad \text{where } \mathrm{v}(p; P_B) \equiv \left(\frac{p}{P_B}\right)^{-(\eta-\epsilon)} \quad \text{and } \mathcal{E} \equiv P^{\epsilon} C. \qquad [2.10]$$

The aggregate component of the firm-level demand function is $\mathcal{E}$. The function $\mathrm{v}(p; P_B)$, referred to as the *purchase multiplier*, 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. In a model with standard Dixit-Stiglitz preferences, the actions of other firms are subsumed exclusively into $\mathcal{E}$, and this term proportionately scales demand; here, there is an additional channel through P_B via which other firms' actions matter, and one that affects demand from loyal customers and bargain hunters differently. Consequently, P_B does not have a uniform effect on demand at all prices.

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

$$\mathscr{R}(q; P_B, \mathcal{E}) \equiv q \mathscr{D}^{-1}(q; P_B, \mathcal{E}), \quad \text{with price } p = \mathscr{D}^{-1}(q; P_B, \mathcal{E}), \qquad [2.11]$$

where $\mathscr{D}^{-1}(q; P_B, \mathcal{E})$ is the inverse demand function corresponding to [2.10].

The profit-maximization problem for a firm consists of choosing the distribution of prices used across shopping moments. Let $F(p)$ be a general distribution function for prices. This distribution function is chosen to maximize profits

$$\mathscr{P} = \int_p \mathscr{R}\left(\mathscr{D}(p; P_B, \mathcal{E}); P_B, \mathcal{E}\right) dF(p) - \mathscr{C}\left(\int_p \mathscr{D}(p; P_B, \mathcal{E}) dF(p); W\right), \qquad [2.12]$$

where the first integral aggregates revenue $\mathscr{R}(q; P_B, \mathcal{E})$ over all shopping moments, and second term

[8] It is assumed for simplicity that firms can only hold inventories within a time period.

[9] There is a continuum of bargain hunters, each of which is a customer for all brands of a product type, so the two terms in the demand function are commensurable.

is the total cost $\mathscr{C}(Q; W)$ of producing the whole batch of output Q, which is equal to demand aggregated over all moments.

Consider a discrete distribution of prices $\{p_i\}$ with weights $\{\omega_i\}$.[10] The first-order conditions for maximizing profits [2.12] with respect to prices p_i and weights ω_i are

$$\mathscr{R}'\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) = \mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{and} \tag{2.13a}$$

$$\mathscr{R}\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) = \aleph + \mathscr{D}(p_i; P_B, \mathcal{E})\mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{if } \omega_i > 0; \quad \text{and} \tag{2.13b}$$

$$\mathscr{R}\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) \leq \aleph + \mathscr{D}(p_i; P_B, \mathcal{E})\mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{if } \omega_i = 0, \tag{2.13c}$$

where $\aleph$ is the Lagrangian multiplier attached to the constraint $\sum_j \omega_j = 1$. Equation [2.13a] is the usual marginal revenue equals marginal cost condition, which must hold for any price that receives positive weight. As discussed later, [2.13b] requires a firm to be indifferent between any prices receiving positive weight, and [2.13c] requires any price not used to be weakly dominated by some price receiving positive weight.

Observe that the first-order conditions are the same for all firms, therefore a price distribution over shopping moments that maximizes profits for one firm equally well maximizes profits for any other firm. Moreover, having chosen a price distribution, given that the demand function is the same at all shopping moments, random draws of prices from this distribution at each moment are consistent with profit maximization. Finally, note that randomization by firms makes all households indifferent between all shopping moments, as was claimed earlier.

3 Equilibrium with flexible prices

There are two steps to characterizing the equilibrium. The first is the profit-maximizing pricing policy of an individual firm conditional on the behaviour of others. The second is the strategic interaction among firms. The latter turns out to be essential for understanding the results.

3.1 Profit-maximizing price distributions

Firms choose a price distribution across shopping moments. If households had standard Dixit-Stiglitz preferences, which imply a constant price elasticity of demand, then the marginal revenue function would be strictly decreasing in quantity sold and the profit function would be strictly concave in price. Thus choosing a single price for all shopping moments would be strictly preferable to any price distribution.

However, in the model presented here, firms may prefer to randomize across shopping moments, that is, choose a non-degenerate price distribution. The reason is that the model features a price

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

elasticity that decreases with price, potentially leading to a non-monontonic marginal revenue, in which case the profit function ceases to be globally concave. This can be seen from the following identity:

$$\text{Marginal revenue} \equiv \left(1 - \frac{1}{\text{Price elasticity}}\right) \times \text{Price}.$$

With the price elasticity decreasing in price, the two terms on the right-hand side move in opposite directions.

As demand in the model comes from two different sources, loyal customers and bargain hunters, and these groups have different price sensitivities, the price elasticity of demand changes with the composition of a 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 the price elasticity is high.[11]

The price elasticity $\zeta(p; P_B)$ implied by the demand function $\mathscr{D}(p; P_B, \mathcal{E})$ in [2.10] is

$$\zeta(p; P_B) = \frac{\lambda\epsilon + (1-\lambda)\eta \mathrm{v}(p; P_B)}{\lambda + (1-\lambda)\mathrm{v}(p; P_B)}. \qquad [3.1]$$

This price elasticity is a weighted average of ϵ and η, with the weight on the larger elasticity η increasing with the purchase multiplier $\mathrm{v}(p; P_B)$, as defined in [2.10]. The higher is the price p, the lower is the purchase multiplier, and the smaller is the price elasticity.[12]

Marginal revenue is non-monotonic when η is sufficiently large relative to ϵ. This case is depicted in Figure 2. For very low prices, the price 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 an intermediate region there is a smooth transition between ϵ and η, and this increase in price elasticity can be large enough to make marginal revenue positively sloped, although it has its usual negative slope outside this intermediate range.

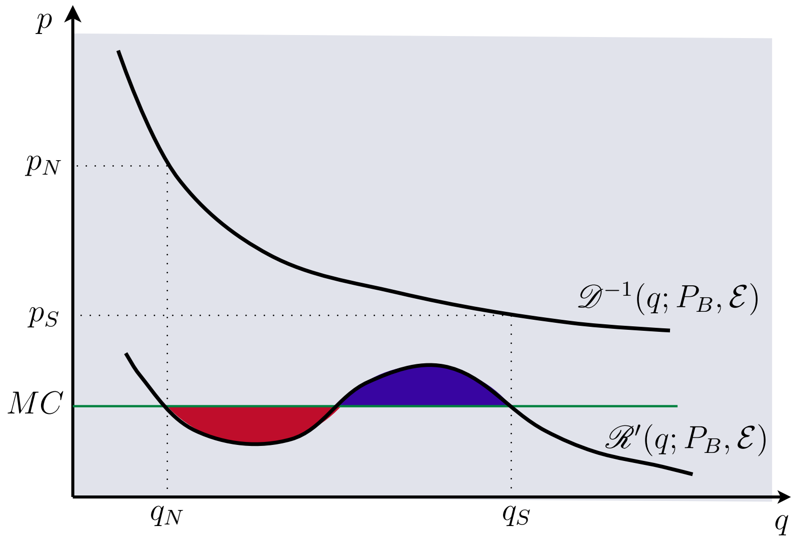
For some parameters ϵ, η and λ, firms find it optimal to choose a distribution with two prices: a *normal* high price, and a low *sale* price. Denote these two prices respectively by p_N and p_S, and let $q_N = \mathscr{D}(p_N; P_B, \mathcal{E})$ and $q_S = \mathscr{D}(p_S; P_B; \mathcal{E})$ be the quantities demanded at a single shopping moment at these prices. The frequency of sales (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 [2.13a]–[2.13b]. By eliminating the Lagrangian multiplier $\aleph$ from [2.13b], profit maximization requires:

$$\mathscr{R}'(q_N; P_B, \mathcal{E}) = \mathscr{R}'(q_S; P_B, \mathcal{E}) = \frac{\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}(q_N; P_B, \mathcal{E})}{q_S - q_N} = \mathscr{C}'\left(sq_S + (1-s)q_N; W\right). \qquad [3.2]$$

[11] This change in price elasticity along the demand function is a less extreme version of a "kinked" demand curve. The difference between the demand function in this paper and the "smoothed-kink" of Kimball (1995) is that there, the elasticity increases with price, whereas here it decreases with price. The behaviour of the price elasticity here is a consequence of aggregation, not a direct assumption.

[12] More generally, it can be shown that the price elasticity of demand is everywhere decreasing in price when demand is aggregated from any distribution of constant-elasticity individual demand functions.

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



Notes: Schematic representation of demand function $\mathscr{D}(p; P_B, \mathcal{E})$ from [2.10] and marginal revenue function $\mathscr{R}'(p; P_B, \mathcal{E})$ from [2.11] when η is sufficiently large relative to ϵ.

There are three requirements for the optimality of this price distribution and these are represented graphically in Figure 2. First, marginal revenue must be equalized at both normal and sale prices.[13] Second, 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. This is represented in the figure by the equality of the two shaded areas bounded between the marginal revenue function and the horizontal line MC, and between the quantities q_N and q_S. Finally, marginal revenue and average extra revenue must both be equal to the marginal cost of producing total output (for simplicity, the marginal cost function is not shown in the figure).

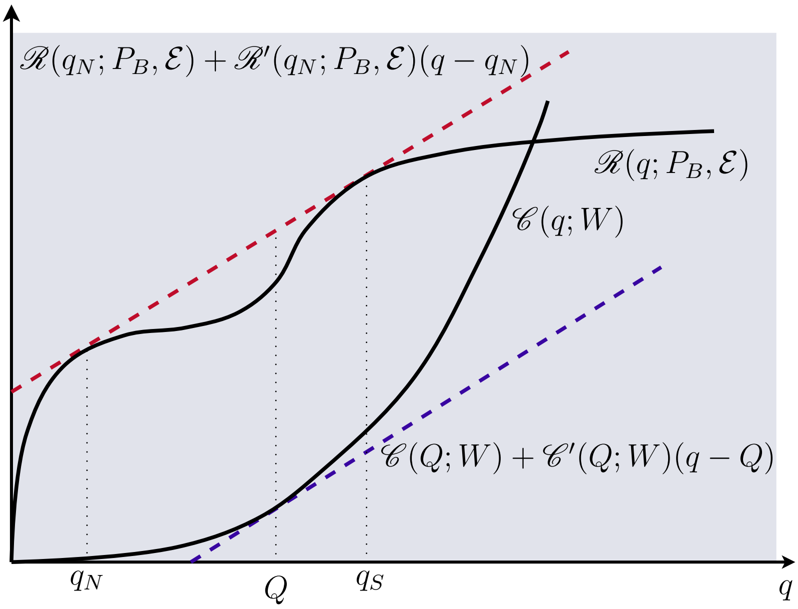
Firms have a choice at which shopping moment they sell each unit of their output, so switching a unit from one moment to another must not increase total revenue, thus marginal revenue must be equalized at all prices used at some shopping moment. Furthermore, firms must be indifferent between holding 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.

The full set of first-order conditions in [3.2] is depicted using the revenue and total cost functions in Figure 3. As firms can charge different prices at different shopping moments, the set of achievable total revenues is convexified. This raises attainable revenue in the range between q_N and q_S. The first two conditions for profit maximization in [3.2] require that the 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 common tangent itself. This slope is then associated with a total quantity sold $Q = sq_S + (1-s)q_N$ where marginal cost equals the common value of marginal revenue, which in

[13]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.

turn corresponds to a value of the sale frequency s.

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



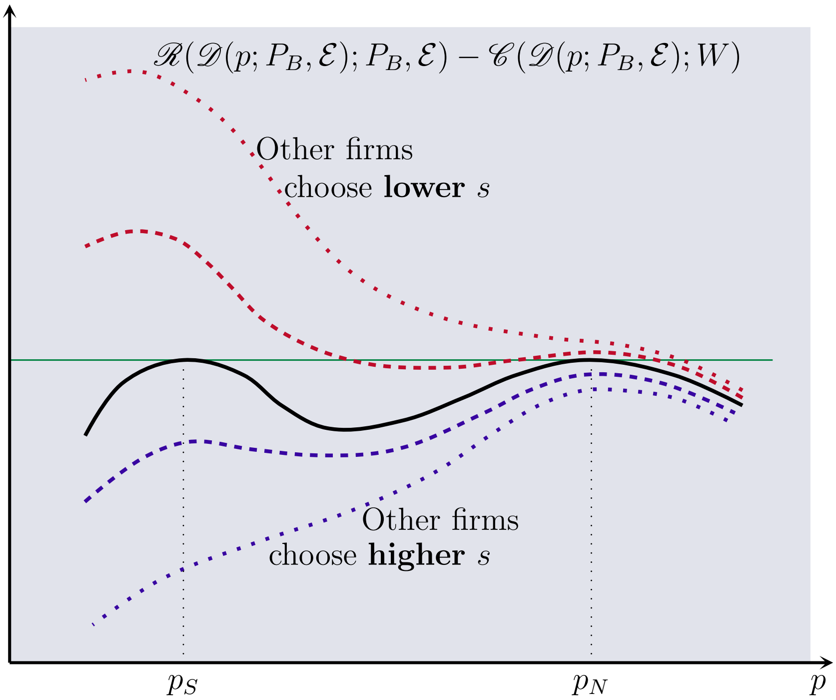
Notes: Schematic representation of the revenue function $\mathscr{R}(q; P_B, \mathcal{E})$ from [2.11] and total cost function $\mathscr{C}(Q; W)$ from [2.9], when η is sufficiently large relative to ϵ.

3.2 Strategic interaction

The figures depicting the first-order conditions for the choice of two prices may leave the impression that this is an unlikely case because it is necessary that both prices p_N and p_S simultaneously maximize profits. In particular, in the case of constant marginal cost, the first-order conditions in Figure 3 require the constant slope of the total cost function exactly to equal the slope of the tangent line to the revenue function, which may appear to hold only for a measure-zero set of parameters. However, this reasoning completely neglects the impact of other firms' actions, and the resulting strategic interaction among firms.

The effects of this strategic interaction are best illustrated in Figure 4. The figure plots the profits of a given firm as a function of its price at a single shopping moment in the simple case of constant marginal cost. Take the prices p_S and p_N that maximize profits from Figure 3. The solid curve in Figure 4 depicts the case where both prices simultaneously maximize profits, with both local maxima being of the same height. Let s denote the average sales frequencies of other firms. As s increases, profits at the sale price fall relative to profits at the higher normal price, which leads any individual firm strictly to prefer selling all its output at the normal price. Likewise, a lower s induces firms to sell only at the sale price. It is this strategic effect that guarantees a unique equilibrium in two prices for a wide range of parameters. In relation to Figure 3, the decisions of other firms about sales change the slope of the tangent line to the revenue function, bringing it into line with marginal cost in equilibrium.[14]

Figure 4: *Profits at a single moment, as affected by other firms' sale frequencies*



Notes: Schematic representation of profits as a function of the price charged at a single shopping moment, in the case where the total cost function $\mathscr{C}(Q; W)$ is linear, η is sufficiently large relative to ϵ, and λ is not too close to 0 or 1.

The reason why profits at one price relative to the other are affected by others' sales decisions in the way shown in Figure 4 is apparent from looking at the demand function in [2.10]. For high prices, the first term in λ corresponding to demand from loyal customers is dominant, while for low prices, the second term $(1-\lambda)\mathrm{v}(p; P_B)$ corresponding to demand from bargain hunters is more important. This is because the purchase multiplier $\mathrm{v}(p; P_B)$ is decreasing in price p, as demand from bargain hunters is much more sensitive to price. The strategic dimension of this equation comes from the presence of P_B. As other firms increase s, P_B falls, which has a negative impact on $\mathrm{v}(p; P_B)$ through demand from bargain hunters, but no effect on demand from loyal customers.[15] Therefore, other firms' sales decisions have a strong effect on profits from selling at low prices, but only a weak effect on profits at high prices.

Conditional on the marginal revenue function being non-monontonic, this strategic argument for sales depends only on a sufficient mixture of the two types of customer (a value of λ not very close to zero or one). If there were very few of one type of customer then the maximum attainable profits from a price aimed at the other type might always be larger irrespective of other firms' actions.

[14] The argument here is based on the case of constant marginal cost, but similar reasoning applies in the general case.

[15] Changing s also affects P, but this has a proportional effect on both groups' demand and hence on profits at all prices.

This is because the value of λ influences the relative height of the two local maxima of profits in addition to the pricing strategies of other firms.

The logic of the argument developed here implies that sales are strategic substitutes. The problem of choosing the profit-maximizing frequency of sales is essentially one of a firm deciding how much to target its loyal customers versus the 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.

Thus the varying composition of demand at different prices that gives rise to an equilibrium with sales also leads to strategic substitutability in sales decisions. This central feature of the model turns out to have important implications for monetary policy analysis.

3.3 Discussion

Although temporary sales have only recently caught the attention of macroeconomists, researchers in marketing have devoted a great deal of time and effort to them. This substantial literature is summarized by Neslin (2002). Most of the explanations for temporary sales rely on heterogeneity in the response of customers to price changes, for example, loyal customers versus bargain hunters (Narasimhan, 1988), or informed versus uninformed shoppers (Varian, 1980). Other explanations are based on behavioural aspects of consumer choice (Thaler, 1985) or habits (Nakamura and Steinsson, 2009). In Kehoe and Midrigan (2008), sales arise because temporary price changes are cheaper than changes to a product's regular price, in an environment where firms are subject to large and transitory idiosyncratic shocks. However, to the best of our knowledge, this proposed explanation for sales has not been entertained in the marketing literature.

In a recent study using a large retail-price dataset, Nakamura (2008) finds that most price variation is idiosyncratic, in that it is not common to stores in the same geographical area. This is particularly true of products for which there are frequent temporary sales. This evidence is consistent with randomization in the timing of sales as in the model here, but not with idiosyncratic shocks to costs or demand at the product level. The fact that many price changes are common to retailers of the same chain reinforces this point, as it is difficult to conceive of shocks specific to a product that affect only one chain, but all across a country.

Considering the conventionally assumed price elasticities in macroeconomics and the magnitude of sale discounts, it is unlikely that temporary sales would be a sensible strategy to react to idiosyncratic shocks that drive up inventories. Using a standard price elasticity of around 6, a discount of 25% would imply a fivefold increase in quantity sold. For a lower elasticity of 3, this discount still implies an increase in quantity sold of 137%. Idiosyncratic shocks would have to be huge to generate so much surplus inventory in a short space of time.

This paper captures the motivation for sales based on customer heterogeneity, but in a simple and tractable general equilibrium model suitable for addressing macroeconomic questions. While the ability of customer heterogeneity to explain temporary sales has been widely recognized, its implications for macroeconomics had not been analysed before.

By not making a distinction between producers and retailers, the model here shows that the total profits available to firms along the chain from producer to retailer are maximized using a pricing strategy involving temporary sales. The model abstracts from the division of these profits between producers and retailers. Empirical studies reveal that some sales are initiated by retailers, others by producers.

In addition to temporary sales, the phenomenon of clearance sales has also been analysed. Understanding the implications of clearance sales requires developing a different model (perhaps along the lines suggested by Lazear, 1986). But the typical price pattern shown in Figure 1, which is responsible for the bulk of the divergence between the estimated duration of a price spell in Bils and Klenow (2004) and Nakamura and Steinsson (2008), reflects temporary sales rather than clearance sales.

3.4 Characterizing the equilibrium

The following theorem gives existence and uniqueness results for the equilibrium of the model in a stationary environment where preferences, technology and the money supply are constant. All macroeconomic aggregates (though not individual prices) are constant, so time subscripts are dropped here.

Theorem 1 *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*

$$\eta > (3\epsilon - 1) + 2\sqrt{2\epsilon(\epsilon - 1)} \qquad [3.3]$$

holds, and everywhere decreasing otherwise. When elasticities ϵ and η are such that the above non-monotonicity condition holds, there exist thresholds $\underline{\lambda}(\epsilon, \eta)$ and $\overline{\lambda}(\epsilon, \eta)$ such that $0 < \underline{\lambda}(\epsilon, \eta) < \overline{\lambda}(\epsilon, \eta) < 1$ determining the type of equilibrium as follows:

(i) If ϵ and η satisfy the non-monotonicity condition [3.3] and $\lambda \in (\underline{\lambda}(\epsilon, \eta), \overline{\lambda}(\epsilon, \eta))$ then there exists a two-price equilibrium, and no other equilibria exist.

(ii) If ϵ and η violate the non-monontonicity condition [3.3] or $\lambda \notin (\underline{\lambda}(\epsilon, \eta), \overline{\lambda}(\epsilon, \eta))$ then there exists a one-price equilibrium, and no other equilibria exist.

Proof See appendix A.3 ■

Necessary and sufficient conditions for a two-price equilibrium with sales are that loyal customers and bargain hunters are sufficiently different (η is above a threshold depending on ϵ), and that there is a sufficient mixture of these two types of customer (λ is not too close to zero or one). The intuition for both of these conditions has already been discussed. Note that whether the cost function is strictly convex or not (and its curvature if so) plays no role in determining whether a two-price equilibrium prevails.

The model contains two types of consumer, but including more types would not necessarily generate a greater number of prices in equilibrium. From Figure 2, having more prices chosen in equilibrium requires more undulations of similar amplitude in the marginal revenue function, which is possible, but does not necessarily follow on augmenting the model with extra consumer types (even with a continuum of types). This is for the same reason that with two types of insufficiently different consumer, or where one consumer type predominates, the unique equilibrium might be in one price with no sales.

Now the two-price equilibrium is characterized. The total physical quantity of output sold by firms is $Q = sq_S + (1-s)q_N$ and the corresponding marginal cost is denoted by $X \equiv \mathscr{C}'(Q; W)$. Each of the markups on marginal cost associated with the two prices must satisfy the usual optimality condition in terms of the price elasticity of demand. What is new here is that two markups can satisfy this condition simultaneously. The optimal markup at price p is $\mu(p; P_B) = \zeta(p; P_B)/(\zeta(p; P_B) - 1)$. Using the price elasticity $\zeta(p; P_B)$ from [3.1], the first-order conditions for p_S and p_N are

$$p_S = \mu(p_S; P_B)X, \quad \text{and} \quad p_N = \mu(p_N; P_B)X, \quad \text{with} \quad \mu(p; P_B) = \frac{\lambda\epsilon + (1-\lambda)\eta\mathrm{v}(p; P_B)}{\lambda(\epsilon - 1) + (1-\lambda)(\eta - 1)\mathrm{v}(p; P_B)}. \tag{3.4}$$

The optimal markup function $\mu(p; P_B)$ depends on the parameters ϵ, η and λ, and the purchase multiplier $\mathrm{v}(p; P_B)$ from [2.10]. Let $v_S \equiv \mathrm{v}(p_S; P_B)$ and $v_N \equiv \mathrm{v}(p_N; P_B)$ denote the purchase multipliers at the two prices, and $\mu_S \equiv \mu(p_S; P_B)$ and $\mu_N \equiv \mu(p_N; P_B)$ the associated optimal markups:

$$\mu_S = \frac{\lambda\epsilon + (1-\lambda)\eta v_S}{\lambda(\epsilon - 1) + (1-\lambda)(\eta - 1)v_S}, \quad \text{and} \quad \mu_N = \frac{\lambda\epsilon + (1-\lambda)\eta v_N}{\lambda(\epsilon - 1) + (1-\lambda)(\eta - 1)v_N}. \tag{3.5}$$

The first-order condition for the sale frequency s is

$$(\mu_S - 1)q_S = (\mu_N - 1)q_N. \tag{3.6}$$

Given that a fraction s of all prices are at p_S and the remaining $1 - s$ are at p_N at any shopping moment, equation [2.7] implies the bargain hunters' price index is

$$P_B = \left(sp_S^{1-\eta} + (1-s)p_N^{1-\eta}\right)^{\frac{1}{1-\eta}}, \tag{3.7}$$

which is used to calculate the purchase multipliers and determine the optimal markups μ_S and μ_N.

