CENTRE for ECONOMIC PERFORMANCE

CEP Discussion Paper No 887

August 2008

Sales and Monetary Policy

Bernardo Guimaraes and Kevin Sheedy

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Abstract
A striking fact about prices is the prevalence of ``sales": large temporary price cuts followed by a return exactly to the former price. This paper builds a macroeconomic model with a rationale for sales based on firms facing consumers with different price sensitivities. Even if firms can vary 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 of prices at the micro level due to sales does not translate into flexibility at the macro level.

Keywords: sales; monetary policy; nominal rigidities
JEL Classification: E3, E5

This paper was produced as part of the Centre's Macro Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Acknowledgements
We thank Andy Levin, Ananth Ramanarayanan, and especially Rachel Ngai for helpful discussions and suggestions, and seminar participants at Bristol University, FGV - Rio, the London School of Economics, the Paris School of Economics, the CEP/ESRC Monetary Policy conference, and the 2008 Texas Monetary Conference for their comments.

Bernardo Guimaraes is an Associate of the Macro Programme at the Centre for Economic Performance and Lecturer (Assistant Professor) in the Department of Economics, London School of Economics. Kevin Sheedy is an Associate of the Macro Programme at the Centre for Economic Performance, London School of Economics. He is also a Lecturer in the Department of Economics, LSE.

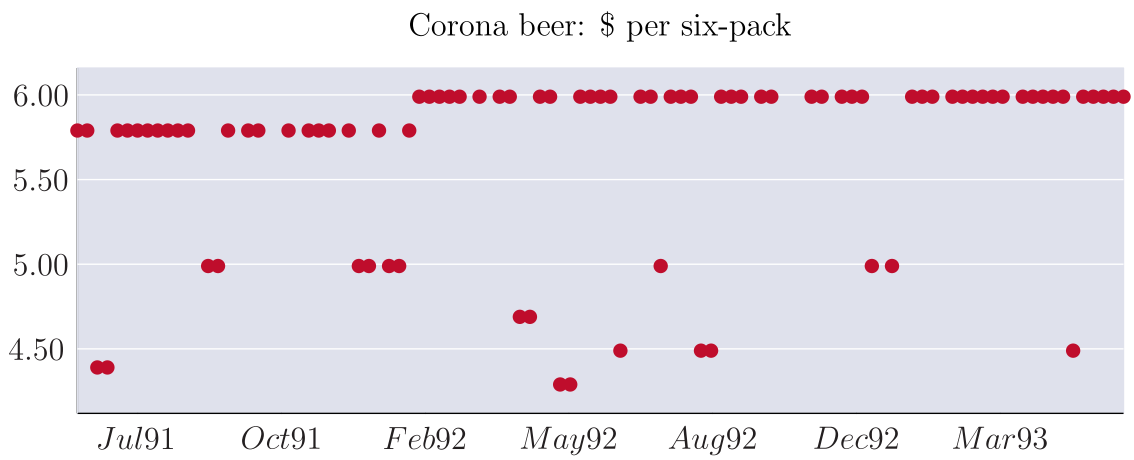
Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

ISBN 978-0-85328-294-5

1 Introduction

A striking fact about prices is that many price changes are "sales": large temporary price cuts followed by a return exactly to the former price. 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 "sale" price are frequent, the "normal" price itself changes far less often. Consequently, empirical estimates of price stickiness significantly diverge when sales are treated differently. Bils and Klenow (2004) count sales as price changes and find that the median duration of a price spell across the whole consumer price index is around 4 months; by disregarding sales, Nakamura and Steinsson (2007) find a median duration of around 9 months.[2] Quantitative models deliver radically different estimates of the real effects of monetary policy depending on which of these two numbers is used. Hence an understanding of sales is needed to answer the question of how large those real effects should be.