In finding the stationary 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 frequency 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 $\chi \equiv q_S/q_N$.

Proposition 1 *Suppose parameters ϵ, η and λ are such that there is a unique two-price equilibrium.*

(i) The first-order conditions in [3.5] and [3.6] are necessary and sufficient to characterize the equilibrium price distribution (μ_S, μ_N, s).

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

(iii) The equilibrium values of s, υ_S and υ_N are functions only of the parameters ϵ, η and λ.

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

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

Proof See appendix A.4. ∎

The first part of the proposition shows that even though firms are maximizing a non-concave objective function, the first-order conditions are necessary and sufficient. The second and third parts establish the separation of the equilibrium for the microeconomic variables from the broader macroeconomic equilibrium, that is, the parameters ϵ, η and λ alone determine μ, χ and s.[16] The final part shows that the equilibrium sales frequency 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.

Proposition 1 also establishes that the purchase multipliers υ_S and υ_N and the markups μ_S and μ_N are determined by parameters ϵ, η and λ, hence finding the macroeconomic equilibrium is straightforward. The aggregate price level P is obtained by combining equation [2.6] and the demand function [2.7], and making use of the definition of the purchase multiplier $\mathrm{v}(p; P_B)$ from [2.10]:

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

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

$$x = \left(s(\lambda + (1-\lambda)\upsilon_S)\mu_S^{1-\epsilon} + (1-s)(\lambda + (1-\lambda)\upsilon_N)\mu_N^{1-\epsilon}\right)^{\frac{1}{\epsilon-1}}. \tag{3.8}$$

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:

$$q_S = (\lambda + (1-\lambda)\upsilon_S)\,\varrho_S^{-\epsilon} Y, \quad \text{and} \quad q_N = (\lambda + (1-\lambda)\upsilon_N)\,\varrho_N^{-\epsilon} Y. \tag{3.9}$$

Given that total physical output is $Q = sq_S + (1-s)q_N$, the ratio of Y to Q, denoted by $\mathit{\Delta}$, is

$$\mathit{\Delta} \equiv \frac{1}{s(\lambda + (1-\lambda)\upsilon_S)\varrho_S^{-\epsilon} + (1-s)(\lambda + (1-\lambda)\upsilon_N)\varrho_N^{-\epsilon}}, \tag{3.10}$$

[16] A solution method for μ, χ and s is described in appendix A.1.

which satisfies $0 < \mathit{\Delta} < 1$. The production function [2.8], cost function [2.9], and labour supply function [2.3] imply a positive relationship between real marginal cost x and aggregate output Y:

$$x = \frac{\nu_h\left(\mathcal{F}^{-1}(Y/\mathit{\Delta})\right)}{\upsilon_c(Y)\mathcal{F}'\left(\mathcal{F}^{-1}(Y/\mathit{\Delta})\right)}. \tag{3.11}$$

As the equilibrium real marginal cost x is already known from [3.8], the equation above uniquely determines output Y. Since the cash-in-advance constraint [2.2b] binds, the aggregate price level P is then given by $P = M/Y$. Finally, the interest rate is $i = (1-\beta)/\beta$.

4 Flexible sales with sticky normal prices

4.1 Staggered adjustment of normal prices

The model now developed allows firms costlessly to vary their sales frequencies and sale discounts, but adjustment times of their *normal* prices are staggered according to the Calvo (1983) pricing model. These assumptions are consistent with the stylized facts from micro price data discussed earlier. If there are in practice costs of adjusting sales through either frequency or discount size, this exercise will provide an upper bound for price flexibility in the aggregate.

The assumption of Calvo adjustment times for normal prices is made for simplicity. Of course, the choice of an alternative model of price adjustment, for example, state-dependent adjustment times for normal prices, 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 obtained below.

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 frequency and sale price for a firm at time t that last changed its normal price ℓ periods ago (referred to as a vintage-ℓ firm). The reset price $R_{N,t}$ is chosen to maximize the present value of a resetting firm calculated using the profit function [2.12] and stochastic discount factor $\mathscr{A}_{t+\ell|t}$:

$$\max_{R_{N,t}} \sum_{\ell=0}^{\infty} \phi_p^{\ell} \mathbb{E}_t \left[\mathscr{A}_{t+\ell|t} \left\{ \begin{array}{l} s_{\ell,t+\ell} p_{S,\ell,t+\ell} \mathscr{D}(p_{S,\ell,t+\ell}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) + (1-s_{\ell,t+\ell}) R_{N,t} \mathscr{D}(R_{N,t}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) \\ -\mathscr{C}\Big(s_{\ell,t+\ell}\mathscr{D}(p_{S,\ell,t+\ell}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) + (1-s_{\ell,t+\ell})\mathscr{D}(R_{N,t}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}); W_{t+\ell}\Big) \end{array} \right\} \right]. \tag{4.1}$$

Using the demand function [2.10], the total quantity $Q_{\ell,t}$ sold by a vintage-ℓ firm at time t is

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

The nominal marginal cost of such a firm is $X_{\ell,t} \equiv \mathscr{C}'(Q_{\ell,t}; W_t)$.

The first-order condition for the reset price $R_{N,t}$ maximizing the firm value [4.1] is

$$\sum_{\ell=0}^{\infty} \phi_p^\ell \mathbb{E}_t \left[(1 - s_{\ell,t+\ell}) \mathfrak{V}_{t+\ell|t} \left\{ \frac{R_{N,t}}{P_{t+\ell}} - \mu(R_{N,t}; P_{B,t+\ell}) \frac{X_{\ell,t+\ell}}{P_{t+\ell}} \right\} \right] = 0, \qquad [4.2]$$

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

This condition weights the sequence of one-period optimality conditions for the normal price over the expected lifetime of the price using a discount factor $\mathfrak{V}_{t+\ell|t}$. The profit-maximizing sales frequencies $s_{\ell,t}$ and sale prices $p_{S,\ell,t}$ are chosen to maximize profits [4.1] at all times, yielding first-order conditions:

$$\frac{p_{S,\ell,t} q_{S,\ell,t} - R_{N,t-\ell} q_{N,\ell,t}}{q_{S,\ell,t} - q_{N,\ell,t}} = X_{\ell,t}, \quad \text{and } \; p_{S,\ell,t} = \mu(p_{S,\ell,t}; P_{B,t}) X_{\ell,t}. \qquad [4.3]$$

Firms' pricing behaviour is aggregated as follows. Using equations [2.6], [2.7] and [2.10], an expression for the aggregate price level is

$$P_t = \left((1 - \phi_p) \sum_{\ell=0}^{\infty} \phi_p^\ell \left\{ \begin{array}{c} s_{\ell,t}(\lambda + (1 - \lambda) \mathrm{v}(p_{S,\ell,t}, P_{B,t})) p_{S,\ell,t}^{1-\epsilon} \\ + (1 - s_{\ell,t})(\lambda + (1 - \lambda) \mathrm{v}(R_{N,t-\ell}, P_{B,t})) R_{N,t-\ell}^{1-\epsilon} \end{array} \right\} \right)^{\frac{1}{1-\epsilon}}, \qquad [4.4]$$

and the bargain hunters' price index from [2.7] is given by

$$P_{B,t} = \left((1 - \phi_p) \sum_{\ell=0}^{\infty} \phi_p^\ell \left\{ s_{\ell,t} p_{S,\ell,t}^{1-\eta} + (1 - s_{\ell,t}) R_{N,t-\ell}^{1-\eta} \right\} \right)^{\frac{1}{1-\eta}}. \qquad [4.5]$$

Total labour demand from all firms is

$$H_t = \sum_{\ell=0}^{\infty} (1 - \phi_p) \phi_p^\ell H_{\ell,t}, \qquad [4.6]$$

where $H_{\ell,t} = \mathcal{F}^{-1}(Q_{\ell,t})$ is the amount of labour employed by a vintage-ℓ firm.

4.2 A Phillips curve with sales

Monetary policy is analysed by log linearizing the model around the flexible-price stationary equilibrium characterized in section 3. Denote log deviations of variables from their flexible-price stationary equilibrium value using the corresponding sans serif letters.

To study the dynamic implications of the sales model, it is helpful to derive a Phillips curve for aggregate inflation that can be compared to the New Keynesian Phillips curve resulting from a standard model with Calvo pricing. It turns out that the model with sales also yields a simple Phillips curve.[17]

[17] All the log deviations of the special features of the sales equilibrium (sale discount, sales frequency, quantity ratio, price distortions) are proportional in equilibrium to the log deviation of real marginal cost. This feature makes

Theorem 2 *Consider parameter values ϵ, η and λ for which the economy has a two-price equilibrium. Let $\pi_t \equiv P_t/P_{t-1}$ be the inflation rate for the aggregate price level [4.4].*

(i) The first-order conditions for the sale discount and the sale frequency imply

$$\mathsf{p}_{S,\ell,t} = \mathsf{X}_{\ell,t}, \quad \text{and } \mathsf{X}_{\ell,t} = \mathsf{P}_{B,t}, \tag{4.7}$$

which yield $\mathsf{p}_{S,\ell,t} = \mathsf{P}_{S,t}$, $\mathsf{X}_{\ell,t} = \mathsf{X}_t$, and thus $\mathsf{Q}_{\ell,t} = \mathsf{Q}_t$. The first-order condition for the reset price implies

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

(ii) The Phillips curve linking inflation $\pi_t = \mathsf{P}_t - \mathsf{P}_{t-1}$ and real marginal cost $\mathsf{x}_t = \mathsf{X}_t - \mathsf{P}_t$ is

$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \frac{1}{1-\psi}\left(\kappa\mathsf{x}_t + \psi\left(\boldsymbol{\Delta}\mathsf{x}_t - \beta\mathbb{E}_t\boldsymbol{\Delta}\mathsf{x}_{t+1}\right)\right), \tag{4.8}$$

where $\kappa \equiv \left((1-\phi_p)(1-\beta\phi_p)\right)/\phi_p$, and the coefficient ψ is a function only of ϵ, η, and λ. By solving forwards, inflation can also be expressed as

$$\pi_t = \frac{\kappa}{1-\psi}\sum_{\ell=0}^{\infty}\beta^{\ell}\mathbb{E}_t\mathsf{x}_{t+\ell} + \frac{\psi}{1-\psi}\boldsymbol{\Delta}\mathsf{x}_t. \tag{4.9}$$

(iii) The coefficient ψ satisfies $0 \le \psi \le 1$, but $\psi = 1$ can only occur if the sale discount is zero [$\mu = 1$], or goods are never off sale [$s = 1$], or the GDP share transacted at the normal price is zero [$(1-s)p_N q_N/(s p_S q_N + (1-s)p_N q_N) = 0$]. The value of ψ is strictly decreasing in λ.

Proof See appendix A.5. ∎

The first part of the theorem reflects the fact 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 [4.5] increases. This leads a given firm optimally to increase its total quantity sold by holding more sales to the point where marginal cost X has risen one-for-one in percentage terms with P_B.

The condition linking the bargain hunters' price index P_B and marginal cost X is novel. As has been discussed in section 3.2, a rise in P_B disproportionately benefits a firm selling at its sale price relative to one selling at its normal price. On the other hand, a rise in costs disproportionately hurts firms selling at low prices where demand is higher. No other variables (including the aggregate price level P) have this asymmetric effect, and since both P_B and X are nominal variables, the relationship between them must be one-for-one.[18]

The optimal sale price features a constant markup on marginal cost, at least locally, and the equation determining the optimal reset price is the same as in any standard application of Calvo

the model particularly tractable. More details on the decomposition of aggregate inflation movements are provided by Lemma 4 in appendix A.5.

[18] The individual sale and normal prices themselves only have second-order effects on profits by the envelope theorem.

pricing. The optimal adjustment of sales has the consequence that all firms produce the same total quantity, and thus have the same level of marginal cost.

The Phillips curve with sales in equation [4.8] would reduce to the standard New Keynesian Phillips curve $\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa x_t$ in the case that $\psi = 0$.[19] On the other hand, the case of a fully-flexible price level (a vertical short-run Phillips curve) is equivalent to $\psi = 1$. With parameters consistent with sales in equilibrium, ψ always lies strictly between these extremes. While varying sale frequencies and discounts can always generate the same average price change as a given adjustment of normal prices, in equilibrium, firms never find these to be perfect substitutes and so flexibility in sales never replicates full price flexility.

The effect of a positive value of ψ is to increase the response of inflation to real marginal cost to some extent when compared to the New Keynesian Phillips curve. This is best seen by looking at the solved-forwards version of the Phillips curve with sales in [4.9], where there are two distinct differences relative to the solved-forwards version of the standard New Keynesian Phillips curve: $\pi_t = \kappa \sum_{\ell=0}^{\infty} \beta^{\ell} \mathbb{E}_t X_{t+\ell}$. 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 $1 - \phi_p$. The second is the presence of a term in the growth rate of real marginal cost Δx_t. The growth rate appears in addition to the level because the extra margin of price adjustment operates through temporary sales rather than persistent changes to normal prices.

The analysis here is based on assumptions congruent with the micro pricing evidence (sticky normal prices, flexible sales), but are there good reasons for firms to set prices in this way? In the model, deviations of the sale and normal prices from their profit-maximizing levels would not be equally costly to firms. Both the price elasticity of demand and the quantity sold at a given shopping moment are higher at the sale price than at the normal price. This implies that for a given percentage deviation from their profit-maximizing levels, the benefits from reoptimizing the sale price would be higher than for the normal price.[20]

4.3 A DSGE model with sales

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

As in Erceg, Henderson and Levin (2000), firms hire differentiated labour inputs. So hours H in the production function [2.8] is now the composite labour input

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

where $H(\imath)$ is hours of type-$\imath$ labour supplied 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 minimum monetary cost of hiring one unit of the composite labour input H is denoted

[19] See Woodford (2003) for a derivation and discussion of the standard New Keynesian Phillips curve.
[20] This point is discussed further in an earlier working paper (Guimaraes and Sheedy, 2008).

by W, and this is now the relevant wage index appearing in firms' cost function [2.9].

Each household (supplying a particular type of labour) has a probability $1 - \phi_w$ of being able to adjust its money wage in any given time period. Since households have equal initial financial wealth and expected lifetime income, as asset markets are complete and utility [2.1] is additively separable between hours and consumption, households are fully insured and hence have equal consumption in equilibrium. Consumption is the only source of expenditure, so goods market equilibrium requires $C_t = Y_t$. Thus by using [2.3] and [2.4], and by noting that [2.2b] is binding, the following intertemporal IS equation and money-market equilibrium condition are obtained:

$$\beta(1+i_t)\mathbb{E}_t\left[\frac{\upsilon_c(Y_{t+1})}{\upsilon_c(Y_t)}\frac{1}{\pi_{t+1}}\right] = 1, \quad \text{and } \; Y_t = \frac{M_t}{P_t}. \tag{4.10}$$

The wage setting and wage index equations are as in Erceg, Henderson and Levin (2000).[21]

Finally, the model is closed by specifying a rule for monetary policy. The growth rate of the money supply M_t is assumed to follow the first-order autoregressive process

$$\frac{M_t}{M_{t-1}} = \left(\frac{M_{t-1}}{M_{t-2}}\right)^{\mathfrak{p}} \exp\left\{(1-\mathfrak{p})\mathsf{e}_t\right\}, \quad \text{where } \; \mathsf{e}_t \sim \text{i.i.d.}(0, \Omega_m). \tag{4.11}$$

4.4 Calibration

The distinguishing parameters of the sales model are the two elasticities ϵ and η and the fraction λ of loyal customers. As shown in Proposition 1, 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 frequency of sales s. Furthermore, the model has a convenient block recursive structure in that only ϵ, η and λ need to be known to determine these observables. There are thus three parameters that can be matched to data on just these three variables.[22]

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 frequency s. The baseline values of these variables are taken from Nakamura and Steinsson (2008). 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 that are sales (weighted by expenditure) is 7.4%, so $s = 0.074$ is used here.[23] They also report that the median difference between the logarithms of the normal and sale prices 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 research provides information about the quantity ratio. Papers typically report a range of estimates conditional on factors other than price that affect the impact

[21] See appendix A.6 for details of these equations.

[22] It is also possible to match these three parameters using data on the average markup instead of the quantity ratio. This approach would be in line with typical practice in macroeconomics, but the strategy adopted here is more direct.

[23] The sales frequency s is for the whole economy. Certain sectors have higher frequencies of sales and some sectors have none. The implications of such heterogeneity are considered in section 5.

of a price promotion, for example, advertising. The baseline value of the quantity ratio is obtained from the study by Narasimhan, 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 sale discount taken from Nakamura and Steinsson (2008) 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 $\chi = 5$.[24]

The three facts about sales are then used to find matching values of the three unknown parameters.[25] The results are shown in Table 1.

Table 1: *Calibration of the model of sales*

Description	Notation	Value
Stylized facts		
Ratio of the sale-price markup to the normal-price 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†
Frequency of sales	s	0.074*
Parameters		
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.
* *Source*: Nakamura and Steinsson (2008)
† *Source*: Narasimhan, Neslin and Sen (1996)

The remainder of the calibration is standard, drawing on conventional values from the DSGE literature. The parameter values selected are shown in Table 2. One time 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 θ_c is chosen to match a coefficient of relative risk aversion of 3, and the Frisch elasticity of labour supply θ_h is set to 0.7, which lies in the range of estimates found in the literature (Hall, 2009). The production function is $\mathcal{F}(H) = AH^{\alpha}$, where α is the elasticity of output with respect to hours. The value of α is chosen to match a labour share of 0.667. This production function implies that the elasticity γ of marginal cost with respect to output is given by $\gamma = (1-\alpha)/\alpha$. So $\alpha = 0.667$ yields $\gamma = 0.5$. The elasticity of substitution between labour inputs ς is taken from Christiano, Eichenbaum and Evans (2005). The probability ϕ_p of stickiness of the normal price is set to match an average price-spell duration of 9 months, which is taken from

[24] This quantity ratio is very close to what would be consistent with a price elasticity of 6 over the relevant range of the demand function. Levin and Yun (2009) find that substitution by consumers on the extensive margin between brands alone can account for elasticities of approximately this size.

[25] A procedure for calculating the equilibrium values of μ, χ, and s is described in appendix A.1. 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. An extensive grid search over the elasticities ϵ and η was used to verify that no other values are consistent with the targets for μ and χ. Proposition 1 demonstrates that given ϵ and η, there is always one and only one value of λ matching the target sales frequency s.

Nakamura and Steinsson (2008). 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 persistence parameter of money-supply growth $\mathfrak{p}$ is chosen to match the first-order autocorrelation coefficient of M1 growth in the U.S. from 1960:1 to 1999:12.

Table 2: *Calibration of the DSGE model parameters*

Description	Notation	Value
Preference parameters		
Subjective discount factor	β	0.9975
Intertemporal elasticity of substitution in consumption	θ_c	0.333
Frisch elasticity of labour supply	θ_h	$0.7^{\dagger}$
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	$20^{\ddagger}$
Nominal rigidities		
Probability of stickiness of normal prices	ϕ_p	$0.889^{\S}$
Probability of wage stickiness	ϕ_w	0.889
Monetary policy		
First-order serial correlation of the money-supply growth rate	$\mathfrak{p}$	$0.536^{\sharp}$

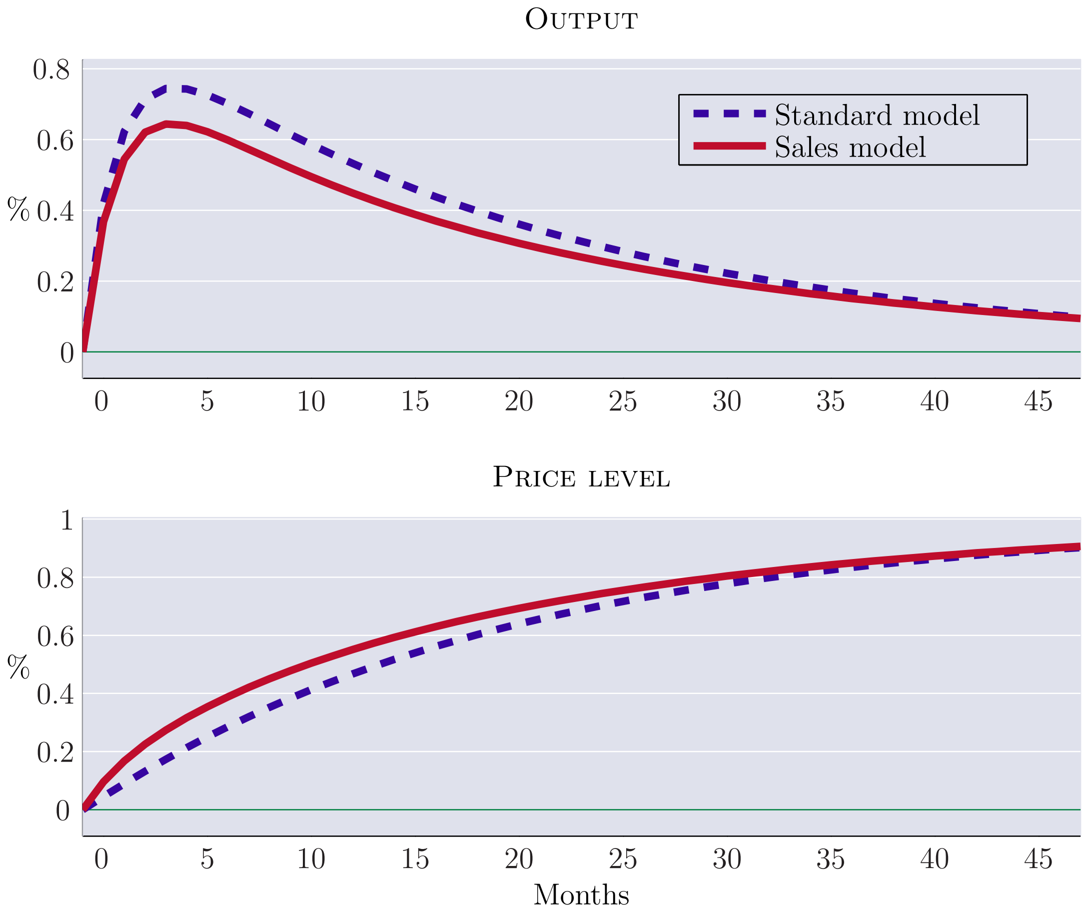
Notes: Monthly calibration.
† *Source*: Hall (2009)
‡ *Source*: Christiano, Eichenbaum and Evans (2005)
§ *Source*: Nakamura and Steinsson (2008)
$^{\sharp}$ *Source*: Authors' calculations using data on M1 for the period 1960:1–1999:1. Series `M1SL` from Federal Reserve Economic Data (`http://research.stlouisfed.org/fred2`).

4.5 Dynamic simulations

This section calculates the impulse responses of output and the price level to monetary policy shocks in the calibrated DSGE model with sales. 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 where price-adjustment times are staggered according to the Calvo (1983) model. With Calvo pricing, a standard New Keynesian Phillips curve is obtained. The latter model is set up so that it is otherwise identical to the DSGE model with sales.

The calibrated parameters of the DSGE model with sales are given in Table 1 and Table 2. For the standard DSGE model without sales, the same parameter values from Table 2 are used, with the probability of price stickiness ϕ_p applying to a firm's single price, rather than to its normal price in the sales model. In place of the parameters ϵ, η and λ, the standard model requires only a calibration of its constant price elasticity of demand ξ (the elasticity of substitution in the usual Dixit-Stiglitz aggregator). This is chosen to match the average markup (in the sense of the reciprocal of real marginal cost) from the calibrated sales model. For the baseline calibration this implies $\xi = 3.77$.

Figure 5: *Impulse responses to a persistent shock to money growth*



Notes: The model is as described in section 4. Parameter values are given in Table 1 and Table 2.

Figure 5 plots the impulse response functions of aggregate output and the price level to a serially-correlated money growth shock in both the sales model and the standard model without sales. 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 react to the shock by varying sales without cost. The ratio of the cumulated responses of output between the two models is 0.89.

Strategic substitutability in sales decisions is fundamental to understanding the real effects of monetary policy in the sales model. On the one hand, firms have an incentive to reduce sales in response to a positive monetary shock, essentially mimicking an increase in price. On the other hand, owing to strategic substitutability in sales, as other firms reduce their sales, an individual firm has a strong incentive to target the bargain hunters, who are being neglected by others. Thus there are two conflicting effects on sales and the price level after a monetary shock. One tends towards money neutrality, while the other tends towards money having real effects.

Quantitatively, finding the right balance between targeting their two groups of customers turns out to be much more important to firms' profits than using sales as a means of changing their average prices. Since there is a substantial gap between sale and normal prices on average, a relatively modest response of sales to a monetary shock would suffice to raise the price level in line with the money supply. However, strong strategic substitutability dissuades firms from adjusting sales in this way because all firms would need to respond in the same way to the aggregate monetary shock.[26]

The role of strategic substitutability can be isolated by considering instead an idiosyncratic demand shock to one single firm. Since this one firm is negligible, no other firms react, so the bargain hunters' price index P_B does not change. From the first-order condition in [4.7], the marginal cost of the affected firm must remain unchanged. Hence, the total quantity the firm produces is insulated from the demand shock through its adjustment of sales. This is in stark contrast to the small response of sales to aggregate demand shocks where strategic considerations dominate.

The robustness of these results is checked by performing a sensitivity analysis with respect to the key empirical targets used to calibrate the model: the markup ratio, the quantity ratio, and the sales frequency. A range of values for each around its baseline value from Table 1 is considered. One target is varied at a time while the others are held constant. The sensitivity analysis is extended to include the elasticity of output with respect to hours to explore the implications of different degrees of curvature of firms' cost functions.

Figure 6 depicts the ratio of the cumulated impulse response of output in the model with sales to that in the standard model as a function of each target, performing exactly the same monetary policy experiment described earlier.

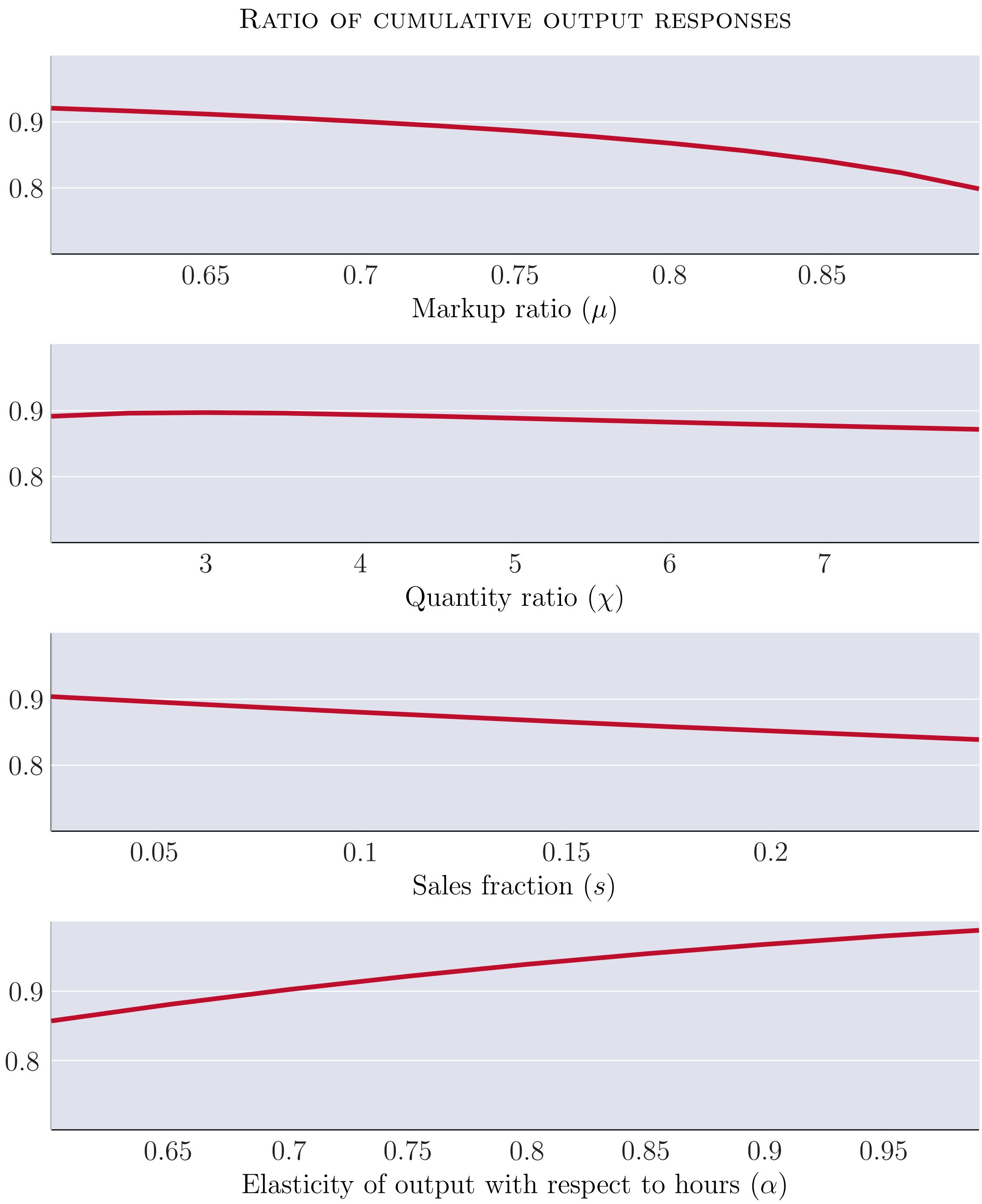
The impulse responses are not particularly sensitive to the calibration targets. The quantity ratio χ is the target for which the literature yields the widest range of estimates. But nonetheless, varying χ from 2 to 8 implies that the ratio of cumulated output responses lies only between 0.87 and 0.9. For the other targets, more precise data are available. By considering markup ratios from 0.65 to 0.85 (a wide band around the baseline value), the response ratio between the models varies from 0.84 to 0.91. Similarly, a wide range of sales frequencies from 0.05 to 0.15 yields ratios between 0.86 and 0.9.

Finally, for values of the elasticity of output with respect to hours above the baseline, all the way up to one, the ratio of cumulated output responses is higher than 0.89. In particular, as the elasticity gets close to one, the ratio approaches 0.99. This implies that when the cost function is close to being linear, the real effects of monetary policy are essentially the same in the model with fully flexible sales as in the standard model with no sales at all.

The intuition for this finding is that when the cost function is linear, marginal cost does not depend on the quantity of output produced. So a rise in aggregate demand, which if accommodated

[26] Although firms' sales are reacting only slightly to monetary shocks, the losses from failing to adjust the normal price more frequently are considerably smaller than they would otherwise be in a model without sales. The possibility of adjusting sales implies that the quantity produced by an individual firm would be exactly the same had this firm the option of adjusting its normal price in addition to adjusting its sales, as is shown in Theorem 2. Hence there are no undesirable fluctuations in marginal cost, and so the further gains from adjusting the normal price are smaller. This point is discussed further in an earlier working paper (Guimaraes and Sheedy, 2008).

Figure 6: *Sensitivity analysis for the real effects of monetary shocks*

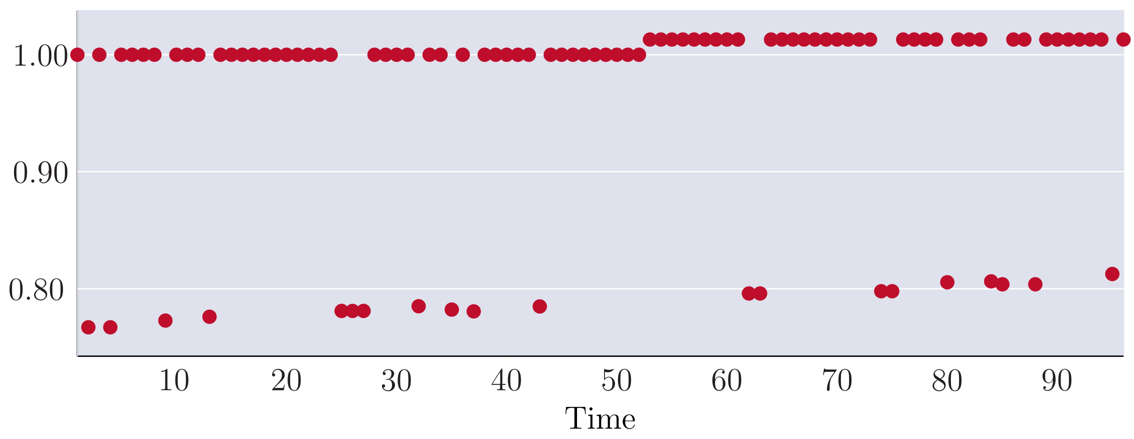


Notes: For each graph, the results are obtained by fixing the other targets at their baseline values as given in Table 1 (together with $\alpha = 2/3$) and choosing matching values of the parameters ϵ, η and λ as explained in section 4.4.

would increase the quantity sold, no longer provides firms with a reason to reduce sales. Hence all that matters for sales decisions is striking the right balance between targeting loyal customers and bargain hunters.

Figure 7 shows an example of an individual price path in the model with sales generated using

Figure 7: *A typical individual price path generated by the model*



Notes: Obtained 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.

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.

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.

Behind the findings of this section lies the fact that the equilibrium distribution of prices reacts little to monetary shocks. So while the occurrence of sales means that there is much more nominal flexibility of individual prices, the rationale for sales implies that there is an endogenous real rigidity constraining the adjustment of the relative prices in firms' price distributions.

5 Sectoral heterogeneity in sales

The model presented thus far assumes all sectors of the economy have the same pattern of sales. But sales are in fact concentrated in some sectors, and rare or non-existent in others. This creates a divergence between estimates of the frequency of sales using data covering the whole economy (Nakamura and Steinsson, 2008) and those based on scanner data from supermarkets. These findings suggest a multi-sector model is more empirically appropriate for analysing the implications of sales.

The model of section 2 is extended to include two sectors. In one sector, households have homogeneous Dixit-Stiglitz preferences over brands of product types, so no sales will occur in equilibrium. In the other sector, household preferences over brands are heterogeneous, with some mixture of loyal and bargain hunting behaviour, which will give rise to sales in equilibrium. This extension is simple and tractable.

A measure σ of product types are in the sale sector, with $\mathscr{T}$ now denoting the set of just these product types, and where household preferences are as described in section 2. The remaining set of product types with measure $1-\sigma$ in the non-sale sector is denoted by $\mathscr{H}$. The new composite good C replacing that in equation [2.5] is

$$C \equiv \left(\int_{\Lambda} c(\tau, \mathcal{B}(\tau))^{\frac{\epsilon-1}{\epsilon}} d\tau + \int_{\mathscr{T}\setminus\Lambda} \left(\int_{\mathscr{B}} c(\tau, \mathrm{b})^{\frac{\eta-1}{\eta}} d\mathrm{b} \right)^{\frac{\eta(\epsilon-1)}{\epsilon(\eta-1)}} d\tau + \int_{\mathscr{H}} \left(\int_{\mathscr{B}} c(\tau, \mathrm{b})^{\frac{\xi-1}{\xi}} d\mathrm{b} \right)^{\frac{\xi(\epsilon-1)}{\epsilon(\xi-1)}} d\tau \right)^{\frac{\epsilon}{\epsilon-1}}, \qquad [5.1]$$

where ξ is the homogeneous elasticity of substitution between brands in the non-sale sector for all households.

Two restrictions are imposed. First, the elasticity ξ is chosen to ensure the markup in the non-sale sector is equal to the economy's average markup (in the sense of the reciprocal of real marginal cost). This entails choosing $\xi = 1/(1-x)$, where x is calculated for the sale sector as in [3.8], as if it encompassed the whole economy. Second, the relative contributions of the sale and non-sale sectors to GDP must be proportional to σ and $1-\sigma$. Given that the sale sector features price distortions, it is not possible to satisfy these two restrictions when the production function is the same in both sectors. Consequently, a slight adjustment is made to the non-sale sector production function $\mathfrak{F}(H)$, but one which ensures it has the same elasticity of output with respect to hours α and elasticity of marginal cost with respect to output γ as the production function $\mathcal{F}(H)$ in the sale sector. These conditions are satisfied only when $\mathfrak{F}(H) = \Delta\mathcal{F}(\Delta^{-1}H)$, where Δ is the stationary equilibrium price distortion (ratio of Y to Q) in the sale sector from [3.10] (again, calculated as if this sector encompassed the whole economy). Since Δ is close to one in practice, the difference between the production functions is very small.

The characteristics of a sale when one occurs (the discount size, and the extra amount purchased) are the same here as in the earlier one-sector model. Proposition 1 shows that the markup ratio μ and the quantity ratio χ depend solely on the elasticities ϵ and η. So neither μ and χ, nor ϵ and η, change when moving from the one-sector to the two-sector model. The two-sector model allows for the sale sector to have an above-average frequency of sales s, while holding constant the average sales frequency $\bar{s} = \sigma s$ for the whole economy. The higher frequency within the sale sector is matched by a lower value of λ there than in the one-sector model. Finally, the extent of nominal rigidity (excluding sales) is equal across sectors, in the sense that price stickiness in the non-sale sector is the same as normal-price stickiness in the sale sector.

Proposition 2 *Let $\Psi(s; \epsilon, \eta)$ be the Phillips curve coefficient ψ from Theorem 2 implied by a sale frequency s, with parameters ϵ and η consistent with μ and χ, and with λ implicitly adjusted to match s, as if the sale sector encompassed the whole economy.*

(i) The function $\Psi(s; \epsilon, \eta)$ is strictly increasing and strictly concave in s.

(ii) In the case of constant marginal cost ($\gamma = 0$), the Phillips curve for aggregate inflation in the two-sector model is of exactly the same form as that in Theorem 2 with ψ replaced by the

weighted average of $\Psi(s;\epsilon,\eta)$ *(for the sale sector) and* 0 *(for the non-sale sector) using weights* σ *and* $1-\sigma$*. This weighted average is less than* $\Psi(\sigma s;\epsilon,\eta)$ *for all* $\sigma < 1$*.*

Proof See appendix A.8. ■

The first finding states that $\Psi(s;\epsilon,\eta)$, which can be interpreted as the amount of price-level flexibility resulting from adjustment of sales, is increasing in the frequency with which sales occur, but at a diminishing rate. In a multi-sector context, what matters for aggregate price flexibility is mainly the weighted average of the value of ψ across sectors. Therefore, by Jensen's inequality, an economy with an unequal distribution of sales across sectors implies a lower average value of ψ, and thus a flatter aggregate Phillips curve where monetary policy has larger real effects, than an economy with just one sector, but the same average sales frequency. The second finding makes this intuition precise when marginal cost is constant.[27]

Table 3: *Calibration of the two-sector model*

Description	Notation	Value
Stylized facts		
Frequency of sales in the sale sector	s	0.29*
Aggregated frequency of sales	$\bar{s}$	0.074†
Parameters		
Fraction of loyal customers for each brand in the sale sector	λ	0.735
Size of the sale sector	σ	0.255

Notes: The stylized facts for μ and χ are as in Table 1 along with the matching parameters values for ϵ and η.
* *Source*: Eichenbaum, Jaimovich and Rebelo (2008)
† *Source*: Nakamura and Steinsson (2008)

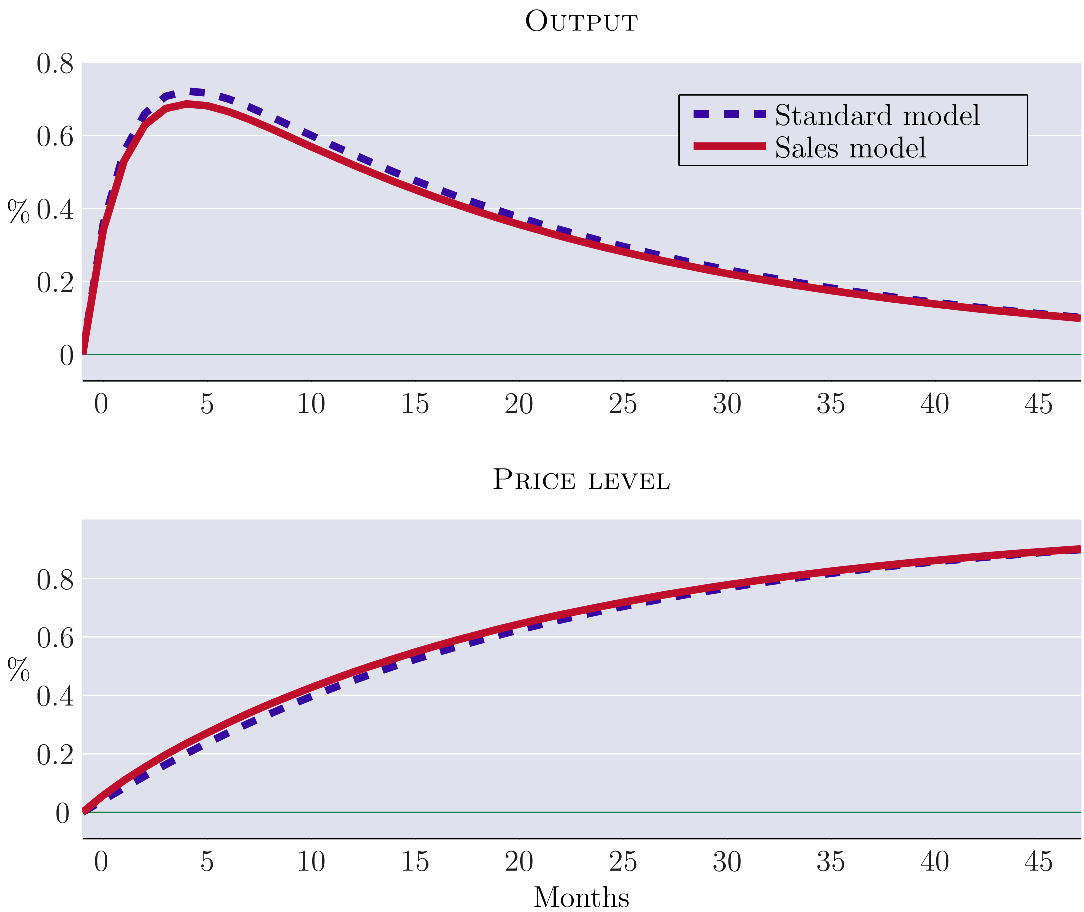
The two-sector model is now calibrated to establish the magnitude of the effect of sectoral heterogeneity on the earlier findings. The only change to the earlier calibration is that the sale frequency in the sale sector s is targeted in addition to the average sale frequency $\bar{s}$ for the whole economy. The two targets are matched by adjusting λ and σ appropriately.

Eichenbaum, Jaimovich and Rebelo (2008) study data from a major U.S. retailer and find that prices are below their "reference" level 29% of the time on average. Hence, the target value for s is 0.29, which yields σ, the size of the sale sector, equal to 0.255 when the economy-wide sale frequency must be the same as the one-sector calibration (Table 1). The calibration exercise is summarized in Table 3.

Figure 8 shows the impulse responses to the same monetary policy experiment described in section 4 for the two-sector model with sales and the standard model without any sales. The difference between the impulse responses is even smaller than before. The ratio of the cumulated

[27] The equations of the two-sector model in the general case $\gamma \neq 0$ are presented in appendix A.7. The analysis can easily be extended to an n-sector model.

Figure 8: *Impulse responses to a persistent shock to money growth in the two-sector model*



Notes: The model is as described in section 5. Parameter values are from Table 2 and Table 3.

responses of output is now 0.96, in contrast to 0.89 in the one-sector model. This shows that sales are essentially irrelevant for monetary policy analysis in the two-sector model.

6 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 cuts 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 proposed in this paper is used to understand the implications for monetary policy analysis of flexibility in sales alongside stickiness in normal prices. Explaining the occurrence of sales in a framework based on consumer heterogeneity entails strategic substitutability of sales decisions. But it is exactly because sales are strategic substitutes that they barely react to aggregate shocks, including monetary policy shocks. This is in spite of firms having a direct incentive to adjust sales when their normal prices are sticky. Firms would adjust sales in response to idiosyncratic shocks: only aggregate shocks lead to a tension between adjustment through sales and strategic considerations.