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

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

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

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

But this simple framework for analysing sales is not complete. No reason has been presented for why firms would want to choose a pricing strategy in which sales discounts play a significant role. In the IO literature, the most prominent theories of sales are based on incomplete information about prices and consumer preferences. Leading examples include Salop and Stiglitz (1977), Salop and Stiglitz (1982), Varian (1980), Sobel (1984), and Lazear (1986). This paper builds a general-equilibrium macroeconomic model with sales that draws upon the rationale proposed in the IO literature. Despite substantial heterogeneity at the micro level, the model can be easily aggregated to study macroeconomic questions.[3]

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

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

Having built a model of sales, the key question of the paper is: for the purposes of monetary policy analysis, does it matter that the normal price is sticky amidst all the flexibility in sales seen

[3]Another possibility is that the normal price and sale discount would be chosen ex ante to give firms flexibility to respond to shocks if it were too costly to choose a completely new price. This seems implausible, and it is difficult to see why the best way to achieve this insurance would correspond to the observed pattern of sales. Furthermore, to the best of our knowledge, this explanation of sales has not been proposed in the IO literature.

in Figure 1? In sharp contrast to the simple framework discussed first where flexibility in sales implied money was approximately neutral, monetary policy has strong real effects in the IO-based model of sales when the normal price is sticky but sales decisions are completely flexible.

The strong real effects of monetary policy are due to sales being strategic substitutes. Following an expansionary shock to monetary policy, an individual firm has an incentive to decrease its sales, thus increasing the price it sells at on average. However, if all other firms were to follow this course of action then the bargain hunters would be relatively neglected compared to the loyal consumers, increasing the returns to targeting the bargain hunters for any one firm. This leads firms in equilibrium not to adjust sales by much in response to a monetary shock because all firms are trying to respond in the same way. Therefore the aggregate price level adjusts by less and monetary policy shocks have larger real effects.

The model can be calibrated to match some simple facts about sales, and thus assess quantitatively the real effects of monetary policy. If the normal price is completely sticky and sales decisions are completely flexible then the elasticity of output with respect to an unanticipated change in the money supply is around 0.9, and the elasticity of the price level is around 0.1. The flexibility of sales seen at the level of individual prices contributes little toward flexibility of the aggregate price level. Therefore the real effects of monetary policy in a model with a sticky normal price and fully flexible sales are very similar to those found in a model with a single sticky price. These numerical results turn out to be not particularly sensitive to the calibrated parameters.

This analysis treats the normal price as sticky, an assumption in line with the stylized facts as illustrated in Figure 1. A branch of the macroeconomics literature has proposed many justifications for price rigidity, some of which can be applied to explain why the normal price is not continuously readjusted. While these features are not directly incorporated into the model, there are three findings of the model relevant to this issue. First, if a firm had to monitor continuously either its normal price or its sale price, it would choose the latter. Second, deviations of actual from desired normal prices are small even though the model features very large individual price changes, so the losses from failing to make such adjustments are much smaller than might be supposed simply from looking at the size of the average price change. Third, the absolute size of any reoptimization costs needed to justify a constant normal price is much lower than in an otherwise comparable model where the normal price is the only price.

This paper then constructs a dynamic version of the model with sales where firms' normal prices are reoptimized at staggered intervals. This dynamic extension is tractable and an expression for the resulting Phillips curve is derived analytically. It is embedded into a complete dynamic stochastic general equilibrium model and the results of simulations are compared to the same DSGE model incorporating a standard New Keynesian Phillips curve. A quantitative analysis reveals that the difference between the real effects of monetary policy in the two models is small, and thus in line with the findings of the static 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 (2007). In their model, firms face different menu costs depending on whether they make temporary or permanent price changes. Coupled with large but transitory idiosyncratic shocks, this mechanism gives rise to sales in equilibrium.

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

2 Benchmark model

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

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

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

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

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

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

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

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

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

where the price level P is

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

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

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

The utility-maximizing choice of real money balances implies:

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

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

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

Each good is made by a single firm subject to production function $Q = \mathcal{F}(H)$, where Q is output sold at across all points in time and H is hours of labour hired. The production function is differentiable, strictly increasing and concave.