The findings of this paper indicate that in a world with both sticky normal prices and flexible sales, it is 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

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, 14

Calvo, G. A. (1983), "Staggered prices in a utility-maximizing framework", *Journal of Monetary Economics*, **12(3)**:383–398 (September). 17, 23

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). 22, 23

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, 29

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). 20, 21, 58

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

Guimaraes, B. and Sheedy, K. D. (2008), "Sales and monetary policy", Discussion paper 6940, Centre for Economic Policy Research. 20, 25

Hall, R. E. (2009), "Reconciling cyclical movements in the marginal value of time and the marginal product of labor", *Journal of Political Economy*, **117(2)**:281–323 (April). 22, 23

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. (2008), "Temporary price changes and the real effects of monetary policy", Working paper 14392, National Bureau of Economic Research. 1, 3, 13

Kimball, M. S. (1995), "The quantitative analytics of the basic Neomonetarist model", *Journal of Money, Credit and Banking*, **27(4 part 2)**:1241–1277 (November). 9

Klenow, P. J. and Kryvtsov, O. (2008), "State-dependent or time-dependent pricing: Does it matter for recent U.S. inflation?", *Quarterly Journal of Economics*, **123(3)**:863–904. 1

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

Levin, A. and Yun, T. (2009), "Reconsidering the microeconomic foundations of price-setting behavior", Working paper, Federal Reserve Board. 22

Nakamura, E. (2008), "Pass-through in retail and wholesale", *American Economic Review*, **98(2)**:430–437 (May). 13

Nakamura, E. and Steinsson, J. (2008), "Five facts about prices: A reevaluation of menu cost models", *Quarterly Journal of Economics*, **123(4)**:1415–1464 (November). 1, 14, 21, 22, 23, 27, 29

——— (2009), "Price setting in forward-looking customer markets", Working paper, Columbia University. 13

Narasimhan, C. (1988), "Competitive promotional strategies", *Journal of Business*, **61(4)**:427–449 (October). 2, 13

Narasimhan, C., Neslin, S. and Sen, S. (1996), "Promotional elasticities and category characteristics", *Journal of Marketing*, **60(2)**:17–30 (April). 22

Neslin, S. A. (2002), "Sales promotion", in *Handbook of Marketing*, SAGE. 13

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

Thaler, R. (1985), "Mental accounting and consumer choice", *Marketing Science*, **4(3)**:199–214 (Summer). 13

VARIAN, H. (1980), "A model of sales", *American Economic Review*, **70(4)**:651–659 (September). 2, 13

WOODFORD, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, New Jersey. 20

A Technical appendix

A.1 Solving the model

Steady state

Finding the steady state of the model characterized in section 3.4 requires solving only one equation numerically. For parameters ϵ and η satisfying condition [3.3], the markup ratio μ is a root of the equation $\mathfrak{R}(\mu;\epsilon,\eta)=0$, where $\mathfrak{R}(\mu;\epsilon,\eta)$ is the determinant

$$\mathfrak{R}(\mu;\epsilon,\eta)\equiv\begin{vmatrix}\mathfrak{a}_0(\mu;\epsilon,\eta) & \mathfrak{a}_1(\mu;\epsilon,\eta) & \mathfrak{a}_2(\eta) & 0 & 0\\ 0 & \mathfrak{a}_0(\mu;\epsilon,\eta) & \mathfrak{a}_1(\mu;\epsilon,\eta) & \mathfrak{a}_2(\eta) & 0\\ 0 & 0 & \mathfrak{a}_0(\mu;\epsilon,\eta) & \mathfrak{a}_1(\mu;\epsilon,\eta) & \mathfrak{a}_2(\eta)\\ \mathfrak{b}_0(\mu;\epsilon,\eta) & \mathfrak{b}_1(\mu;\epsilon,\eta) & \mathfrak{b}_2(\mu;\epsilon,\eta) & \mathfrak{b}_3(\eta) & 0\\ 0 & \mathfrak{b}_0(\mu;\epsilon,\eta) & \mathfrak{b}_1(\mu;\epsilon,\eta) & \mathfrak{b}_2(\mu;\epsilon,\eta) & \mathfrak{b}_3(\eta)\end{vmatrix}, \quad [A.1.1]$$

and where the functions in the matrix are given by:

$$\mathfrak{a}_0(\mu;\epsilon,\eta)\equiv\epsilon(\epsilon-1)\mu^{\eta-\epsilon}; \quad [A.1.2a]$$

$$\mathfrak{a}_1(\mu;\epsilon,\eta)\equiv\eta(\epsilon-1)\left(\frac{1-\mu^{\eta-\epsilon+1}}{1-\mu}\right)+\epsilon(\eta-1)\left(\frac{\mu^{\eta-\epsilon}-\mu}{1-\mu}\right); \quad [A.1.2b]$$

$$\mathfrak{a}_2(\eta)\equiv\eta(\eta-1); \quad [A.1.2c]$$

$$\mathfrak{b}_0(\mu;\epsilon,\eta)\equiv(\epsilon-1)\left(\frac{\mu^{2(\eta-\epsilon)}-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right); \quad [A.1.2d]$$

$$\mathfrak{b}_1(\mu;\epsilon,\eta)\equiv(\eta-1)\left(\frac{\mu^{2(\eta-\epsilon)}-\mu^{\eta}}{1-\mu^{\eta}}\right)+2(\epsilon-1)\left(\frac{\mu^{\eta-\epsilon}-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right); \quad [A.1.2e]$$

$$\mathfrak{b}_2(\mu;\epsilon,\eta)\equiv(\epsilon-1)\left(\frac{1-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right)+2(\eta-1)\left(\frac{\mu^{\eta-\epsilon}-\mu^{\eta}}{1-\mu^{\eta}}\right); \quad [A.1.2f]$$

$$\mathfrak{b}_3(\eta)\equiv(\eta-1). \quad [A.1.2g]$$

When searching for a root, it is necessary to restrict attention to economically meaningful solutions. These correspond to positive real values of the function

$$\mathfrak{z}(\mu;\epsilon,\eta)\equiv\frac{-\mathfrak{a}_1(\mu;\epsilon,\eta)-\sqrt{\mathfrak{a}_1(\mu;\epsilon,\eta)^2-4\mathfrak{a}_2(\eta)\mathfrak{a}_0(\mu;\epsilon,\eta)}}{2\mathfrak{a}_2(\eta)}. \quad [A.1.3]$$

Under the conditions stated in Theorem 1, there exists a unique economically meaningful solution of the equation $\mathfrak{R}(\mu;\epsilon,\eta)=0$.

Having obtained the markup ratio μ, the quantity ratio χ is

$$\chi=\mu^{-\epsilon}\left(\frac{1+\mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu;\epsilon,\eta)}{1+\mathfrak{z}(\mu;\epsilon,\eta)}\right), \quad [A.1.4]$$

and the sales frequency s is

$$s = \frac{\left(\frac{\lambda}{1-\lambda}\mathfrak{z}(\mu;\epsilon,\eta)\right)^{-\left(\frac{\eta-1}{\eta-\epsilon}\right)} - 1}{\mu^{-(\eta-1)} - 1}. \tag{A.1.5}$$

This expression for the sales frequency is economically meaningful when λ lies between the bounds $\underline{\lambda}(\epsilon,\eta)$ and $\overline{\lambda}(\epsilon,\eta)$ referred to in Theorem 1, which are given by:

$$\underline{\lambda}(\epsilon,\eta) \equiv \frac{1}{1+\mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu;\epsilon,\eta)}, \quad \text{and} \quad \overline{\lambda}(\epsilon,\eta) \equiv \frac{1}{1+\mathfrak{z}(\mu;\epsilon,\eta)}. \tag{A.1.6}$$

An expression for real marginal cost x (the reciprocal of the average markup) is

$$x = \left(\lambda\left((1+\mathfrak{z}(\mu;\epsilon,\eta)) + \left((\mu^{1-\epsilon}-1) + (\mu^{1-\eta}-1)\mathfrak{z}(\mu;\epsilon,\eta)\right)s\right)\right)^{\frac{1}{\epsilon-1}}\left(\frac{(\epsilon-1)+(\eta-1)\mathfrak{z}(\mu;\epsilon,\eta)}{\epsilon+\eta\mathfrak{z}(\mu;\epsilon,\eta)}\right), \tag{A.1.7}$$

and the degree of price distortion $\mathit{\Delta} = Y/Q$ is given by:

$$\mathit{\Delta} = \frac{\left(\lambda\left((1+\mathfrak{z}(\mu;\epsilon,\eta)) + \left((\mu^{1-\epsilon}-1)+(\mu^{1-\eta}-1)\mathfrak{z}(\mu;\epsilon,\eta)\right)s\right)\right)^{\frac{\epsilon}{\epsilon-1}}}{\lambda\left((1+\mathfrak{z}(\mu;\epsilon,\eta)) + ((\mu^{-\epsilon}-1)+(\mu^{-\eta}-1)\mathfrak{z}(\mu;\epsilon,\eta))s\right)}. \tag{A.1.8}$$

DSGE model

The system of log-linearized equations of the model from section 4 is

$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \frac{1}{1-\psi}\left(\kappa \mathsf{x}_t + \psi\left(\mathbf{\Delta}\mathsf{x}_t - \beta\mathbb{E}_t\mathbf{\Delta}\mathsf{x}_{t+1}\right)\right); \tag{A.1.9a}$$

$$\mathsf{x}_t = \frac{1}{1+\gamma\delta}\mathsf{w}_t + \frac{\gamma}{1+\gamma\delta}\mathsf{Y}_t; \tag{A.1.9b}$$

$$\pi_{W,t} = \beta\mathbb{E}_t\pi_{W,t+1} + \frac{(1-\phi_w)(1-\beta\phi_w)}{\phi_w}\frac{1}{1+\varsigma\theta_h^{-1}}\left(\left(\theta_c^{-1} + \frac{1}{1+\gamma\delta}\frac{\theta_h^{-1}}{\alpha}\right)\mathsf{Y}_t - \left(1+\frac{\delta}{1+\gamma\delta}\frac{\theta_h^{-1}}{\alpha}\right)\mathsf{w}_t\right); \tag{A.1.9c}$$

$$\mathbf{\Delta}\mathsf{w}_t = \pi_{W,t} - \pi_t; \tag{A.1.9d}$$

$$\mathsf{Y}_t = \mathbb{E}_t\mathsf{Y}_{t+1} - \theta_c\left(\mathsf{i}_t - \mathbb{E}_t\pi_{t+1}\right); \tag{A.1.9e}$$

$$\mathbf{\Delta}\mathsf{Y}_t = \mathbf{\Delta}\mathsf{M}_t - \pi_t; \tag{A.1.9f}$$

$$\mathbf{\Delta}\mathsf{M}_t = \mathfrak{p}\mathbf{\Delta}\mathsf{M}_{t-1} + (1-\mathfrak{p})\mathsf{e}_t. \tag{A.1.9g}$$

The Phillips curve equation is from Theorem 2 and derivations of the other equations are given in appendix A.6. All the coefficients apart from ψ, κ and δ are as defined in Table 2. Formulæ for ψ, κ and δ are

$$\psi = 1 - \frac{(1-s)\left(1-\left(\frac{\eta-1}{\epsilon-1}\right)\left(\frac{\mu^{1-\epsilon}-1}{\mu^{1-\eta}-1}\right)\right)}{(1+\mathfrak{z}(\mu;\epsilon,\eta)) + ((\mu^{1-\epsilon}-1)+(\mu^{1-\eta}-1)\mathfrak{z}(\mu;\epsilon,\eta))s}, \quad \kappa = \frac{(1-\phi_p)(1-\beta\phi_p)}{\phi_p}, \quad \text{and} \tag{A.1.10a}$$

$$\delta = \frac{s\chi\epsilon(1-\mu)}{s\chi+(1-s)} \tag{A.1.10b}$$

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

The solution of the system [A.1.9] can be obtained using standard methods for solving expectational difference equations.

The standard model without sales is a special case of [A.1.9] with the following parameter restrictions:

$$\psi = 0, \quad \delta = 0, \quad \xi = \frac{1}{1-x}, \quad \text{and} \quad \kappa = \frac{1}{1+\xi\gamma}\frac{(1-\phi_p)(1-\beta\phi_p)}{\phi_p},$$

where the Phillips curve then reduces to the standard New Keynesian Phillips curve.

A.2 Properties of the demand, revenue and marginal revenue functions

The structure of household consumption preferences introduced in section 2.2 implies that firms face a demand curve $q = \mathscr{D}(p; P_B, \mathcal{E})$ of the form given in equation [2.10] 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

$$\mathcal{D}(\rho) \equiv \lambda\rho^{-\epsilon} + (1-\lambda)\rho^{-\eta}, \tag{A.2.1}$$

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 [2.7], 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 [2.10]:

$$\rho \equiv \frac{p}{P_B}, \quad \text{and} \quad \mathfrak{q} \equiv \frac{P_B^{\epsilon}}{\mathcal{E}}q. \tag{A.2.2}$$

With these definitions, the original demand function [2.10] is stated in terms of the relative demand function [A.2.1] as follows:

$$\mathscr{D}(p; P_B, \mathcal{E}) = \frac{\mathcal{E}}{P_B^{\epsilon}}\mathcal{D}\left(\frac{p}{P_B}\right). \tag{A.2.3}$$

The relative demand function [A.2.1] 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 revenue function $\mathcal{R}(\mathfrak{q})$, defined in terms of the relative demand function, is

$$\mathcal{R}(\mathfrak{q}) \equiv \mathfrak{q}\mathcal{D}^{-1}(\mathfrak{q}). \tag{A.2.4}$$

Using the inverse demand function $\rho = \mathcal{D}^{-1}(\mathfrak{q})$, an equivalent expression for the revenue function 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.2.1]:

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

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 revenue function $\mathcal{R}(\mathfrak{q})$ from [A.2.4] using the inverse function theorem, and substituting demand function [A.2.1] yields an expression for marginal revenue:

$$\mathcal{R}'\left(\mathcal{D}(\rho)\right) = \left(\frac{(\epsilon-1)\lambda + (\eta-1)(1-\lambda)\rho^{\epsilon-\eta}}{\epsilon\lambda + \eta(1-\lambda)\rho^{\epsilon-\eta}}\right)\rho. \tag{A.2.5}$$

Because $\epsilon > 1$ and $\eta > 1$, it follows that $\mathcal{R}'(\mathfrak{q}) > 0$ for all $\mathfrak{q}$, so 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.2.5] 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.2.3] establishes the original demand function $\mathscr{D}(p; P_B, \mathcal{E})$ in [2.10] is connected to the relative demand function $\mathcal{D}(\rho)$ in [A.2.1], there are similar relations between the original inverse demand function

$\mathscr{D}^{-1}(\mathfrak{q}; P_B, \mathcal{E})$, original 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.2.3]:

$$\mathscr{D}^{-1}(q; P_B, \mathcal{E}) = P_B \mathcal{D}^{-1}\left(\frac{qP_B^{\epsilon}}{\mathcal{E}}\right). \qquad \text{[A.2.6]}$$

Equations [2.11], [A.2.4] and [A.2.6] justify the following connections between the revenue functions and their derivatives:

$$\mathscr{R}(q; P_B, \mathcal{E}) = P_B^{1-\epsilon}\mathcal{E}\mathcal{R}\left(\frac{qP_B^{\epsilon}}{\mathcal{E}}\right), \quad \mathscr{R}'(q; P_B, \mathcal{E}) = P_B\mathcal{R}'\left(\frac{qP_B^{\epsilon}}{\mathcal{E}}\right), \quad \text{and } \mathscr{R}''(q; P_B, \mathcal{E}) = \frac{P_B^{1+\epsilon}}{\mathcal{E}}\mathcal{R}''\left(\frac{qP_B^{\epsilon}}{\mathcal{E}}\right). \qquad \text{[A.2.7]}$$

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.2.4] using the relative demand function [A.2.1], and suppose that $\eta > \epsilon > 1$.*

(i) If $\lambda = 0$ or $\lambda = 1$ or condition [3.3] 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 [3.3] 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.2.5] 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.2.5] to obtain

$$\mathcal{D}'(\rho)\mathcal{R}''(\mathcal{D}(\rho)) = \frac{\eta(\eta-1)\left(\frac{1-\lambda}{\lambda}\rho^{\epsilon-\eta}\right)^2 - \left((\eta-\epsilon)^2 - \eta(\epsilon-1) - \epsilon(\eta-1)\right)\left(\frac{(1-\lambda)}{\lambda}\rho^{\epsilon-\eta}\right) + \epsilon(\epsilon-1)}{\left(\epsilon + \eta\left(\frac{1-\lambda}{\lambda}\rho^{\epsilon-\eta}\right)\right)^2}, \qquad \text{[A.2.8]}$$

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 $\mathcal{D}^{-1}(\mathfrak{q})$:

$$\mathcal{Z}(\mathfrak{q}) \equiv \frac{1-\lambda}{\lambda}\left(\mathcal{D}^{-1}(\mathfrak{q})\right)^{\epsilon-\eta}, \qquad \text{[A.2.9]}$$

and use this together with [A.2.8] to write the derivative of marginal revenue as follows:

$$\mathcal{R}''(\mathfrak{q}) = \frac{\eta(\eta-1)\left(\mathcal{Z}(\mathfrak{q})\right)^2 - \left((\eta-\epsilon)^2 - \eta(\epsilon-1) - \epsilon(\eta-1)\right)\mathcal{Z}(\mathfrak{q}) + \epsilon(\epsilon-1)}{\mathcal{D}'\left(\mathcal{D}^{-1}(\mathfrak{q})\right)\left(\epsilon + \eta\mathcal{Z}(\mathfrak{q})\right)^2}. \qquad \text{[A.2.10]}$$

Since $\mathcal{D}'\left(\mathcal{D}^{-1}(\mathfrak{q})\right) < 0$ for all $\mathfrak{q}$, and the remaining term in the denominator of [A.2.10] is strictly positive, the sign of $\mathcal{R}''(\mathfrak{q})$ is the opposite of that of the quadratic function

$$\mathscr{Q}(z) \equiv \eta(\eta-1)z^2 - \left((\eta-\epsilon)^2 - \eta(\epsilon-1) - \epsilon(\eta-1)\right)z + \epsilon(\epsilon-1), \qquad \text{[A.2.11]}$$

evaluated at $z = \mathcal{Z}(\mathfrak{q})$. The aim is to find a region where marginal revenue is upward sloping, which corresponds to $\mathscr{Q}(z)$ 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 $\mathscr{Q}(z) = 0$ 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 $\mathscr{Q}(0) = \epsilon(\epsilon-1) > 0$ it then follows that $\mathscr{Q}(z) > 0$ 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):

$$\mathcal{G}_p(\eta;\epsilon) \equiv \eta^2 - (4\epsilon-1)\eta + \epsilon(\epsilon+1), \quad \text{and} \quad \mathcal{G}_r(\eta;\epsilon) \equiv \eta^2 - 2(3\epsilon-1)\eta + (\epsilon+1)^2. \qquad \text{[A.2.12]}$$

By comparing $\mathcal{G}_p(\eta;\epsilon)$ to the coefficient of z in [A.2.11], the sum of the roots $\mathscr{Q}(z)=0$ is positive if and only if $\mathcal{G}_p(\eta;\epsilon)>0$ since $\eta>1$. Then the discriminant of the quadratic $\mathscr{Q}(z)$ is factored in terms of $\mathcal{G}_r(\eta;\epsilon)$ as follows:

$$\left((\eta-\epsilon)^2 - \eta(\epsilon-1) - \epsilon(\eta-1)\right)^2 - 4\epsilon\eta(\epsilon-1)(\eta-1) = (\eta-\epsilon)^2\mathcal{G}_r(\eta;\epsilon), \qquad \text{[A.2.13]}$$

and as $\eta\neq\epsilon$, the equation $\mathscr{Q}(z)=0$ has two distinct real roots if and only if $\mathcal{G}_r(\eta;\epsilon)>0$. To summarize, the quadratic $\mathscr{Q}(z)$ 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.2.12] 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

$$\hat{\eta}(\epsilon) - \epsilon = \frac{2\epsilon - (2\epsilon^2-1)}{2\epsilon-1}. \qquad \text{[A.2.14]}$$

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.2.12], 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.2.14] 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 $\mathscr{Q}(z)=0$ from [A.2.11] has two distinct positive real roots $\underline{z}$ and $\overline{z}$ with $\underline{z}<\overline{z}$. $\mathscr{Q}(z)<0$ 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 $\mathscr{Q}(z)>0$. If $\eta\leq\eta^*(\epsilon)$ then $\mathscr{Q}(z)>0$ 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.2.10] and [A.2.11] that $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing for all $\mathfrak{q}\geq0$.

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.2.9] is strictly increasing. Its inverse is

$$\mathcal{Z}^{-1}(z) = \mathcal{D}\left(\left(\frac{\lambda}{1-\lambda}z\right)^{\frac{1}{\epsilon-\eta}}\right), \tag{A.2.15}$$

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.2.10] and [A.2.11] 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 [3.3], 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 revenue function $\mathcal{R}(\mathfrak{q})$ defined in [A.2.4], suppose that $0 < \lambda < 1$, and ϵ and η are such that non-monotonicity condition [3.3] 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

$$\mathcal{R}'(\mathfrak{q}_S) = \mathcal{R}'(\mathfrak{q}_N) = \frac{\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}(\mathfrak{q}_N)}{\mathfrak{q}_S - \mathfrak{q}_N}. \tag{A.2.16}$$

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

$$\mathcal{R}''(\mathfrak{q}_S) < 0, \quad \mathcal{R}''(\mathfrak{q}_N) < 0, \quad \mathcal{R}(\mathfrak{q}_S)/\mathfrak{q}_S > \mathcal{R}'(\mathfrak{q}_S), \quad \textit{and } \mathcal{R}(\mathfrak{q}_N)/\mathfrak{q}_N > \mathcal{R}'(\mathfrak{q}_N). \tag{A.2.17}$$

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

$$\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_S) + \mathcal{R}'(\mathfrak{q}_S)(\mathfrak{q} - \mathfrak{q}_S) = \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\mathfrak{q} - \mathfrak{q}_N). \tag{A.2.18}$$

Proof (i) When $0 < \lambda < 1$ and condition [3.3] 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:

$$\mathfrak{q}_1(\varkappa) \equiv (1-\varkappa)\underline{\mathfrak{q}}_1 + \varkappa\overline{\mathfrak{q}}_1, \tag{A.2.19}$$

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 $\varkappa \in [0,1]$:

$$\mathcal{R}'(\mathfrak{q}_1(\varkappa)) \equiv \mathcal{R}'(\mathfrak{q}_2(\varkappa)) \equiv \mathcal{R}'(\mathfrak{q}_3(\varkappa)). \tag{A.2.20}$$

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

$$\mathfrak{q}_2(0) = \mathfrak{q}_3(0) = \bar{\mathfrak{q}}, \quad \text{and} \;\; \mathfrak{q}_1(1) = \mathfrak{q}_2(1) = \underline{\mathfrak{q}}. \tag{A.2.21}$$

Continuity and differentiability of $\mathcal{R}'(\mathfrak{q})$ and of $\mathfrak{q}_1(\varkappa)$ from [A.2.19] 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.2.20] it follows that

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

As Lemma 1 establishes $\mathcal{R}(\mathfrak{q})$ is concave on $[\underline{\mathfrak{q}}_1, \bar{\mathfrak{q}}_1]$ and $[\underline{\mathfrak{q}}_3, \bar{\mathfrak{q}}_3]$, and convex on $[\underline{\mathfrak{q}}_2, \bar{\mathfrak{q}}_2]$, 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 $F(\varkappa)$ in terms of the following integrals:

$$F(\varkappa) \equiv \int_{\mathfrak{q}_2(\varkappa)}^{\mathfrak{q}_3(\varkappa)} \big(\mathcal{R}'(\mathfrak{q}) - \mathcal{R}'(\mathfrak{q}_2(\varkappa))\big)\, d\mathfrak{q} - \int_{\mathfrak{q}_1(\varkappa)}^{\mathfrak{q}_2(\varkappa)} \big(\mathcal{R}'(\mathfrak{q}_2(\varkappa)) - \mathcal{R}'(\mathfrak{q})\big)\, d\mathfrak{q}. \tag{A.2.22}$$

From continuity and differentiability of $\mathfrak{q}_1(\varkappa)$, $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$, it follows that $F(\varkappa)$ is also continuous for all $\varkappa \in [0,1]$ and differentiable for all $\varkappa \in (0,1)$. Evaluating $F(\varkappa)$ at the endpoints of the interval $[0,1]$ and making use of [A.2.21] yields

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

where the first inequality follows because $\mathcal{R}'(\mathfrak{q}) < \overline{\mathcal{R}'}$ for all $\underline{\mathfrak{q}}_1 < \mathfrak{q} < \bar{\mathfrak{q}}_2$, and the second because $\mathcal{R}'(\mathfrak{q}) > \underline{\mathcal{R}'}$ for all $\underline{\mathfrak{q}}_2 < \mathfrak{q} < \bar{\mathfrak{q}}_3$. Differentiating $F(\varkappa)$ in [A.2.22] using Leibniz's rule and substituting the definitions from [A.2.20] leads to the following result:

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

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 $F(0) < 0$, $F(1) > 0$, and $F(\varkappa)$ is continuous and strictly increasing in $\varkappa$, there exists a unique $\varkappa^* \in (0,1)$ such that $F(\varkappa^*) = 0$.

The solution of the system of equations [A.2.16] 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 $F(\varkappa) = 0$ obtained above. From [A.2.20], it follows immediately that $\mathcal{R}'(\mathfrak{q}_N) = \mathcal{R}'(\mathfrak{q}_S)$, establishing the first equality in [A.2.16]. Since $F(\varkappa^*) = 0$, the definition of $F(\varkappa)$ in equation [A.2.22] implies

$$\int_{\mathfrak{q}_2(\varkappa^*)}^{\mathfrak{q}_S} \big(\mathcal{R}'(\mathfrak{q}) - \mathcal{R}'(\mathfrak{q}_2(\varkappa^*))\big)\, d\mathfrak{q} = \int_{\mathfrak{q}_N}^{\mathfrak{q}_2(\varkappa^*)} \big(\mathcal{R}'(\mathfrak{q}_2(\varkappa^*)) - \mathcal{R}'(\mathfrak{q})\big)\, d\mathfrak{q}, \tag{A.2.23}$$

which is rearranged to deduce

$$\int_{\mathfrak{q}_N}^{\mathfrak{q}_S} \mathcal{R}'(\mathfrak{q}) d\mathfrak{q} = (\mathfrak{q}_S - \mathfrak{q}_N)\mathcal{R}'(\mathfrak{q}_2(\varkappa^*)). \tag{A.2.24}$$

Equation [A.2.20] implies $\mathcal{R}'(\mathfrak{q}_2(\varkappa^*)) = \mathcal{R}'(\mathfrak{q}_N) = \mathcal{R}'(\mathfrak{q}_S)$, which together with the above establishes that

$$\mathcal{R}'(\mathfrak{q}_S) = \mathcal{R}'(\mathfrak{q}_N) = \frac{\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}(\mathfrak{q}_N)}{\mathfrak{q}_S - \mathfrak{q}_N}. \tag{A.2.25}$$

Thus, the values of $\mathfrak{q}_N$ and $\mathfrak{q}_S$ are indeed a solution of the system of equations in [A.2.16].

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.2.16] 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.2.20] 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.2.16]. 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.2.23] and [A.2.24] show that the second equality in [A.2.16] 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.2.20] 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

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

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.2.25] 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.2.16] 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:

$$\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\mathfrak{q} - \mathfrak{q}_N) \quad \forall \mathfrak{q} \in [0, \underline{\mathfrak{q}}], \quad \text{and} \ \mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_S) + \mathcal{R}'(\mathfrak{q}_S)(\mathfrak{q} - \mathfrak{q}_S) \quad \forall \mathfrak{q} \in [\overline{\mathfrak{q}}, \infty), \tag{A.2.26}$$

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.2.16] is

$$\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}'(\mathfrak{q}_S)\mathfrak{q}_S = \mathcal{R}(\mathfrak{q}_N) - \mathcal{R}'(\mathfrak{q}_N)\mathfrak{q}_N. \tag{A.2.27}$$

By evaluating the first inequality in [A.2.26] at $\mathfrak{q} = 0$, where $\mathcal{R}(0) = 0$, and making use of the equation above it is deduced that

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

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.2.17].

(iii) By applying the inequalities in [A.2.26] 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:

$$\mathcal{R}(\underline{\mathfrak{q}}) \leq \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\underline{\mathfrak{q}} - \mathfrak{q}_N), \quad \text{and} \ \mathcal{R}(\overline{\mathfrak{q}}) \leq \mathcal{R}(q_N) + \mathcal{R}'(\mathfrak{q}_N)(\overline{\mathfrak{q}} - \mathfrak{q}_N). \tag{A.2.28}$$

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:

$$\mathcal{R}(\mathfrak{q}) \equiv \mathcal{R}\left(\left(\frac{\overline{\mathfrak{q}} - \mathfrak{q}}{\overline{\mathfrak{q}} - \underline{\mathfrak{q}}}\right)\underline{\mathfrak{q}} + \left(\frac{\mathfrak{q} - \underline{\mathfrak{q}}}{\overline{\mathfrak{q}} - \underline{\mathfrak{q}}}\right)\overline{\mathfrak{q}}\right) \leq \left(\frac{\overline{\mathfrak{q}} - \mathfrak{q}}{\overline{\mathfrak{q}} - \underline{\mathfrak{q}}}\right)\mathcal{R}(\underline{\mathfrak{q}}) + \left(\frac{\mathfrak{q} - \underline{\mathfrak{q}}}{\overline{\mathfrak{q}} - \underline{\mathfrak{q}}}\right)\mathcal{R}(\overline{\mathfrak{q}}), \tag{A.2.29}$$

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.2.28] using as weights the coefficients from [A.2.29] yields $\mathcal{R}(\mathfrak{q}) \leq$

$\mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\mathfrak{q} - \mathfrak{q}_N)$ for all $\mathfrak{q} \in (\underline{\mathfrak{q}}, \bar{\mathfrak{q}})$. This finding, together with the inequalities in [A.2.26] and the equations [A.2.25] and [A.2.27], 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.2.18] is established, which completes the proof. ■

The existence and uniqueness of the solution of equations [A.2.16] has been demonstrated given condition [3.3] 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.2.16] (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.2.1]. In addition, define the markup ratio $\mu \equiv \mu_S/\mu_N = \rho_S/\rho_N$ and the quantity ratio $\chi \equiv \mathfrak{q}_S/\mathfrak{q}_N$.*

(i) The markup ratio $\mu \equiv \rho_S/\rho_N$ is the only solution of the equation $\mathfrak{R}(\mu; \epsilon, \eta) = 0$ from [A.1.1] with $0 < \mu < 1$ and where $\mathfrak{z}(\mu; \epsilon, \eta)$ in [A.1.3] 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 equal to the expression in equation [A.1.4]. Hence χ depends only on parameters ϵ and η.

(iii) The equilibrium markups μ_S and μ_N from [3.5] depend only on ϵ and η and are given by

$$\mu_S = \frac{\epsilon + \eta\mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu; \epsilon, \eta)}{(\epsilon - 1) + (\eta - 1)\mu^{-(\eta-\epsilon)}\mathfrak{z}(\mu; \epsilon, \eta)}, \quad \textit{and} \;\; \mu_N = \frac{\epsilon + \eta\mathfrak{z}(\mu; \epsilon, \eta)}{(\epsilon - 1) + (\eta - 1)\mathfrak{z}(\mu; \epsilon, \eta)}, \qquad \text{[A.2.30]}$$

where the function $\mathfrak{z}(\mu; \epsilon, \eta)$ is given in [A.1.3].

(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:

$$\rho_N = \left(\frac{\lambda}{1-\lambda}\mathfrak{z}(\mu; \epsilon, \eta)\right)^{-\frac{1}{\eta-\epsilon}}, \quad \textit{and} \;\; \rho_S = \left(\frac{\lambda}{1-\lambda}\mathfrak{z}(\mu; \epsilon, \eta)\right)^{-\frac{1}{\eta-\epsilon}}\mu, \qquad \text{[A.2.31]}$$

where $\mathfrak{q}_N = \mathcal{D}(\rho_N)$ and $\mathfrak{q}_S = \mathcal{D}(\rho_S)$ using the relative demand function $\mathcal{D}(\rho)$ from [A.2.1].

Proof (i) Using the expression for marginal revenue from [A.2.5], the first equality in [A.2.16] 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

$$\mu = \left(\frac{\epsilon + \eta\mu^{-(\eta-\epsilon)}z}{\epsilon + \eta z}\right)\left(\frac{(\epsilon - 1) + (\eta - 1)z}{(\epsilon - 1) + (\eta - 1)\mu^{-(\eta-\epsilon)}z}\right). \qquad \text{[A.2.32]}$$

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.2.32] 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:

$$\eta(\eta-1)z^2 + \left(\eta(\epsilon-1)\left(\frac{1-\mu^{\eta-\epsilon+1}}{1-\mu}\right) + \epsilon(\eta-1)\left(\frac{\mu^{\eta-\epsilon}-\mu}{1-\mu}\right)\right)z + \epsilon(\epsilon-1)\mu^{\eta-\epsilon} = 0. \qquad \text{[A.2.33]}$$

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.1.2] are obtained directly from [A.2.33].

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.2.16]. The definition of the revenue function $\mathcal{R}(\mathfrak{q})$ in [A.2.4] 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

$$\mathfrak{q}_S\left(\rho_S - \mathcal{R}'(\mathfrak{q}_S)\right) = \mathfrak{q}_N\left(\rho_N - \mathcal{R}'(\mathfrak{q}_N)\right). \qquad \text{[A.2.34]}$$

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.2.16]), equation [A.2.34] becomes

$$\chi = \left(\frac{\rho_N}{\mathcal{R}'(\mathcal{D}(\rho_N))} - 1\right) \Big/ \left(\frac{\rho_S}{\mathcal{R}'(\mathcal{D}(\rho_S))} - 1\right). \qquad \text{[A.2.35]}$$

The formula for marginal revenue $\mathcal{R}'(\mathcal{D}(\rho))$ in [A.2.5] 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.2.35] 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

$$\chi = \left(\frac{1+z}{1+\mu^{-(\eta-\epsilon)}z}\right)\left(\frac{(\epsilon-1) + (\eta-1)\mu^{-(\eta-\epsilon)}z}{(\epsilon-1) + (\eta-1)z}\right). \qquad \text{[A.2.36]}$$

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.2.1], 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

$$\chi = \mu^{-\epsilon}\left(\frac{1+\mu^{-(\eta-\epsilon)}z}{1+z}\right). \qquad \text{[A.2.37]}$$

Putting together the two expressions for the quantity ratio χ in [A.2.36] and [A.2.37], μ and z must satisfy the equation

$$\left(\frac{1+z}{1+\mu^{-(\eta-\epsilon)}z}\right)\left(\frac{(\epsilon-1) + (\eta-1)\mu^{-(\eta-\epsilon)}z}{(\epsilon-1) + (\eta-1)z}\right) = \mu^{-\epsilon}\left(\frac{1+\mu^{-(\eta-\epsilon)}z}{1+z}\right). \qquad \text{[A.2.38]}$$

Since the quantity ratio χ is finite, none of the terms in the denominators of [A.2.36] or [A.2.37] can be zero, so [A.2.38] 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:

$$
\begin{aligned}
(\eta-1)z^3 + \left((\epsilon-1)\left(\frac{1-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right)+2(\eta-1)\left(\frac{\mu^{\eta-\epsilon}-\mu^{\eta}}{1-\mu^{\eta}}\right)\right)z^2 \\
+\left((\eta-1)\left(\frac{\mu^{2(\eta-\epsilon)}-\mu^{\eta}}{1-\mu^{\eta}}\right)+2(\epsilon-1)\left(\frac{\mu^{\eta-\epsilon}-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right)\right)z \\
+(\epsilon-1)\left(\frac{\mu^{2(\eta-\epsilon)}-\mu^{2\eta-\epsilon}}{1-\mu^{\eta}}\right)=0. \quad \text{[A.2.39]}
\end{aligned}
$$

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.1.2] are obtained directly from [A.2.39].

These steps demonstrate that starting from a solution $\mathfrak{q}_S$ and $\mathfrak{q}_N$ of [A.2.16], the quadratic and the cubic equations [A.2.33] and [A.2.39] 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.1.2] 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.1.1], 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.2.16] 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 positive real numbers. 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 $\mathfrak{z}(\mu;\epsilon,\eta)$ in [A.1.3] (the smaller of the two roots of $\mathfrak{Q}(z;\mu,\epsilon,\eta)=0$) 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.1.2] as $\mu\to0$ and $\mu\to1$ and applying L'Hôpital's rule, which yields

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

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.2.16] 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.1.2] 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.1.3].

The equilibrium quantity ratio χ is obtained by substituting $z=\mathfrak{z}(\mu;\epsilon,\eta)$ into [A.2.37]. 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.2.2], the formula for the purchase multipliers in [2.10] implies $\upsilon_N = \rho_N^{\epsilon-\eta}$ and $\upsilon_S = \mu^{\epsilon-\eta}\upsilon_N$. Using the fact that $z \equiv ((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, and dividing numerator and denominator of the expression in [3.4] by λ yields [A.2.30].

(iv) The expressions for the relative prices ρ_S and ρ_N in [A.2.31] 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.3 Proof of Theorem 1

Non-monotonicity of the marginal revenue function

Using the relationship between the revenue function $\mathscr{R}(q;P_B,\mathcal{E})$ and its equivalent $\mathcal{R}(\mathfrak{q})$ defined in [A.2.4] using the relative demand function $\mathcal{D}(\rho)$ from [A.2.1], the corresponding two marginal revenue functions $\mathscr{R}'(q;P_B,\mathcal{E})$ and $\mathcal{R}'(\mathfrak{q})$ are proportional according to [A.2.7]. Lemma 1 demonstrates that $\mathcal{R}'(\mathfrak{q})$ has the described pattern of non-monotonicity under the conditions $0<\lambda<1$ and [3.3], and is otherwise a decreasing function of q.

Existence of a two-price equilibrium

For a two-price equilibrium to exist, first-order conditions [3.2] 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 [2.7], and $\mathcal{E}=P^{\epsilon}Y$ is the measure of aggregate expenditure from [2.10].

The necessary conditions for the two-price equilibrium are now restated in terms of the relative demand function $\mathcal{D}(\rho)$ defined in [A.2.1], and its associated total and marginal revenue functions $\mathcal{R}(\mathfrak{q})$ and $\mathcal{R}'(\mathfrak{q})$, as defined in [A.2.4] and analysed in appendix A.2. 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.2.2]. Using the relationships in [A.2.3] and [A.2.7], the first two optimality conditions in [3.2] 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.2.16] 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 [3.3] 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.2.16].

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 [3.7] 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 frequency s must satisfy $s\in(0,1)$.

Assume the parameters are such that ϵ and η satisfy [3.3], 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.2.16] 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.2.31] by substituting the equilibrium value of μ into the function $\mathfrak{z}(\mu; \epsilon, \eta)$ defined in [A.1.3]. 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.2.31], this condition is equivalent to:

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

Define lower and upper bounds for λ conditional on ϵ and η using the formulæ in [A.1.6] together with the equilibrium value of μ (which is a function only of ϵ and η) and the function $\mathfrak{z}(\mu; \epsilon, \eta)$ from [A.1.3]. 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 definitions of the bounds on λ, the inequality is equivalent to λ lying in the interval:

$$\underline{\lambda}(\epsilon, \eta) < \lambda < \overline{\lambda}(\epsilon, \eta). \qquad \text{[A.3.4]}$$

This restriction on λ is necessary and sufficient for the existence of an equilibrium sales frequency $s \in (0, 1)$ satisfying [A.3.2]. The equivalence is demonstrated by substituting the expressions for ρ_S and ρ_N from [A.2.31] 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:

$$s = \frac{\left(\frac{\lambda}{1-\lambda}\mathfrak{z}(\mu; \epsilon, \eta)\right)^{-\left(\frac{\eta-1}{\eta-\epsilon}\right)} - 1}{\mu^{-(\eta-1)} - 1}. \qquad \text{[A.3.5]}$$

Recalling the equivalence of inequalities [A.3.3] and [A.3.4], 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 [3.3] 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 frequency $s \in (0, 1)$ consistent with the first two equalities in [3.2]. Lemma 3 then demonstrates that the two purchase multipliers υ_S and υ_N and the two optimal markups μ_S and μ_N are determined. Equations [3.1] and [3.4] show that using the optimal markups in [3.5] is equivalent to satisfying the remaining first-order condition involving marginal cost in [3.2]. The other variables relevant to the macroeconomic equilibrium are then determined as discussed in section 3.4.

Confirming that the two-price equilibrium exists then requires checking that the remaining first-order condition [2.13c] 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.2.7] and the results of Lemma 2 in [A.2.17] the following inequalities are deduced:

$$\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}'(q_S; P_B, \mathcal{E})q_S > 0, \quad \text{and} \quad \mathscr{R}(q_N; P_B, \mathcal{E}) - \mathscr{R}'(q_N; P_B, \mathcal{E})q_N > 0. \qquad \text{[A.3.6]}$$

Since $s \in (0, 1)$, the Lagrangian multiplier $\aleph$ from first-order conditions [2.13b]–[2.13c] 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.3.6]. By combining this expression for the Lagrangian multiplier with the first-order condition [2.13c]:

$$\mathscr{R}(q; P_B, \mathcal{E}) \leq \mathscr{R}(q_N; P_B, \mathcal{E}) + \mathscr{R}'(q_N; P_B, \mathcal{E})(q - q_N) = \mathscr{R}(q_S; P_B, \mathcal{E}) + \mathscr{R}'(q_S; P_B, \mathcal{E})(q - q_S), \qquad \text{[A.3.7]}$$

which is required to hold for all $q \geq 0$. Appealing to the result of Lemma 2 in [A.2.18] and again using [A.2.7] verifies the inequality.

The assumptions about the production function [2.8] ensure that the total cost function $\mathscr{C}(Q; W)$ in

[2.9] is continuously differentiable and convex, so for all $q \geq 0$:

$$\mathscr{C}(q; W) \geq \mathscr{C}(Q; W) + \mathscr{C}'(Q; W)(q - Q), \tag{A.3.8}$$

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 $G(q) \equiv 1 - F(\mathscr{D}(p; P_B, \mathcal{E}))$ 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 [2.12]:

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

Applying the inequalities involving the revenue and total cost functions from [A.3.7] and [A.3.8] to the expression for profits yields:

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

The first-order conditions [3.2] 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 [3.3] 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.2.1] 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 the marginal revenue function 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 $(\overline{\mathfrak{q}}, \infty)$, strictly convex on $(\underline{\mathfrak{q}}, \overline{\mathfrak{q}})$, and from Lemma 2 that $\mathfrak{q}_N < \underline{\mathfrak{q}} < \overline{\mathfrak{q}} < \mathfrak{q}_S$.

Consider first the case where $\underline{\mathfrak{q}} < 1 < \overline{\mathfrak{q}}$. 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.2.2], 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.2.7] implies $\mathscr{R}''(q_1; P_B, \mathcal{E}) > 0$. Therefore, for sufficiently small $\varepsilon > 0$, the profits $\mathscr{P}$ from selling quantity $q_1 - \varepsilon$ at one half of shopping moments and $q_1 + \varepsilon$ at the other half exceed the profits from offering one price and hence one quantity at all shopping moments:

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

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 ε of moments and price p_1 at the remaining fraction $1-\varepsilon$ of moments. Profits $\mathscr{P}$ from the hybrid strategy are given by:

$$\mathscr{P} = (1-\varepsilon)\mathscr{R}(q_1; P_B, \mathcal{E}) + \varepsilon\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{C}\left((1-\varepsilon)q_1 + \varepsilon q_S; W\right). \quad \text{[A.3.9]}$$

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

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

where $\mathscr{O}\left(\varepsilon^2\right)$ denotes second- and higher-order terms in ε. 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 [2.13] 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:

$$\mathscr{P} = (\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{C}(q_1; W)) + \varepsilon(q_S - q)\left(\frac{\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}(q_1; P_B, \mathcal{E})}{q_S - q_1} - \mathscr{R}'(q_1; P_B, \mathcal{E})\right) + \mathscr{O}\left(\varepsilon^2\right). \quad \text{[A.3.10]}$$

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.2.16] can be expressed as follows:

$$\int_{\mathfrak{q}_N}^{1} \mathcal{R}'(\mathfrak{q})d\mathfrak{q} + \mathcal{R}(\mathfrak{q}_S) - \mathcal{R}(\mathfrak{q}_1) = (\mathfrak{q}_S - \mathfrak{q}_N)\mathcal{R}'(\mathfrak{q}_N). \quad \text{[A.3.11]}$$

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:

$$\int_{\mathfrak{q}_N}^{1} \mathcal{R}'(\mathfrak{q})d\mathfrak{q} < (1 - \mathfrak{q}_N)\mathcal{R}'(\mathfrak{q}_N). \quad \text{[A.3.12]}$$

Noting that $\mathcal{R}'(\mathfrak{q}_N) > \mathcal{R}'(1)$ because of $\mathfrak{q}_N < 1 < \underline{\mathfrak{q}}$, and substituting [A.3.12] into [A.3.11] and rearranging yields:

$$\frac{\mathcal{R}(\mathfrak{q}_S) - \mathcal{R}(1)}{\mathfrak{q}_S - 1} > \mathcal{R}'(\mathfrak{q}_N) > \mathcal{R}'(1), \quad \text{[A.3.13]}$$

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.2.2], and the links between the functions $\mathcal{R}(\mathfrak{q})$ and $\mathscr{R}(q; P_B, \mathcal{E})$ as set out in [A.2.7]:

$$\frac{\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}(q_1; P_B, \mathcal{E})}{q_S - q_1} > \mathscr{R}'(q_1; P_B, \mathcal{E}). \quad \text{[A.3.14]}$$

Therefore, by comparing this inequality with [A.3.10] and noting $q_S > q_1$, it follows for sufficiently small $\varepsilon > 0$ 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 ε of shopping moments and price $p_1 = P_B$ at the remaining fraction $1-\varepsilon$, 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.3.9]–[A.3.10] leads to an expression for profits $\mathscr{P}$ resulting from this hybrid strategy:

$$\mathscr{P} = (\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{C}(q_1; W)) + \varepsilon(q_1 - q_N)\left(\mathscr{R}'(q_1; P_B, \mathcal{E}) - \frac{\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{R}(q_N; P_B, \mathcal{E})}{q_1 - q_N}\right) + \mathscr{O}\left(\varepsilon^2\right). \quad \text{[A.3.15]}$$

Appealing to the properties of $\mathcal{R}(\mathfrak{q})$ for $\mathfrak{q} > \bar{\mathfrak{q}}$ and following similar steps to those in [A.3.11]–[A.3.13] 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.3.14]:

$$\mathscr{R}'(q_1; P_B, \mathcal{E}) > \frac{\mathscr{R}(q_1; P_B, \mathcal{E}) - \mathscr{R}(q_N; P_B, \mathcal{E})}{q_1 - q_N}. \qquad \text{[A.3.16]}$$

Since $q_1 > q_N$, for sufficiently small $\varepsilon > 0$, [A.3.15] and [A.3.16] 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 [3.3], Lemma 1 implies that marginal revenue $\mathscr{R}'(q; P_B, \mathcal{E})$ is strictly decreasing for all q. This is equivalent to 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 [3.3], 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.1.6] 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 $[\overline{\mathfrak{q}}, \infty)$ (containing $\mathfrak{q}_S$), which establishes that $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(1) + \mathcal{R}'(1)(\mathfrak{q} - 1)$ for all $\mathfrak{q} \in [\overline{\mathfrak{q}}, \infty)$. 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 revenue function $\mathscr{R}(q; P_B, \mathcal{E})$ using [A.2.2] and [A.2.7] yields the following for all q:

$$\mathscr{R}(q; P_B, \mathcal{E}) \leq \mathscr{R}(q_1; P_B, \mathcal{E}) + \mathscr{R}'(q_1; P_B, \mathcal{E})(q - q_1). \qquad \text{[A.3.17]}$$

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.3.17] holds in both cases. The convexity of the total cost function $\mathscr{C}(q; W)$ together with [A.3.17] 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.2.2]. 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.2.2]. Using the first-order conditions [2.13] together with [A.2.2] and [A.2.7], it follows that $\mathfrak{q}_1$ and $\mathfrak{q}_2$ must satisfy the system of equations [A.2.16] 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 1

(i) The first-order conditions are of course necessary. For sufficiency, note using the argument in the proof of Theorem 1 that the first-order conditions in [3.2] are equivalent to the equations in [A.3.1]. As Lemma 3 shows, the equations in [A.3.1] have a unique solution. Since an equilibrium is known to exist by Theorem 1, the first-order conditions must also be sufficient.