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

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

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

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

$$\mathscr{P} = sp_S\left(\frac{p_S}{P}\right)^{-\varepsilon} Y + (1-s)p_N\left(\frac{p_N}{P}\right)^{-\varepsilon} Y - W\mathcal{F}^{-1}\left(s\left(\frac{p_S}{P}\right)^{-\varepsilon} Y + (1-s)\left(\frac{p_N}{P}\right)^{-\varepsilon} Y\right) \, . \tag{2.5}$$

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

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

Proposition 1 *Given predetermined prices p_N and p_S, and predetermined wage W, if firms choose sales fraction s to maximize profits* [2.5], *as long as $s \in (0,1)$ before and after the monetary shock, firms' output Q is unaffected by the shock.*

Proof The first-order condition with respect to s is

$$\frac{1}{\mathcal{F}'\left(\mathcal{F}^{-1}(Q)\right)} = \frac{1}{W}\left(\frac{p_S^{1-\varepsilon} - p_N^{1-\varepsilon}}{p_S^{-\varepsilon} - p_N^{-\varepsilon}}\right) . \qquad [2.6]$$

Note that the marginal product of labour, and therefore the quantity produced Q, depend only on predetermined variables which are not affected by the realization of the monetary shock. ■

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

The intuition is the following: higher s means that (i) revenues are higher because demand is price elastic; (ii) costs are higher because quantity sold is higher; and (iii) the marginal cost of production increases because the production function is concave. If p_S, p_N and W are fixed, both an increase in the price level P and an increase in output Y multiply the demand for goods at the sale and normal prices by the same factors, and so effects (i) and (ii) exactly cancel out. Therefore profit maximization requires keeping the marginal cost of production, and hence quantity produced Q, constant.

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

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

3 The model of sales

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

3.1 Households

Each household $\imath \in \mathscr{H}$ has the same utility function [2.1] over their composite good $C(\imath)$, real money balances $m(\imath)$ and hours $H(\imath)$ as was used in section 2. The budget constraint [2.2] and aggregate demand [2.4] are also as before. The only change introduced here is in the specification of each household's composite good.

The utility-maximizing trade-off between consumption and hours is given by

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

making use of the first-order condition [2.3] for optimal real balances, which implies $C(\imath) = m(\imath)$.

3.2 Composite goods

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

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

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

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

where $c(\tau, b)$ is the household's consumption of brand b of product type τ. It is assumed that $\eta > \epsilon > 1$, so bargain hunters are more likely to substitute between different brands than households are to substitute between different product types. Households have a zero elasticity of substitution between brands of product types for which they are loyal to a particular brand.

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

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

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

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

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

The expenditure-minimizing demand functions are:

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

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

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

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

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

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

Consequently, the loyal set Λ and the set of types $\mathscr{T}\backslash\Lambda$ for which a household is a bargain hunter have measures λ and $1-\lambda$ respectively for all households. It is assumed that $0<\lambda<1$, so all households are loyal to a brand for some product types and bargain hunters for others. When this paper refers to consumers as either loyal or bargain hunters, note that this designation is for a specific product type only.

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

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

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

3.3 Firms

Each brand b of each product type τ is owned and produced by a single firm, indexed by $(\tau,b)\in(\mathscr{T}\times\mathscr{B})$. All firms have the same production technology. With H hours of labour, a firm can produce physical output Q of its good according to the production function

$$Q=\mathcal{F}(H)\,, \qquad [3.7]$$

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

$$\mathscr{C}(Q;W)=W\mathcal{F}^{-1}(Q)\,. \qquad [3.8]$$

This cost function is differentiable, strictly increasing and convex in Q and satisfies $\mathscr{C}(0;W)=0$ as a result of the properties of the production function [3.7].