(ii) Lemma 3 shows that μ, χ, μ_S and μ_N are uniquely determined as functions of ϵ and η when the inequality [3.3] is satisfied, as is necessary for the two-price equilibrium to exist.

(iii) Lemma 3 implicitly determines the purchase multipliers v_S and v_N using the expressions for $\rho_S \equiv p_S/P_B$ and $\rho_N = p_N/P_B$ in [A.2.31] and the fact that $v_S = (p_S/P_B)^{-(\eta-\epsilon)}$ and $v_N = (p_N/P_B)^{-(\eta-\epsilon)}$ from [2.10]. Hence Lemma 3 shows that these variables depend only on ϵ, η and λ. In conjunction with equation [3.7], knowledge of ρ_S and ρ_N from [A.2.31] yields a linear equation for s after dividing both sides of [3.7] by P_B. This shows that it too only depends on ϵ, η and λ.

(iv) Substituting the bounds for λ from [A.1.6] into equation [A.3.5] proves the first two results. Differentiating [A.3.5] with respect to λ yields the third result. This completes the proof.

A.5 Proof of Theorem 2

Log linearizations

The notational convention adopted here is that a variable without a time subscript denotes its flexible-price steady-state value as determined in section 3, and the corresponding sans serif letter denotes the log deviation of the variable from its steady-state value (except for the sales frequency s, where it denotes just the deviation from steady state, and the inflation rate, where it denotes the log deviation of the gross rate).

Consider first the demand function faced by firms. The levels of demand $q_{S,\ell,t}$ and $q_{N,\ell,t}$ at the sale and normal prices are obtained from [3.9], which have the following log-linearized forms:

$$\mathsf{q}_{S,\ell,t} = \left(\frac{(1-\lambda)v_S}{\lambda + (1-\lambda)v_S}\right)\mathsf{v}_{S,\ell,t} - \epsilon(\mathsf{p}_{S,\ell,t} - \mathsf{P}_t) + \mathsf{Y}_t, \quad \text{and} \qquad \text{[A.5.1a]}$$

$$\mathsf{q}_{N,\ell,t} = \left(\frac{(1-\lambda)v_N}{\lambda + (1-\lambda)v_N}\right)\mathsf{v}_{N,\ell,t} - \epsilon(\mathsf{R}_{N,t-\ell} - \mathsf{P}_t) + \mathsf{Y}_t, \qquad \text{[A.5.1b]}$$

where the expressions are given in terms of log deviations of the purchase multipliers $v_{S,\ell,t}$ and $v_{N,\ell,t}$ from [2.10]:

$$\mathsf{v}_{S,\ell,t} = -(\eta-\epsilon)\left(\mathsf{p}_{S,\ell,t} - \mathsf{P}_{B,t}\right), \quad \text{and} \quad \mathsf{v}_{N,\ell,t} = -(\eta-\epsilon)\left(\mathsf{R}_{N,t-\ell} - \mathsf{P}_{B,t}\right). \qquad \text{[A.5.2]}$$

By substituting the purchase multipliers into the demand functions [A.5.1], the following expressions are found:

$$\mathsf{q}_{S,\ell,t} = -\left(\frac{\lambda\epsilon + (1-\lambda)\eta v_S}{\lambda + (1-\lambda)v_S}\right)\mathsf{p}_{S,\ell,t} + (\eta-\epsilon)\left(\frac{(1-\lambda)v_S}{\lambda + (1-\lambda)v_S}\right)\mathsf{P}_{B,t} + \epsilon\mathsf{P}_t + \mathsf{Y}_t, \quad \text{and} \qquad \text{[A.5.3a]}$$

$$\mathsf{q}_{N,\ell,t} = -\left(\frac{\lambda\epsilon + (1-\lambda)\eta v_N}{\lambda + (1-\lambda)v_N}\right)\mathsf{R}_{N,t-\ell} + (\eta-\epsilon)\left(\frac{(1-\lambda)v_N}{\lambda + (1-\lambda)v_N}\right)\mathsf{P}_{B,t} + \epsilon\mathsf{P}_t + \mathsf{Y}_t. \qquad \text{[A.5.3b]}$$

From equation [3.4], the log-linearized optimal markups at given sale and normal prices are:

$$\mathsf{\mu}_{S,\ell,t} = -\mathfrak{c}_S\mathsf{v}_{S,\ell,t}, \quad \text{with} \;\; \mathfrak{c}_S \equiv \frac{\lambda(1-\lambda)(\eta-\epsilon)v_S}{(\lambda\epsilon + (1-\lambda)\eta v_S)\left(\lambda(\epsilon-1) + (1-\lambda)(\eta-1)v_S\right)}, \quad \text{and} \qquad \text{[A.5.4a]}$$

$$\mathsf{\mu}_{N,\ell,t} = -\mathfrak{c}_N\mathsf{v}_{N,\ell,t}, \quad \text{with} \;\; \mathfrak{c}_N \equiv \frac{\lambda(1-\lambda)(\eta-\epsilon)v_N}{(\lambda\epsilon + (1-\lambda)\eta v_N)\left(\lambda(\epsilon-1) + (1-\lambda)(\eta-1)v_N\right)}, \qquad \text{[A.5.4b]}$$

which are given in terms of the purchase multipliers from [A.5.2]. Overall demand $Q_{\ell,t} = s_{\ell,t}q_{S,\ell,t} + (1-s_{\ell,t})q_{N,\ell,t}$ is log-linearized as follows:

$$\mathsf{Q}_{\ell,t} = \left(\frac{\chi-1}{s\chi + (1-s)}\right)\mathsf{s}_{\ell,t} + \left(\frac{s\chi}{s\chi + (1-s)}\right)\mathsf{q}_{S,\ell,t} + \left(\frac{1-s}{s\chi + (1-s)}\right)\mathsf{q}_{N,\ell,t}. \qquad \text{[A.5.5]}$$

Define the following weighted averages of variables across the distribution of normal-price vintages. First, the average sale frequency:

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

Now the average normal price, the average quantity sold, and the purchase multiplier associated with the normal price:

$$\mathsf{P}_{N,t} \equiv (1-\phi_p)\sum_{\ell=0}^{\infty}\phi_p^{\ell}\mathsf{R}_{N,t-\ell}, \quad \mathsf{q}_{N,t} \equiv (1-\phi_p)\sum_{\ell=0}^{\infty}\phi_p^{\ell}\mathsf{q}_{N,\ell,t}, \quad \text{and} \quad \mathsf{v}_{N,t} \equiv (1-\phi_p)\sum_{\ell=0}^{\infty}\phi_p^{\ell}\mathsf{v}_{N,\ell,t}. \quad \text{[A.5.6]}$$

Finally, the average sale price and associated average quantity and purchase multiplier:

$$\mathsf{P}_{S,t} \equiv (1-\phi_p)\sum_{\ell=0}^{\infty}\phi_p^{\ell}\mathsf{p}_{S,\ell,t}, \quad \mathsf{q}_{S,t} \equiv (1-\phi_p)\sum_{\ell=0}^{\infty}\phi_p^{\ell}\mathsf{q}_{S,\ell,t}, \quad \text{and} \quad \mathsf{v}_{S,t} \equiv (1-\phi_p)\sum_{\ell=0}^{\infty}\phi_p^{\ell}\mathsf{v}_{S,\ell,t}. \quad \text{[A.5.7]}$$

The bargain hunters' price index $P_{B,t}$ as given in [4.5] is log-linearized as follows:

$$\mathsf{P}_{B,t} = \vartheta_B\mathsf{P}_{S,t} + (1-\vartheta_B)\mathsf{P}_{N,t} - \varphi_B\mathsf{s}_t, \quad \text{where} \quad \text{[A.5.8]}$$
$$\vartheta_B \equiv \left(\frac{s}{s+(1-s)\mu^{\eta-1}}\right), \quad \text{and} \quad \varphi_B \equiv \frac{1}{\eta-1}\left(\frac{1-\mu^{\eta-1}}{s+(1-s)\mu^{\eta-1}}\right),$$

using the averages defined above. The coefficients satisfy $0 \leq \vartheta_B \leq 1$ and $\varphi_B \geq 0$. By analogy with the expression for $P_{B,t}$ in [4.5], define a price index $P_{L,t}$ corresponding to the average purchase price for a hypothetical loyal customer:

$$P_{L,t} = \left((1-\phi_p)\sum_{\ell=0}^{\infty}\phi_p^{\ell}\left\{s_{\ell,t}p_{S,\ell,t}^{1-\epsilon} + (1-s_{\ell,t})R_{N,t-\ell}^{1-\epsilon}\right\}\right)^{\frac{1}{1-\epsilon}}. \quad \text{[A.5.9]}$$

This has the following log linearization:

$$\mathsf{P}_{L,t} = \vartheta_L\mathsf{P}_{S,t} + (1-\vartheta_L)\mathsf{P}_{N,t} - \varphi_L\mathsf{s}_t, \quad \text{where} \quad \text{[A.5.10]}$$
$$\vartheta_L \equiv \left(\frac{s}{s+(1-s)\mu^{\epsilon-1}}\right), \quad \text{and} \quad \varphi_L \equiv \frac{1}{\epsilon-1}\left(\frac{1-\mu^{\epsilon-1}}{s+(1-s)\mu^{\epsilon-1}}\right),$$

where the coefficients satisfy $0 \leq \vartheta_L \leq 1$ and $\varphi_L \geq 0$.

Note that [4.4], [4.5] and [A.5.9] imply that the price level P_t can be expressed in terms of $P_{L,t}$ and $P_{B,t}$:

$$P_t = \left(\lambda P_{L,t}^{1-\epsilon} + (1-\lambda)P_{B,t}^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}},$$

which can be log linearized to yield:

$$\mathsf{P}_t = (1-\varpi)\mathsf{P}_{L,t} + \varpi\mathsf{P}_{B,t}, \quad \text{where} \quad \varpi = \frac{(1-\lambda)}{(1-\lambda)+\lambda\hbar^{\epsilon-1}}, \quad \text{and} \quad \hbar = \frac{\left(s+(1-s)\mu^{\epsilon-1}\right)^{\frac{1}{\epsilon-1}}}{\left(s+(1-s)\mu^{\eta-1}\right)^{\frac{1}{\eta-1}}}, \quad \text{[A.5.11]}$$

with $\hbar$ being a bargain hunter's cost of consumption relative to a loyal customer, that is $\hbar \equiv P_B/P_L$, and ϖ denoting the weight on the bargain hunters' price index in the overall aggregate price level ($0 \leq \varpi \leq 1$). It is convenient to express the price level P_t in terms of the averages $\mathsf{P}_{S,t}$, $\mathsf{P}_{N,t}$ and s_t:

$$\mathsf{P}_t = \vartheta_P\mathsf{P}_{S,t} + (1-\vartheta_P)\mathsf{P}_{N,t} - \varphi_P\mathsf{s}_t, \quad \text{where} \quad \vartheta_P = (1-\varpi)\vartheta_L + \varpi\vartheta_B, \quad \text{and} \quad \varphi_P = (1-\varpi)\varphi_L + \varpi\varphi_B. \quad \text{[A.5.12]}$$

Note that $0 \leq \vartheta_P \leq 1$ and $\varphi_P \geq 0$ follow from the properties of the coefficients ϑ_B, ϑ_L, φ_L, φ_B and ϖ.

The log linearization of the production function [2.8] is

$$\mathsf{Q}_{\ell,t} = \alpha\mathsf{H}_{\ell,t}, \quad \text{where} \quad \alpha \equiv \frac{\mathcal{F}^{-1}(Q)\mathcal{F}'(\mathcal{F}^{-1}(Q))}{\mathcal{F}(\mathcal{F}^{-1}(Q))}. \quad \text{[A.5.13]}$$

The nominal marginal cost function corresponding to [2.9] has the following log-linear form:

$$\mathsf{X}_{\ell,t} = \gamma \mathsf{Q}_{\ell,t} + \mathsf{W}_t, \quad \text{where } \gamma \equiv \frac{Q\mathscr{C}''(Q;W)}{\mathscr{C}'(Q;W)} = \left(-\frac{\mathcal{F}^{-1}(Q)\mathcal{F}''(\mathcal{F}^{-1}(Q))}{\mathcal{F}'(\mathcal{F}^{-1}(Q))}\right)\left(\frac{Q}{\mathcal{F}^{-1}(Q)\mathcal{F}'(\mathcal{F}^{-1}(Q))}\right). \quad \text{[A.5.14]}$$

(i) The log-linearized first-order condition for the sales frequency (the first equation in [4.3]) is

$$(\chi - 1)\mathsf{X}_{\ell,t} = \mu_S \chi \mathsf{p}_{S,\ell,t} - \mu_N \mathsf{R}_{N,t-\ell} + (\mu_S - 1)\chi(\mathsf{q}_{S,\ell,t} - \mathsf{q}_{N,\ell,t}), \quad \text{[A.5.15]}$$

where the fact that $\chi = (\mu_N - 1)/(\mu_S - 1)$ is used to simplify the expression. By using equation [A.5.3]:

$$\begin{aligned}(\chi - 1)\mathsf{X}_{\ell,t} = &\left(\mu_S - (\mu_S - 1)\left(\frac{\lambda\epsilon + (1-\lambda)\eta v_S}{\lambda + (1-\lambda)v_S}\right)\right)\chi \mathsf{p}_{S,\ell,t} \\ &- \left(\mu_N - (\mu_N - 1)\left(\frac{\lambda\epsilon + (1-\lambda)\eta v_N}{\lambda + (1-\lambda)v_N}\right)\right)\mathsf{R}_{N,t-\ell} \\ &+ (\eta - \epsilon)\left(\frac{(1-\lambda)v_S}{\lambda + (1-\lambda)v_S} - \frac{(1-\lambda)v_N}{\lambda + (1-\lambda)v_N}\right)(\mu_S - 1)\chi \mathsf{P}_{B,t}.\end{aligned}$$

Given the expressions for μ_S and μ_N in [3.4], the coefficients of both $\mathsf{p}_{S,\ell,t}$ and $\mathsf{R}_{N,\ell,t}$ in the above are zero. Since $\chi > 1$, this equation implies $\mathsf{X}_{\ell,t}$ is independent of $\mathsf{p}_{S,\ell,t}$ and $\mathsf{R}_{N,t-\ell}$. Using $\chi = (\mu_N - 1)/(\mu_S - 1)$ yields:

$$\begin{aligned}(\chi - 1)\mathsf{X}_{\ell,t} = (\chi - 1)\mathsf{P}_{B,t} &- \left(1 - (\eta - \epsilon)\left(\frac{(1-\lambda)v_S}{\lambda + (1-\lambda)v_S}\right)(\mu_S - 1)\right)\chi \mathsf{P}_{B,t} \\ &+ \left(1 - (\eta - \epsilon)\left(\frac{(1-\lambda)v_N}{\lambda + (1-\lambda)v_N}\right)(\mu_N - 1)\right)\mathsf{P}_{B,t}. \quad \text{[A.5.16]}\end{aligned}$$

After substituting the expressions for μ_S and μ_N from [3.5], the above equation reduces to

$$(\chi - 1)\mathsf{X}_{\ell,t} = (\chi - 1)\mathsf{P}_{B,t} + (\epsilon - 1)\left((\mu_S - 1)\chi - (\mu_N - 1)\right)\mathsf{P}_{B,t},$$

and noting that the coefficient on the final term is zero, it follows that $\mathsf{X}_{\ell,t} = \mathsf{P}_{B,t}$ for all ℓ. Hence, all firms have the same marginal cost, $\mathsf{X}_t = \mathsf{P}_{B,t}$, irrespective of their normal-price vintage.

The optimal $p_{S,\ell,t}$ is characterized by the second equation in [4.3]. In log-linear terms it is

$$\mathsf{p}_{S,\ell,t} = \upmu_{S,\ell,t} + \mathsf{X}_t.$$

By substituting the expressions for the log-linearized optimal sale markup from [A.5.4] and the sale purchase multiplier from [A.5.2], and using $\mathsf{X}_t = \mathsf{P}_{B,t}$:

$$(1 - (\eta - \epsilon)\mathfrak{c}_S)(\mathsf{p}_{S,\ell,t} - \mathsf{X}_t) = 0, \quad \text{[A.5.17]}$$

so $\mathsf{p}_{S,\ell,t} = \mathsf{X}_t$ if the coefficient in the above is different from zero. The expressions for $\mathfrak{c}_S$ from [A.5.4] and μ_S from [3.5] imply

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

Using [A.2.8] 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.2.1] and [A.2.4]. The coefficient in [A.5.17] is strictly positive because $\mathcal{D}'(\rho_S) < 0$ and Lemma 2 shows that $\mathcal{R}''(\mathcal{D}(\rho_S)) < 0$, and therefore $\mathsf{p}_{S,\ell,t} = \mathsf{X}_t$.

Since all firms face the same wage W_t, and as the argument above shows that all have the same nominal marginal cost X_t, the log linearization of nominal marginal cost in [A.5.14] shows that all must produce the same total quantity Q_t when $\gamma > 0$.

The log-linearization of the first-order condition [4.2] for the optimal reset price $R_{N,t}$ simplifies to

$$\sum_{\ell=0}^{\infty}(\beta\phi_p)^{\ell}\mathbb{E}_t\left[\mathsf{R}_{N,t}-\mu_{N,\ell,t+\ell}-\mathsf{X}_{t+\ell}\right]=0, \quad \text{[A.5.18]}$$

where $\mu_{N,\ell,t}$ 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 those results together, it follows that $\mu_{N,\ell,t+\ell}=(\eta-\epsilon)\mathfrak{c}_N\left(\mathsf{R}_{N,t}-\mathsf{P}_{B,t+\ell}\right)$. So by using $\mathsf{X}_t=\mathsf{P}_{B,t}$ and substituting these results into [A.5.18]:

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

An exactly analogous argument to the proof of $1-(\eta-\epsilon)\mathfrak{c}_S>0$ above shows that $1-(\eta-\epsilon)\mathfrak{c}_N>0$ also holds. Hence:

$$\mathsf{R}_{N,t}=(1-\beta\phi_p)\sum_{\ell=0}^{\infty}(\beta\phi_p)^{\ell}\mathbb{E}_t\mathsf{X}_{t+\ell}. \quad \text{[A.5.19]}$$

(ii) By using $\mathsf{P}_{S,t}=\mathsf{X}_t$ and substituting this into [A.5.12] it is demonstrated that

$$\varphi_P\mathsf{s}_t=\vartheta_P(\mathsf{X}_t-\mathsf{P}_t)+(1-\vartheta_P)(\mathsf{P}_{N,t}-\mathsf{P}_t). \quad \text{[A.5.20]}$$

Likewise, by using $\mathsf{P}_{B,t}=\mathsf{X}_t$ and performing similar substitutions in the expression for $\mathsf{P}_{B,t}$ from [A.5.8]:

$$\varphi_B\mathsf{s}_t=(1-\vartheta_B)(\mathsf{P}_{N,t}-\mathsf{X}_t). \quad \text{[A.5.21]}$$

Equation [A.5.20] can be written as

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

and s_t is eliminated using [A.5.21]. After some rearrangement this leads to

$$\mathsf{X}_t-\mathsf{P}_{N,t}=\frac{1}{1-\psi}\mathsf{x}_t, \quad \text{[A.5.22]}$$

where $\mathsf{x}_t=\mathsf{X}_t-\mathsf{P}_t$ is real marginal cost and ψ is defined as follows:

$$\psi=\frac{(1-\vartheta_B)\varphi_P+\vartheta_P\varphi_B}{\varphi_B}. \quad \text{[A.5.23]}$$

Note that the recursive form of the expression for $\mathsf{P}_{N,t}$ in [A.5.6] is

$$\mathsf{P}_{N,t}=\phi_p\mathsf{P}_{N,t-1}+(1-\phi_p)\mathsf{R}_{N,t}, \quad \text{[A.5.24]}$$

and the recursive form of the equation [A.5.19] for $\mathsf{R}_{N,t}$ is:

$$\mathsf{R}_{N,t}=\beta\phi_p\mathbb{E}_t\mathsf{R}_{t+1}+(1-\beta\phi_p)\mathsf{X}_t. \quad \text{[A.5.25]}$$

Then multiplying both sides of the above by $(1-\phi_p)$ and substituting in the recursive equation 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 can be written in terms of normal-price inflation $\pi_{N,t}\equiv\mathsf{P}_{N,t}-\mathsf{P}_{N,t-1}$:

$$\pi_{N,t}=\beta\mathbb{E}_t\pi_{N,t+1}+\kappa(\mathsf{X}_t-\mathsf{P}_{N,t}), \quad \text{[A.5.26]}$$

and where $\kappa = (1-\phi_p)(1-\beta\phi_p)/\phi_p$ is as defined in the statement of the theorem.

Taking the first difference of [A.5.21] yields

$$\Delta s_t = -\frac{(1-\vartheta_B)}{\varphi_B}\left(\Delta X_t - \pi_{N,t}\right). \qquad \text{[A.5.27]}$$

Now use [A.5.12] and make the substitution $P_{S,t} = X_t$ as before, and then take first differences and rearrange:

$$\pi_t = \pi_{N,t} + \vartheta_P\left(\Delta X_t - \pi_{N,t}\right) - \varphi_P \Delta s_t.$$

By eliminating Δs_t from this equation using [A.5.27]:

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

Substituting the first difference of equation [A.5.22] into the above yields

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

Combining this equation with [A.5.22] and [A.5.26] implies

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

which is rearranged to yield the result [4.8]. Recursive forward substitution of equation [4.8] leads to

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

Notice that all $\Delta x_{t+\ell}$ terms apart from Δx_t cancel out because each occurs twice with opposite signs. Hence equation [4.9] is obtained.

(iii) Equation [A.5.23] implies that an expression for $1-\psi$ is

$$1-\psi = \frac{(1-\vartheta_P)\varphi_B - (1-\vartheta_B)\varphi_P}{\varphi_B}. \qquad \text{[A.5.28]}$$

It follows from [A.5.12] that $(1-\vartheta_P) = (1-\varpi)(1-\vartheta_L) + \varpi(1-\vartheta_B)$. Together with the expression for φ_P from the same equation, [A.5.28] implies

$$1-\psi = \frac{\left((1-\varpi)(1-\vartheta_L) + \varpi(1-\vartheta_B)\right)\varphi_B - (1-\vartheta_B)\left((1-\varpi)\varphi_L + \varpi\varphi_B\right)}{\varphi_B},$$

and by rearranging this expression:

$$1-\psi = (1-\varpi)\varphi_L\left(\frac{1-\vartheta_L}{\varphi_L} - \frac{1-\vartheta_B}{\varphi_B}\right). \qquad \text{[A.5.29]}$$

Define the function

$$\Phi(\zeta;\mu) \equiv \frac{\mu^{-\zeta}-1}{\zeta} \qquad \text{[A.5.30]}$$

in terms of the markup ratio μ. An alternative expression for this function is $\Phi(\zeta;\mu) = (e^{(-\log\mu)\zeta}-1)/\zeta$, which shows that it has derivative

$$\Phi'(\zeta;\mu) = \frac{\left((-\log\mu)\zeta - 1\right)e^{(-\log\mu)\zeta} + 1}{\zeta^2}.$$

Now define another function

$$\mathcal{J}(z) \equiv 1 + (z-1)e^z,$$

and note that $\mathcal{J}'(z) = ze^z$. Since $\mathcal{J}(0) = 0$, and $\mathcal{J}'(z) > 0$ for all $z > 0$, it follows that $\mathcal{J}(z) > 0$ for all $z > 0$. Then note

$$\Phi'(\zeta;\mu) = \frac{\mathcal{J}\left((-\log\mu)\zeta\right)}{\zeta^2},$$

which proves that $\Phi(\zeta;\mu)$ is strictly increasing in ζ when $\zeta > 0$ since $0 < \mu < 1$.

The expressions for ϑ_L and φ_L given in [A.5.10] are now used to demonstrate that:

$$\frac{1-\vartheta_L}{\varphi_L} = (1-s)\left(\frac{\epsilon - 1}{(\mu^{-1})^{\epsilon-1} - 1}\right) = \frac{1-s}{\Phi(\epsilon-1;\mu)}. \qquad \text{[A.5.31]}$$

Similarly, the expressions for ϑ_B and φ_B from [A.5.8] yield

$$\frac{1-\vartheta_B}{\varphi_B} = (1-s)\left(\frac{\eta - 1}{(\mu^{-1})^{\eta-1} - 1}\right) = \frac{1-s}{\Phi(\eta-1;\mu)}. \qquad \text{[A.5.32]}$$

These formulæ are then substituted into [A.5.29] to obtain:

$$1-\psi = (1-\varpi)(1-s)\varphi_L\left(\frac{1}{\Phi(\epsilon-1;\mu)} - \frac{1}{\Phi(\eta-1;\mu)}\right).$$

The expression for ψ in [A.5.23] together with the properties of ϑ_B, ϑ_P, φ_B and φ_P derived earlier demonstrates that $\psi \geq 0$. The inequality $\psi \leq 1$ follows from $\Phi(\zeta;\mu)$ being an increasing function of ζ together with $\eta > \epsilon$ and the properties of ϖ and φ_L. Thus, it is established that $0 \leq \psi \leq 1$.

Now use [A.5.31] to obtain the following:

$$1-\psi = (1-\varpi)(1-\vartheta_L)\left(1 - \Theta(\epsilon,\eta;\mu)\right), \quad \text{where } \Theta(\epsilon,\eta;\mu) \equiv \frac{\Phi(\epsilon-1;\mu)}{\Phi(\eta-1;\mu)}. \qquad \text{[A.5.33]}$$

Note that the expression for P_B in [3.7] can be substituted into $v(p_S;P_B)$ from [2.7] to obtain:

$$v_S = \frac{1}{(s + (1-s)\mu^{\eta-1})^{\frac{\eta-\epsilon}{\eta-1}}},$$

and which by combining this with the expression for $\hbar$ from [A.5.11] yields

$$\hbar^{\epsilon-1} = \frac{1}{v_S}\left(\frac{s + (1-s)\mu^{\epsilon-1}}{s+(1-s)\mu^{\eta-1}}\right).$$

Thus, the weight $1-\varpi$ given in [A.5.11] is

$$1-\varpi = \frac{\lambda(s+(1-s)\mu^{\epsilon-1})}{\lambda(s+(1-s)\mu^{\epsilon-1}) + (1-\lambda)v_S(s+(1-s)\mu^{\eta-1})}.$$

Substituting this into [A.5.33] and using the formula for ϑ_L from [A.5.10] implies

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

Since the purchase multipliers are given by $v_N = \rho_N^{-(\eta-\epsilon)}$ and $v_S = \rho_S^{-(\eta-\epsilon)}$, the expressions for ρ_S and ρ_N from Lemma 3 imply that

$$(1-\lambda)v_N = \lambda z, \quad \text{and} \quad (1-\lambda)v_S = \mu^{\epsilon-\eta}\lambda z, \qquad \text{[A.5.34]}$$

where $z = \mathfrak{z}(\mu;\epsilon,\eta)$ is the value of the function in [A.1.3]. Substituting v_S into the expression for $1-\psi$

above yields

$$1-\psi=(1-s)(1-\Theta(\epsilon,\eta;\mu))\frac{\mu^{\epsilon-1}}{(s+(1-s)\mu^{\epsilon-1})+\mu^{\epsilon-\eta}z(s+(1-s)\mu^{\eta-1})}.$$

After further rearrangement this implies

$$1-\psi=\frac{(1-\Theta(\epsilon,\eta;\mu))(1-s)}{(1+z)+((\mu^{1-\epsilon}-1)+(\mu^{1-\eta}-1)z)\,s}. \qquad \text{[A.5.35]}$$

For parameters consistent with a two-price equilibrium, Lemma 3 shows that $z=\mathfrak{z}(\mu;\epsilon,\eta)$ must be a positive real number. The definition of $\Phi(\zeta;\mu)$ in [A.5.30] implies that it is non-negative when $0<\mu<1$ and $\zeta>0$. Since $\eta>\epsilon>1$ and as $\Phi(\zeta;\mu)$ is increasing in ζ, the definition of $\Theta(\epsilon,\eta;\mu)$ in [A.5.33] ensures that $0\leq\Theta(\epsilon,\eta;\mu)\leq 1$. Hence, because all terms in the expression above for $1-\psi$ are positive, the derivative with respect to s (holding ϵ and η constant, and hence μ and z constant by Lemma 3) is negative. Proposition 1 shows that λ and s are negatively related (holding ϵ and η constant), so ψ is strictly decreasing in λ.

By using [A.1.4], it follows that $\mu\chi=\mu^{1-\epsilon}(1+\mu^{\epsilon-\eta}z)/(1+z)$, and hence

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

This expression is substituted into [A.5.35] to yield

$$1-\psi=\frac{(1-\Theta(\epsilon,\eta;\mu))(1-s)}{(1+z)(s\mu\chi+(1-s))}. \qquad \text{[A.5.36]}$$

Note that $\psi=1$ requires the right-hand side of this expression to be zero. There are four terms to consider. First, $s=1$ is the only way the expression can be zero as a result of the $1-s$ term. Now consider the terms in the denominator. Since $\mu=p_S/p_N$ and $\chi=q_S/q_N$, the second term in the denominator is linked to the GDP share transacted at the normal price:

$$\frac{1}{s\mu\chi+(1-s)}=\frac{1}{1-s}\left(\frac{(1-s)p_Nq_N}{sp_Sq_S+(1-s)p_Nq_N}\right).$$

So when $s<1$, $(s\mu\chi+(1-s))\to\infty$ only if $(1-s)p_Nq_N/(sp_Sq_S+(1-s)p_Nq_N)\to 0$, that is, the GDP share traded at the sticky normal price tends to zero. The other term in the denominator is $1+z$, where $z=\mathfrak{z}(\mu;\epsilon,\eta)$, which is the smallest root of the quadratic [A.2.33]. As the proof of Lemma 3 demonstrates, this quadratic must always have two positive real roots in the relevant parameter range. The product of these roots is obtained from the coefficients of the quadratic in [A.2.33]:

$$\left(\frac{\epsilon(\epsilon-1)}{\eta(\eta-1)}\right)\mu^{\eta-\epsilon},$$

which is always less than one, hence $1+z$ is finite, so the only way the denominator of [A.5.36] can approach infinity is through the normal-price GDP share approaching zero.

The final possibility to consider is $\Theta(\epsilon,\eta;\mu)=1$. The function $\Theta(\epsilon,\eta;\mu)$ from [A.5.33] can be written as:

$$\Theta(\epsilon,\eta;\mu)=\left(\frac{\eta-1}{\epsilon-1}\right)\left(\frac{e^{(-\log\mu)(\epsilon-1)}-1}{e^{(-\log\mu)(\eta-1)}-1}\right),$$

and by L'Hôpital's rule:

$$\lim_{\mu\to 1}\Theta(\epsilon,\eta;\mu)=1,$$

for any elasticities ϵ and η such that $1<\epsilon<\eta$, so $\mu=1$ is also a possible way that $\psi=1$ could occur. Now take any other parameters ϵ and η such that $0\leq\mu<1$. The non-monotonicity condition [3.3] is necessary for an equilibrium with $\mu < 1$ to exist. Note that [3.3] implies that ϵ can never approach η in the region of parameters consistent with $\mu < 1$. Since $\Phi(\zeta; \mu)$ is known to be strictly increasing in ζ for any $0 < \mu < 1$, and that η is bounded away from ϵ, it follows that $\Phi(\epsilon - 1; \mu) < \Phi(\eta - 1; \mu)$ and thus $\Theta(\epsilon, \eta; \mu) < 1$ for any $\mu < 1$. This argument establishes that $\mu = 1$ is the only other possible way that $\psi = 1$ can occur, and so completes the proof.

The arguments developed in the proof above lead to the following set of results characterizing the fluctuations in other variables of interest.

Lemma 4 *The Phillips curve in [4.8] is a relationship between aggregate inflation π_t and real marginal cost x_t. Underlying this relationship are the following:*

(i) The average sale discount $P_{N,t} - P_{S,t}$ is determined by real marginal cost x_t. There is a negative relationship between $P_{N,t} - P_{S,t}$ and x_t, and the magnitude of the response of the average sale discount to real marginal cost is decreasing in λ.

(ii) The average quantity ratio $q_{S,t} - q_{N,t}$ is determined by real marginal cost x_t. There is a positive relationship between $q_{S,t} - q_{N,t}$ and x_t, and the magnitude of the response of the average quantity ratio to real marginal cost is decreasing in λ.

(iii) The average sales frequency s_t is determined by real marginal cost x_t. There is a negative relationship between s_t and x_t, and the magnitude of the response of the average sales frequency to real marginal cost is decreasing in λ.

(iv) A firm with a normal price above the average has a sale discount above the average and a sales frequency above the average.

(v) Relative price distortions $Q_t - Y_t$ are negatively related to real marginal cost x_t.

Proof (i) Let $\mu_t = P_{S,t} - P_{N,t}$. Using the result $P_{S,t} = X_t$ from Theorem 2 and [A.5.22], it follows that

$$\mu_t = \frac{1}{1-\psi} x_t. \qquad [A.5.37]$$

The coefficient on x_t is known to be positive because of the inequality for ψ derived in Theorem 2. Its magnitude is decreasing in λ because ψ is negatively related to λ, as shown in Theorem 2.

(ii) Let $\chi_t = q_{S,t} - q_{N,t}$. The log-linearized demand functions and purchase multipliers in [A.5.1] and [A.5.2] imply

$$\chi_t = -\zeta_N (P_{N,t} - P_{S,t}),$$

with ζ_N being the steady-state price elasticity at the normal price, and where $P_{S,t} = P_{B,t}$ has been used. Substitution of the result in [A.5.37] yields

$$\chi_t = \frac{\zeta_N}{1-\psi} x_t.$$

Using the inequality for ψ from Theorem 2 and $\zeta_N > 0$, it follows that the coefficient of x_t in the above is positive. By combining the expression for ζ_N from [3.1] and equation [A.5.34]:

$$\zeta_N = \frac{\epsilon + z\eta}{1+z}. \qquad [A.5.38]$$

Since $z = \mathfrak{z}(\mu; \epsilon, \eta)$, it follows from Lemma 3 that ζ_N is independent of λ. Hence, since Theorem 2 shows that ψ is decreasing in λ, the coefficient of x_t in the equation for χ_t is also decreasing in λ.

(iii) For the average sales frequency s_t, use equation [A.5.21] together with $X_t = P_{S,t}$ and the expression for μ_t in [A.5.37] to obtain:

$$s_t = -\left(\frac{1-\vartheta_B}{\varphi_B}\right)\left(\frac{1}{1-\psi}\right) x_t. \qquad [A.5.39]$$

It has been shown that $0 \leq \vartheta_B \leq 1$, $\varphi_B \geq 0$, and $0 \leq \psi \leq 1$, so it follows that the coefficient of x_t above is negative. By substituting the expressions for $(1-\vartheta_B)/\varphi_B$ from [A.5.32] and $1-\psi$ from [A.5.35] into the above:

$$\left(\frac{1-\vartheta_B}{\varphi_B}\right)\left(\frac{1}{1-\psi}\right) = \frac{(1+z) + \left((\mu^{1-\epsilon}-1) + (\mu^{1-\eta}-1)z\right)s}{\Phi(\eta-1;\mu)\left(1-\Theta(\epsilon,\eta;\mu)\right)}.$$

Since all the terms in the denominator and μ and z in the numerator are independent of λ, it follows that the magnitude of this coefficient is decreasing in λ because s is decreasing in λ.

(iv) Since Theorem 2 implies that $\mathsf{p}_{S,\ell,t} = \mathsf{P}_{S,t}$ for all ℓ, it follows that:

$$(\mathsf{R}_{N,t-\ell} - \mathsf{p}_{S,\ell,t}) - (\mathsf{P}_{N,t} - \mathsf{P}_{S,t}) = (\mathsf{R}_{N,t-\ell} - \mathsf{P}_{N,t}),$$

where $\mathsf{R}_{N,t-\ell} - \mathsf{P}_{N,t}$ clearly has a positive coefficient. A further consequence of $\mathsf{p}_{S,\ell,t} = \mathsf{P}_{S,t}$ is that $\mathsf{q}_{S,\ell,t} = \mathsf{q}_{S,t}$ for all ℓ. The demand function in [A.5.3] implies $\mathsf{q}_{N,\ell,t} - \mathsf{q}_{N,t} = -\zeta_N(\mathsf{R}_{N,t-\ell} - \mathsf{P}_{N,t})$. Together with equation [A.5.5] and the result $\mathsf{Q}_{\ell,t} = \mathsf{Q}_t$ from Theorem 2:

$$\mathsf{s}_{\ell,t} - \mathsf{s}_t = \frac{(1-s)\zeta_N}{\chi - 1}(\mathsf{R}_{N,t-\ell} - \mathsf{P}_{N,t}),$$

with the coefficient on $\mathsf{R}_{N,t-\ell} - \mathsf{P}_{N,t}$ in the above being positive.

(v) Let $\Delta_t = \mathsf{Y}_t - \mathsf{Q}_t$. From the expression for the log-linearized demand function and purchase multipliers in [A.5.1] and [A.5.2], the following individual demand functions are obtained:

$$\mathsf{q}_{S,t} = -\epsilon\mathsf{x}_t + \mathsf{Y}_t, \quad \mathsf{q}_{N,t} = -\epsilon\mathsf{x}_t + \mathsf{Y}_t - \zeta_N(\mathsf{P}_{N,t} - \mathsf{P}_{S,t}),$$

where the results $\mathsf{P}_{S,t} = \mathsf{P}_{B,t} = \mathsf{X}_t$ from Theorem 2 have been used. By substituting these into the expression for total quantity from [A.5.5]:

$$\mathsf{Q}_t = \mathsf{Y}_t - \epsilon\mathsf{x}_t - \zeta_N\left(\frac{(1-s)}{s\chi + (1-s)}\right)(\mathsf{P}_{N,t} - \mathsf{P}_{S,t}) + \left(\frac{\chi-1}{s\chi + (1-s)}\right)\mathsf{s}_t.$$

Substituting [A.5.37] and [A.5.39] in the above expression yields

$$\Delta_t \equiv \mathsf{Y}_t - \mathsf{Q}_t = \left(\epsilon + \frac{1}{(s\chi + (1-s))(1-\psi)}\left((\chi-1)\left(\frac{1-\vartheta_B}{\varphi_B}\right) - (1-s)\zeta_N\right)\right)\mathsf{x}_t.$$

This is written as $\Delta_t = \delta\mathsf{x}_t$, with the coefficient δ of real marginal cost x_t defined by:

$$\delta = \frac{s\chi\mu + (1-s)}{s\chi + (1-s)}\left(\epsilon\frac{s\chi + (1-s)}{s\chi\mu + (1-s)} + \wp\right), \qquad \text{[A.5.40]}$$

and where the term $\wp$ is:

$$\wp = \frac{1}{(1-\psi)(s\mu\chi + (1-s))}\left((\chi-1)\left(\frac{1-\vartheta_B}{\varphi_B}\right) - (1-s)\zeta_N\right).$$

By substituting the expression for $1-\psi$ from [A.5.36] and rearranging:

$$\wp = \frac{1+z}{1-\Theta(\epsilon,\eta;\mu)}\left((\chi-1)\left(\frac{1-\vartheta_B}{\varphi_B(1-s)}\right) - \zeta_N\right).$$

Equation [A.5.32] then implies

$$\wp = \frac{1+z}{1-\Theta(\epsilon,\eta;\mu)}\left(\frac{\chi-1}{\Phi(\eta-1;\mu)} - \zeta_N\right). \qquad \text{[A.5.41]}$$

Noting that equation [A.1.4] can be used to express $\chi - 1$ as follows:

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

and substituting this together with the formula for ζ_N in [A.5.38] into the expression for $\wp$ from [A.5.41]:

$$\wp = \frac{(\mu^{-\epsilon} - 1) + z\left(\mu^{-\eta} - 1\right) - (\epsilon + \eta z)\Phi(\eta - 1; \mu)}{\left(1 - \Theta(\epsilon, \eta; \mu)\right)\Phi(\eta - 1; \mu)}.$$

By using the definitions of the functions $\Phi(\zeta; \mu)$ and $\Theta(\epsilon, \eta; \mu)$ from [A.5.30] and [A.5.33]:

$$\wp = \frac{\epsilon\left(\Phi(\epsilon; \mu) - \Phi(\eta - 1; \mu)\right) + z\eta\left(\Phi(\eta; \mu) - \Phi(\eta - 1; \mu)\right)}{\Phi(\eta - 1; \mu) - \Phi(\epsilon - 1; \mu)}.$$

The expression for δ from [A.5.40] can thus be written as:

$$\delta = \frac{s\chi\mu + (1-s)}{s\chi + (1-s)}\left(\epsilon\frac{s\chi + (1-s)}{s\chi\mu + (1-s)} + \frac{\epsilon\left(\Phi(\epsilon; \mu) - \Phi(\eta - 1; \mu)\right) + z\eta\left(\Phi(\eta; \mu) - \Phi(\eta - 1; \mu)\right)}{\Phi(\eta - 1; \mu) - \Phi(\epsilon - 1; \mu)}\right).$$

The final expression for δ is obtained by adding and subtracting ϵ inside the brackets:

$$\delta = \frac{s\chi\epsilon(1-\mu)}{s\chi + (1-s)} + \frac{s\chi\mu + (1-s)}{s\chi + (1-s)}\left(\frac{\epsilon\left(\Phi(\epsilon; \mu) - \Phi(\epsilon - 1; \mu)\right) + z\eta\left(\Phi(\eta; \mu) - \Phi(\eta - 1; \mu)\right)}{\Phi(\eta - 1; \mu) - \Phi(\epsilon - 1; \mu)}\right).$$

Since the function $\Phi(\zeta; \mu)$ from [A.5.30] is known to be strictly increasing in ζ, it follows that δ is positive. This completes the proof. ■

A.6 DSGE model derivations

Wage-setting behaviour

When each firm chooses its use of the continuum of labour inputs to minimize the cost of obtaining a unit of H from equation [4.6], the minimized cost is given by the wage index

$$W \equiv \left(\int W(\imath)^{1-\varsigma} d\imath\right)^{\frac{1}{1-\varsigma}}, \qquad \text{[A.6.1]}$$

and the cost-minimizing labour demand functions are

$$H(\imath) = \left(\frac{W(\imath)}{W}\right)^{-\varsigma} H. \qquad \text{[A.6.2]}$$

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 [A.6.2]. As shown by Erceg, Henderson and Levin (2000), the first-order condition for this maximization problem is

$$\sum_{\ell=0}^{\infty}(\beta\phi_w)^{\ell}\mathbb{E}_t\left[\frac{W_{t+\ell}^{\varsigma}H_{t+\ell}\upsilon_c(Y_{t+\ell})}{\upsilon_c(Y_t)}\left\{\frac{R_{W,t}}{P_{t+\ell}} - \frac{\varsigma}{\varsigma - 1}\frac{\nu_h\left(R_{W,t}^{-\varsigma}W_{t+\ell}^{\varsigma}H_{t+\ell}\right)}{\upsilon_c(Y_{t+\ell})}\right\}\right] = 0. \qquad \text{[A.6.3]}$$

The wage index W_t in [A.6.1] then evolves according to

$$W_t = \left((1 - \phi_w) \sum_{\ell=0}^{\infty} \phi_w^{\ell} R_{W,t-\ell}^{1-\varsigma} \right)^{\frac{1}{1-\varsigma}}. \qquad \text{[A.6.4]}$$

The presence of market power in wage setting means that the equation [3.11] determining steady-state output Y is replaced by

$$x = \frac{\varsigma}{\varsigma - 1} \frac{\nu_h \left(\mathcal{F}^{-1}(Y/\Delta)\right)}{\upsilon_c(Y)\mathcal{F}' \left(\mathcal{F}^{-1}(Y/\Delta)\right)}.$$

All other equations determining the steady state are unaffected.

Log linearizations

The DSGE model is log linearized around the flexible-price equilibrium characterized in section 3. The notational convention is that a variable without a time subscript denotes its flexible-price steady-state value, and the corresponding sans serif letter denotes the log deviation of the variable from its steady-state value (except for the inflation rate and the nominal interest rate, where it denotes the log deviation of the corresponding gross rates).