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

$$\mathcal{E}(\jmath) \equiv \int_{\mathscr{H}(\jmath)} P(\imath)^{\epsilon} C(\imath) d\imath \ . \tag{3.9}$$

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

$$\mathscr{D}(p; P_B, \mathcal{E}) = (\lambda + (1-\lambda) v(p; P_B))\, p^{-\epsilon} \mathcal{E} \ , \quad \text{where} \ \ v(p; P_B) \equiv \left(\frac{p}{P_B}\right)^{-(\eta-\epsilon)} \ , \tag{3.10}$$

at a moment with bargain hunters' price index P_B for brands of the same type, and an average household expenditure level [3.9] equal to $\mathcal{E}$. Demand is specified in terms of the function $v(p; P_B)$, referred to as the *purchase multiplier*, defined as the ratio of the amount sold to a given measure of bargain hunters relative to the same measure of loyal customers.

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

$$\mathscr{R}(q; P_B, \mathcal{E}) \equiv q \mathscr{D}^{-1}(q; P_B, \mathcal{E}) \ , \quad \text{with} \ \ p = \mathscr{D}^{-1}(q; P_B, \mathcal{E}) \ , \tag{3.11}$$

where $\mathscr{D}^{-1}(q; P_B, \mathcal{E})$ is the inverse demand function corresponding to [3.10].

3.4 Price setting

Now consider the profit-maximization problem for a given firm, which chooses a price for its good at each shopping moment. As will be seen below, the average household expenditure level $\mathcal{E}$, as defined in [3.9], will be the same at all moments. Furthermore, the bargain hunters' price index P_B appearing in demand function [3.10] will be the same for all product types and at all moments. Under these conditions, the profit-maximization problem reduces to choosing a distribution of prices across all possible shopping moments.

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

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

$$\mathscr{P} = \int_p \mathscr{R}\left(\mathscr{D}(p; P_B, \mathcal{E}); P_B, \mathcal{E}\right) dF(p) - \mathscr{C}\left(\int_p \mathscr{D}(p; P_B, \mathcal{E}) dF(p); W\right) \ , \tag{3.12}$$

where $\mathscr{R}(q; P_B, \mathcal{E})$ is total revenue from sales at one moment, defined in [3.11], and $\mathscr{C}(Q; W)$ is the total cost [3.8] of producing the entire output Q of the good sold across all shopping moments.

Consider a discrete distribution of prices $\{p_i\}$ with probabilities $\{\omega_i\}$.[4] The first-order conditions for maximizing [3.12] with respect to prices p_i and probabilities ω_i are

$$\mathscr{R}'\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) = \mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{and} \ , \tag{3.13a}$$

$$\mathscr{R}\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) = \aleph + \mathscr{D}(p_i; P_B, \mathcal{E})\mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{if } \ \omega_i > 0 \ , \tag{3.13b}$$

$$\mathscr{R}\left(\mathscr{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}\right) \leq \aleph + \mathscr{D}(p_i; P_B, \mathcal{E})\mathscr{C}'\left(\sum_j \omega_j \mathscr{D}(p_j; P_B, \mathcal{E}); W\right) \quad \text{if } \ \omega_i = 0 \ , \tag{3.13c}$$

where $\aleph$ is the Lagrangian multiplier attached to the constraint $\sum_i \omega_i = 1$. Equation [3.13a] is the usual marginal revenue equals marginal cost condition, which must hold for any price that receives positive probability. The interpretation of first-order conditions [3.13b] and [3.13c] is discussed later.

3.5 Aggregation

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

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

The common level of consumption $C = Y$ and the common price level P imply that the average household expenditure level [3.9] is the same across all moments, as claimed earlier. This is equal to $\mathcal{E} = P^{\epsilon}Y$ at all moments $\jmath \in \mathscr{M}$. Together with the randomization assumptions for household preferences, this justifies the claim that all firms face the same demand function at all shopping moment, so firms cannot improve upon randomly selecting the moments at which they charge particular prices from their desired distribution.

Finally, note that randomization by firms makes households indifferent between shopping moments, as is assumed.

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

[5] This is true so long as the distribution of firms' price distributions is not different across product types. This will be at all points in this paper, including the dynamic extension of the model.