The log linearization of the intertemporal IS equation in [4.10] is

$$\mathsf{Y}_t = \mathbb{E}_t \mathsf{Y}_{t+1} - \theta_c \left(\mathsf{i}_t - \mathbb{E}_t \pi_{t+1}\right), \quad \text{where } \theta_c \equiv -\left(\frac{Y \upsilon_{cc}(Y)}{\upsilon_c(Y)}\right)^{-1}. \qquad \text{[A.6.5]}$$

The intertemporal elasticity of substitution is θ_c. Money demand is implied by the binding cash-in-advance constraint in [4.10]. It is log linearized as follows:

$$\mathsf{M}_t - \mathsf{P}_t = \mathsf{Y}_t. \qquad \text{[A.6.6]}$$

The money supply rule [4.11] has the following log-linear form:

$$\boldsymbol{\Delta}\mathsf{M}_t = \mathfrak{p}\boldsymbol{\Delta}\mathsf{M}_{t-1} + (1 - \mathfrak{p})\mathsf{e}_t. \qquad \text{[A.6.7]}$$

The log-linearized version of equation [A.6.3] for the utility-maximizing reset wage is

$$\mathsf{R}_{W,t} = (1 - \beta\phi_w) \sum_{\ell=0}^{\infty} (\beta\phi_w)^{\ell} \mathbb{E}_t \left[\left(\frac{1}{1 + \varsigma\theta_h^{-1}} \right) (\mathsf{P}_{t+\ell} + \mathsf{w}_{t+\ell}^*) + \left(\frac{\varsigma\theta_h^{-1}}{1 + \varsigma\theta_h^{-1}} \right) \mathsf{W}_{t+\ell} \right], \qquad \text{[A.6.8]}$$

with w_t^* being the desired real wage in the absence of constraints on wage adjustment:

$$\mathsf{w}_t^* = \theta_h^{-1}\mathsf{H}_t + \theta_c^{-1}\mathsf{Y}, \quad \text{where } \theta_h \equiv \left(\frac{\mathcal{F}^{-1}(Y/\Delta)\nu_{hh}(\mathcal{F}^{-1}(Y/\Delta))}{\nu_h(\mathcal{F}^{-1}(Y/\Delta))} \right)^{-1}. \qquad \text{[A.6.9]}$$

The Frisch elasticity of labour supply is θ_h. Equation [A.6.8] has the following recursive form:

$$\mathsf{R}_{W,t} = \beta\phi_w \mathbb{E}_t \mathsf{R}_{W,t+1} + (1 - \beta\phi_w) \left(\left(\frac{1}{1 + \varsigma\theta_h^{-1}} \right) (\mathsf{P}_t + \mathsf{w}_t^*) + \left(\frac{\varsigma\theta_h^{-1}}{1 + \varsigma\theta_h^{-1}} \right) \mathsf{W}_t \right). \qquad \text{[A.6.10]}$$

The log-linearized wage index [A.6.4] 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:

$$\mathsf{W}_t = \phi_w \mathsf{W}_{t-1} + (1 - \phi_w)\mathsf{R}_{W,t}. \qquad \text{[A.6.11]}$$

Combining the reset wage equation [A.6.10] with the wage index equation [A.6.11] yields an expression for wage inflation $\pi_{W,t} \equiv \mathsf{W}_t - \mathsf{W}_{t-1}$:

$$\pi_{W,t} = \beta\mathbb{E}_t\pi_{W,t+1} + \frac{(1-\phi_w)(1-\beta\phi_w)}{\phi_w}\frac{1}{1+\varsigma\theta_h^{-1}}\left(\mathsf{w}_t^* - \mathsf{w}_t\right), \tag{A.6.12}$$

where w_t^* is defined in [A.6.9].

By averaging over normal-price vintages, equations [A.5.13] and [A.5.14] imply:

$$\mathsf{Q}_t = \alpha\mathsf{H}_t, \quad \text{and} \quad \mathsf{x}_t = \mathsf{w}_t + \gamma\mathsf{Q}_t. \tag{A.6.13}$$

Substituting $\mathsf{Y}_t = \mathsf{Q}_t + \delta\mathsf{x}_t$ from Lemma 4 into the above yields [A.1.9b]. Using equation [A.6.13] to eliminate H_t from [A.6.9] implies:

$$\mathsf{w}_t^* = \frac{\theta_h^{-1}}{\alpha}\mathsf{Q}_t + \theta_c^{-1}\mathsf{Y}_t.$$

Then by using $\mathsf{Q}_t = \mathsf{Y}_t - \delta\mathsf{x}_t$ to eliminate Q_t and substituting in the expression for x_t from [A.1.9b] leads to the following expression for $\mathsf{w}_t^* - \mathsf{w}_t$:

$$\mathsf{w}_t^* - \mathsf{w}_t = \left(\theta_c^{-1} + \frac{1}{1+\gamma\delta}\frac{\theta_h^{-1}}{\alpha}\right)\mathsf{Y}_t - \left(1 + \frac{\delta}{1+\gamma\delta}\frac{\theta_h^{-1}}{\alpha}\right)\mathsf{w}_t.$$

Replacing $\mathsf{w}_t^* - \mathsf{w}_t$ in [A.6.12] with the expression above yields [A.1.9c].

A.7 Two-sector model

DSGE model

The steady state of the two-sector model from section 5 is derived exactly as for the one-sector model by taking the sale sector as representative of the whole economy. This steady state is characterized in section 3.4 and can be computed as described in section A.1.

The system of equations of the two-sector DSGE model with sales is

$$\bar{\pi}_t = \beta\mathbb{E}_t\bar{\pi}_{t+1} + \frac{1}{1-\bar{\psi}}\left(\kappa\mathsf{x}_t + \bar{\psi}\left(\boldsymbol{\Delta}\mathsf{x}_t - \beta\mathbb{E}_t\boldsymbol{\Delta}\mathsf{x}_{t+1}\right)\right) + \left(\frac{1-\sigma}{1-\bar{\psi}}\left(\kappa\rho_t + \boldsymbol{\Delta}\rho_t - \beta\mathbb{E}_t\boldsymbol{\Delta}\rho_{t+1}\right)\right); \tag{A.7.1a}$$

$$\boldsymbol{\Delta}\rho_t = \beta\mathbb{E}_t\boldsymbol{\Delta}\rho_{t+1} + \frac{\kappa}{1+\xi\gamma}\left(\frac{\gamma((1-\psi)\delta+\psi\epsilon-\xi)}{1-\bar{\psi}}\mathsf{x}_t - \frac{\gamma(1-\psi)(\epsilon-(1-\sigma)\delta)+(1-\bar{\psi})+(1-\sigma)\xi\gamma}{1-\bar{\psi}}\rho_t\right); \tag{A.7.1b}$$

$$\mathsf{Y}_t = \bar{\mathsf{Y}}_t + \epsilon\left(\frac{1-\sigma}{1-\bar{\psi}}\right)\left((1-\psi)\rho_t - \psi\mathsf{x}_t\right); \tag{A.7.1c}$$

$$\mathsf{Y}_t = \mathsf{Q}_t + \delta\left(\frac{1-\psi}{1-\bar{\psi}}\right)\left(\mathsf{x}_t + (1-\sigma)\rho_t\right); \tag{A.7.1d}$$

$$\mathsf{x}_t = \mathsf{w}_t + \gamma\mathsf{Q}_t; \tag{A.7.1e}$$

$$\pi_{W,t} = \beta\mathbb{E}_t\pi_{W,t+1} + \frac{(1-\phi_w)(1-\beta\phi_w)}{\phi_w}\frac{1}{1+\varsigma\theta_h^{-1}}(\mathsf{w}_t^* - \mathsf{w}_t); \tag{A.7.1f}$$

$$\mathsf{w}_t^* = \left(\theta_c^{-1} + \frac{\theta_h^{-1}}{\alpha}\frac{\varDelta}{\sigma+(1-\sigma)\varDelta}\right)\bar{\mathsf{Y}}_t + \frac{\theta_h^{-1}}{\alpha}\frac{\sigma}{\sigma+(1-\sigma)\varDelta}\left(\mathsf{Q}_t - \varDelta\mathsf{Y}_t\right); \tag{A.7.1g}$$

$$\boldsymbol{\Delta}\mathsf{w}_t = \pi_{W,t} - \bar{\pi}_t; \tag{A.7.1h}$$

$$\bar{\mathsf{Y}}_t = \mathbb{E}_t\bar{\mathsf{Y}}_{t+1} - \theta_c\left(\mathsf{i}_t - \mathbb{E}_t\bar{\pi}_{t+1}\right); \tag{A.7.1i}$$

$$\boldsymbol{\Delta}\bar{\mathsf{Y}}_t = \boldsymbol{\Delta}\mathsf{M}_t - \bar{\pi}_t; \tag{A.7.1j}$$

$$\boldsymbol{\Delta}\mathsf{M}_t = \mathfrak{p}\boldsymbol{\Delta}\mathsf{M}_{t-1} + (1-\mathfrak{p})\mathsf{e}_t. \tag{A.7.1k}$$

A bar above a variable denotes the log-deviation averaged across both sale and non-sale sectors, using the appropriate weights (σ and $1-\sigma$), and this convention is also employed for the Phillips curve coefficient ψ, with $\bar{\psi}$ denoting the average Phillips curve coefficient $\sigma\psi$. All variables without a bar refer either to economy-wide aggregates, or sale-sector variables as used in earlier sections, as appropriate. The coefficients Δ, ψ, δ, ξ and κ are calculated using the same formulæ as those for the one-sector economy given in appendix A.1 taking the sale sector as representative of the whole economy.

Derivation of the two-sector model

In the following, the notational conventions in addition to those already described are that large script letters denote non-sale sector variables and small script letters denote the corresponding log deviations of the non-sale sector variables.

The aggregate price level is now

$$\bar{P}_t = \left(\sigma P_t^{1-\epsilon} + (1-\sigma)\mathcal{P}_t^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}},$$

which has the log linear form:

$$\bar{\mathsf{P}}_t = \sigma \mathsf{P}_t + (1-\sigma)\mathscr{P}_t. \qquad [A.7.2]$$

The log-linearized price level $\mathscr{P}_t$ in the non-sale sector is a weighted average of past reset prices $\mathscr{R}_t$ in that sector:

$$\mathscr{P}_t = \phi_p \mathscr{P}_{t-1} + (1-\phi_p)\mathscr{R}_t. \qquad [A.7.3]$$

The log-linearized first-order condition for the non-sale sector reset price is standard:

$$\mathscr{R}_t = \beta\phi_p \mathbb{E}_t \mathscr{R}_{t+1} + (1-\beta\phi_p)\left(\frac{1}{1+\xi\gamma}\mathscr{X}_t + \frac{\xi\gamma}{1+\xi\gamma}\mathscr{P}_t\right), \qquad [A.7.4]$$

where ξ is the constant price elasticity in that sector and γ is the elasticity of marginal cost with respect to output at the firm level.

Optimization by households implies the following overall relative demand between the sale and non-sale sectors:

$$\frac{\mathcal{Y}_t}{Y_t} = \left(\frac{\mathcal{P}_t}{P_t}\right)^{-\epsilon},$$

which has the log-linear form:

$$\mathscr{Y}_t - \mathsf{Y}_t = -\epsilon(\mathscr{P}_t - \mathsf{P}_t). \qquad [A.7.5]$$

Define $\rho_t \equiv \mathscr{P}_t - \mathsf{P}_{N,t}$ to be the average relative price between the non-sale sector and the normal prices in the sale sector. Substituting the sale-sector price level equation into the aggregate price level leads to

$$\bar{\mathsf{P}}_t = (1-\bar{\vartheta}_P)\mathsf{P}_{N,t} + \bar{\vartheta}_P \mathsf{P}_{S,t} - \bar{\varphi}_P \mathsf{s}_t + (1-\sigma)\rho_t, \qquad [A.7.6]$$

where $\bar{\vartheta}_P = \sigma\vartheta_P$ and $\bar{\varphi}_P = \sigma\varphi_P$ are defined (by analogy with the aggregate Phillips curve coefficient $\bar{\psi}$).

Real marginal cost x_t for the sale sector is defined in the usual way. By using equation [A.7.6]:

$$\mathsf{x}_t = (1-\bar{\vartheta}_P)(\mathsf{P}_{S,t} - \mathsf{P}_{N,t}) + \bar{\varphi}_P \mathsf{s}_t - (1-\sigma)\rho_t,$$

where $\mathsf{X}_t = \mathsf{P}_{S,t}$ has been substituted. Then by using [A.5.21] to eliminate s_t and rearranging:

$$\mathsf{x}_t = \left(\frac{(1-\bar{\vartheta}_P)\varphi_B - (1-\vartheta_B)\bar{\varphi}_P}{\varphi_B}\right)(\mathsf{P}_{S,t} - \mathsf{P}_{N,t}) - (1-\sigma)\rho_t.$$

Noting that the coefficient in parentheses is $1-\sigma\psi$, which is also equal to $1-\bar{\psi}$ using the definition of $\bar{\psi}$, the equation above can be solved for $\mathsf{P}_{S,t} - \mathsf{P}_{N,t}$:

$$\mathsf{P}_{S,t} - \mathsf{P}_{N,t} = \frac{1}{1-\bar{\psi}}\left(\mathsf{x}_t + (1-\sigma)\rho_t\right). \qquad [A.7.7]$$

Using equation [A.5.21] again, the sales frequency s_t is given by

$$\mathsf{s}_t = -\left(\frac{1-\vartheta_B}{\varphi_B}\right)\left(\frac{1}{1-\bar{\psi}}\right)(\mathsf{x}_t + (1-\sigma)\rho_t)\,. \qquad \text{[A.7.8]}$$

Taking equation [A.5.12] and substituting the expressions for $\mathsf{P}_{S,t} - \mathsf{P}_{N,t}$ and s_t derived above:

$$\mathsf{P}_t - \mathsf{P}_{N,t} = \frac{\psi}{1-\bar{\psi}}(\mathsf{x}_t + (1-\sigma)\rho_t)\,,$$

which uses the formula for ψ derived in Theorem 2. Note that $\mathsf{X}_t - \mathsf{P}_t = (\mathsf{P}_{S,t} - \mathsf{P}_{N,t}) + (\mathsf{P}_{N,t} - \mathsf{P}_t)$, so

$$\mathsf{X}_t - \mathsf{P}_t = \frac{1-\psi}{1-\bar{\psi}}(\mathsf{x}_t + (1-\sigma)\rho_t)\,. \qquad \text{[A.7.9]}$$

Similarly, note that $\mathsf{P}_t - \bar{\mathsf{P}}_t = (\mathsf{P}_t - \mathsf{X}_t) + \mathsf{x}_t$. Then substituting the expression for $\mathsf{X}_t - \mathsf{P}_t$ and simplifying yields:

$$\mathsf{P}_t - \bar{\mathsf{P}}_t = \frac{1-\sigma}{1-\bar{\psi}}(\psi\mathsf{x}_t - (1-\psi)\rho_t)\,. \qquad \text{[A.7.10]}$$

An analogous log-linearization of the cost function in the non-sale sector leads to

$$\mathcal{X}_t = \gamma\mathcal{Q}_t + \mathsf{W}_t,$$

where the assumption about the non-sale sector production function guarantees it has the same elasticity of marginal cost with respect to output as in the sale sector. Note that $\mathcal{Q}_t = \mathcal{Y}_t$ in the non-sale sector since all output in that sector is sold at the same price in the steady state. The derivation of the link between Y_t and Q_t in the sale sector continues to hold subject to P_t being the price level for the sale sector alone:

$$\mathsf{Y}_t = \mathsf{Q}_t + \delta(\mathsf{X}_t - \mathsf{P}_t).$$

Hence the marginal cost differential between the two sectors is

$$\mathcal{X}_t - \mathsf{X}_t = \gamma\left((\mathcal{Y}_t - \mathsf{Y}_t) + \delta(\mathsf{X}_t - \mathsf{P}_t)\right). \qquad \text{[A.7.11]}$$

Using the demand function [A.7.5] and the aggregate price index [A.7.2], relative demand is given by

$$\mathcal{Y}_t - \mathsf{Y}_t = \frac{\epsilon}{1-\sigma}(\mathsf{P}_t - \bar{\mathsf{P}}_t). \qquad \text{[A.7.12]}$$

By substituting this into [A.7.11] and using [A.7.9] and [A.7.10], the marginal cost differential is

$$\mathcal{X}_t - \mathsf{X}_t = \frac{\gamma}{1-\bar{\psi}}\left((\epsilon\psi + \delta(1-\psi))\mathsf{x}_t + (1-\psi)(\delta(1-\sigma)-\epsilon)\rho_t\right).$$

Since price-setting behaviour in the non-sale sector is entirely standard, the usual derivation of the New Keynesian Phillips curve from [A.7.3] and [A.7.4] yields

$$\boldsymbol{\Delta}\mathcal{P}_t = \beta\mathbb{E}_t\boldsymbol{\Delta}\mathcal{P}_{t+1} + \frac{\kappa}{1+\xi\gamma}(\mathcal{X}_t - \mathcal{P}_t).$$

Together with [A.5.26], the differential ρ_t between $\mathcal{P}_t$ and $\mathsf{P}_{N,t}$ is determined by the equation:

$$\boldsymbol{\Delta}\rho_t = \beta\mathbb{E}_t\boldsymbol{\Delta}\rho_{t+1} + \frac{\kappa}{1+\xi\gamma}\left((\mathcal{X}_t - \mathsf{X}_t) - \xi\gamma(\mathsf{P}_{S,t} - \mathsf{P}_{N,t}) - \rho_t\right),$$

which is derived by using $\mathsf{X}_t = \mathsf{P}_{S,t}$. Substituting [A.7.7] and [A.7.11] into the above leads to [A.7.1b] after some rearrangement.

To obtain equation [A.7.1c], note that log-linearized aggregate output is $\bar{\mathsf{Y}}_t = \sigma\mathsf{Y}_t + (1-\sigma)\mathcal{Y}_t$, which is equivalent to $\mathsf{Y}_t - \bar{\mathsf{Y}}_t = -(1-\sigma)(\mathcal{Y}_t - \mathsf{Y}_t)$. Using [A.7.12] and substituting the expression for $\mathsf{P}_t - \bar{\mathsf{P}}_t$ from [A.7.10] yields the result.

Equation [A.7.1d] follows from substituting [A.7.9] into $\mathsf{Y}_t = \mathsf{Q}_t + \delta(\mathsf{X}_t - \mathsf{P}_t)$, which is taken from Lemma 4.

By writing equation [A.7.6] as $\bar{\mathsf{P}}_t = \mathsf{P}_{N,t} + \bar{\vartheta}_P(\mathsf{P}_{S,t} - \mathsf{P}_{N,t}) - \bar{\varphi}_P \mathsf{s}_t + (1-\sigma)\rho_t$, substituting in [A.7.7] and [A.7.8], and then taking first differences:

$$\bar{\pi}_t = \pi_{N,t} + \frac{\bar{\psi}}{1-\bar{\psi}} \boldsymbol{\Delta} \mathsf{x}_t + \left(\frac{1-\sigma}{1-\bar{\psi}}\right) \boldsymbol{\Delta} \rho_t. \qquad \text{[A.7.13]}$$

Then combine equation [A.5.26] with [A.7.7] to obtain:

$$\pi_{N,t} = \beta \mathbb{E}_t \pi_{N,t+1} + \frac{\kappa}{1-\bar{\psi}} \left(\mathsf{x}_t + (1-\sigma)\rho_t\right).$$

Using equation [A.7.13] to write down an expression for $\bar{\pi}_t - \beta\mathbb{E}_t\bar{\pi}_{t+1}$ and substituting for $\pi_{N,t} - \beta\mathbb{E}_t\pi_{N,t+1}$ from above yields the Phillips curve [A.7.1a].

Note that the choice of ξ (which equalizes the average markups in the two sectors) and the production function $\mathfrak{F}(\mathcal{H})$ in the non-sale sector imply that $Y = \mathcal{Y}$, and hence $Q/\mathcal{Q} = \varDelta$. Since the production functions in the two sectors are related by $\mathfrak{F}(\mathcal{H}) = \varDelta\mathcal{F}(\varDelta^{-1}\mathcal{H})$, it follows that $H/\mathcal{H} = \varDelta$. This means that the total labour usage equation $\bar{H}_t = \sigma H_t + (1-\sigma)\mathcal{H}_t$ is log linearized as follows:

$$\bar{\mathsf{H}}_t = \left(\frac{\sigma}{\sigma + (1-\sigma)\varDelta}\right) \mathsf{H}_t + \left(\frac{(1-\sigma)\varDelta}{\sigma + (1-\sigma)\varDelta}\right) \mathcal{H}_t.$$

The log-linearized production functions are the same in the two sectors, so $\mathsf{Q}_t = \alpha\mathsf{H}_t$ and $\mathcal{Q}_t = \alpha\mathcal{H}_t$. By substituting these into the above equation:

$$\bar{\mathsf{H}}_t = \frac{1}{\alpha}\left(\left(\frac{\sigma}{\sigma + (1-\sigma)\varDelta}\right) \mathsf{Q}_t + \left(\frac{(1-\sigma)\varDelta}{\sigma + (1-\sigma)\varDelta}\right) \mathcal{Q}_t\right).$$

By using $\mathcal{Y}_t = \mathcal{Q}_t$ and noting that $\mathcal{Y}_t = (\bar{\mathsf{Y}}_t - \sigma\mathsf{Y}_t)/(1-\sigma)$:

$$\bar{\mathsf{H}}_t = \frac{1}{\alpha} \frac{1}{\sigma + (1-\sigma)\varDelta} \left(\varDelta\bar{\mathsf{Y}}_t + \sigma(\mathsf{Q}_t - \varDelta\mathsf{Y}_t)\right).$$

Substituting this expression into [A.6.9] and rearranging yields [A.7.1g].

A.8 Proof of Proposition 2

(i) Note that Proposition 1 implies μ is only a function of ϵ and η. This is also true of $z = \mathfrak{z}(\mu; \epsilon, \eta)$, as can be seen from equation [A.1.3]. The value of s is then determined by λ (recall that Proposition 1 shows for every $s \in (0,1)$ there is a value of λ generating this s).

Hence, the equilibrium value of ψ can be obtained as a function of s, ϵ and η. This is denoted by $\Psi(s;\epsilon,\eta)$. From [A.5.35], the function is:

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

It has already been shown in Theorem 2 that $\Psi(s;\epsilon,\eta)$ is non-negative. By taking the first derivative with respect to s (holding ϵ and η constant, and hence varying only λ implicitly):

$$\Psi'(s;\epsilon,\eta) = \frac{1-\Theta(\epsilon,\eta;\mu)}{(1+z) + \left((\mu^{1-\epsilon}-1) + (\mu^{1-\eta}-1)z\right)s}\left(1 + \frac{\left((\mu^{1-\epsilon}-1) + (\mu^{1-\eta}-1)z\right)s}{(1+z) + \left((\mu^{1-\epsilon}-1) + (\mu^{1-\eta}-1)z\right)s}\right),$$

which is always strictly positive using the same logic from the proof of Theorem 2. Finally, taking the second derivative yields

$$\Psi''(s;\epsilon,\eta) = -2\frac{(1-\Theta(\epsilon,\eta;\mu))\left((\mu^{1-\epsilon}-1)+(\mu^{1-\eta}-1)z\right)}{(1+z)+((\mu^{1-\epsilon}-1)+(\mu^{1-\eta}-1)z)\,s}\left(1+\frac{\left((\mu^{1-\epsilon}-1)+(\mu^{1-\eta}-1)z\right)s}{(1+z)+((\mu^{1-\epsilon}-1)+(\mu^{1-\eta}-1)z)\,s}\right),$$

which is always strictly negative. This establishes that the function $\Psi(s;\epsilon,\eta)$ is non-negative-valued, strictly increasing and strictly concave.

(ii) The two-sector model's Phillips curve in the general case is given in equation [A.7.1a] following the derivation in appendix A.7. Note that when $\gamma = 0$, the only stable solution of [A.7.1b] is $\rho_t = 0$. By substituting this result into [A.7.1a], it is clear that the resulting equation reduces to the Phillips curve with sales in [4.8] with coefficient $\bar{\psi}$ in place of ψ. Finally, note that $\gamma = 0$ implies that x_t is real marginal cost for both sectors, and hence for the aggregate economy. This completes the proof.