4 Equilibrium with flexible prices and wages

First consider the equilibrium of the economy when all exogenous variables are constant and prices can be adjusted freely as discussed in section 3.4, and wages adjust to clear the labour market. With a constant money supply, and no shifts in the production function [3.7], the general price level and aggregate output are also constant.

The equilibrium pricing strategies chosen by firms depend on the nature of the demand function $\mathscr{D}(p; P_B, \mathcal{E})$ for a firm's brand at a particular moment, as given in equation [3.10]. What is crucial is that demand comes from two different sources: loyal customers and bargain hunters — and these groups have different price sensitivities. Loyal customers do not want to switch to other brands, so the only margin of substitution they have is shifting expenditure to other types of product in their consumption basket. On the other hand, bargain hunters want to find the brands offering the best deals for a particular product type. The price elasticities for these two groups are ϵ and η respectively, and it makes sense to assume $\eta > \epsilon$. This means that the overall demand curve does not have a constant elasticity: the elasticity changes with the composition of the firm's customers. High prices mean that most bargain hunters will 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 competition among brands for these customers means the price elasticity is high.

These arguments suggest that the price elasticity of demand decreases with price. This is confirmed by differentiating demand function $\mathscr{D}(p; P_B, \mathcal{E})$ in [3.10] to obtain the price elasticity $\zeta(p; P_B)$ (in absolute value):

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

This elasticity is a weighted average of ϵ and η, with the weight on the larger elasticity η increasing with the purchase multiplier $v(p; P_B)$, as defined in [3.10]. The higher is price p, the lower is the purchase multiplier, and the smaller is the price elasticity. Such a change in elasticity is simply a less extreme version of a "kinked" demand curve. For very low prices, the elasticity is approximately constant and equal to η because the bargain hunters dominate; and for very high prices, it is approximately constant and equal to ϵ because only loyal customers remain. In the middle of the demand curve there is a smooth transition between η and ϵ.

As is the case with a kinked demand curve, the varying price elasticity of demand means that the marginal revenue curve is not necessarily downward sloping for all prices, even though demand curve [3.10] is downward sloping everywhere. To see this, note that marginal revenue can be expressed in terms of the price and the price elasticity of demand as follows:

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

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

Proposition 2 *Consider the total revenue function $\mathscr{R}(q; P_B, \mathcal{E})$ defined in [3.11] and suppose $\eta > \epsilon > 1$. Then marginal revenue $\mathscr{R}'(q; P_B, \mathcal{E})$ is non-monotonic if and only if $0 < \lambda < 1$ and*

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

hold, and everywhere downward sloping otherwise.

PROOF See appendix A.2. ■

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

With a non-monotonic marginal revenue curve $\mathscr{R}'(q; P_B, \mathcal{E})$, it is possible that more than one price can be associated with the same level of marginal revenue. First-order condition [3.13a] then suggests firms might want to charge different prices at different shopping moments.

As was discussed in the introduction, an interesting case is where firms find it optimal to choose a distribution with just two prices: a "normal" high price, and a low "sale" price. Denote these two prices respectively by p_N and p_S with $p_N > p_S$, and let $q_N = \mathscr{D}(p_N; P_B, \mathcal{E})$ and $q_S = \mathscr{D}(p_S; P_B; \mathcal{E})$ be the quantities demanded by customers at these prices, clearly with $q_S > q_N$. The fraction of shopping moments when the brand is on sale is denoted by s. If $0 < s < 1$ then both prices must satisfy the first-order conditions [3.13a]–[3.13b]. By eliminating the Lagrangian multiplier $\aleph$ from [3.13b], profit maximization requires:

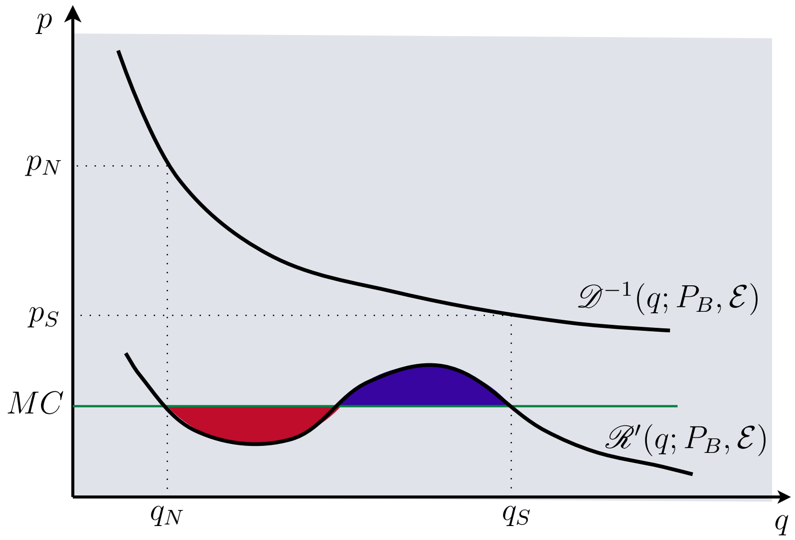
$$\mathscr{R}'(q_N; P_B, \mathcal{E}) = \mathscr{R}'(q_S; P_B, \mathcal{E}) = \frac{\mathscr{R}(q_S; P_B, \mathcal{E}) - \mathscr{R}(q_N; P_B, \mathcal{E})}{q_S - q_N} = \mathscr{C}'\left(sq_S + (1 - s)q_N; W\right) . \qquad [4.4]$$

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

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

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

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



Notes: Schematic representation of demand function [3.10] and marginal revenue function [4.2]. Shown for the case where parameters ϵ and η satisfy [4.3].

The full set of optimality conditions is depicted in Figure 3. The first point to note is that because firms can charge different prices at different moments the total revenue function can be made concave. This raises attainable total revenue in the range between q_N and q_S. The first two optimality conditions in [4.4] require that total revenue 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 the tangent itself. This slope then determines the unique point where marginal cost equals marginal revenue, which yields the total quantity sold and therefore the sales fraction.

The conditions for the existence and uniqueness of this type of two-price equilibrium are given in the following result.

Theorem 1 *Suppose firms choose distributions of prices to maximize profits as given in equation* [3.12].

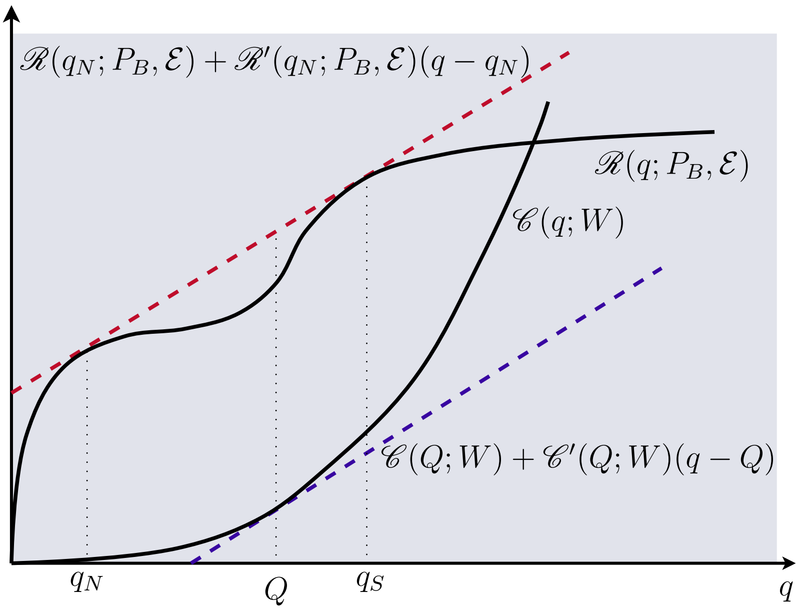
(i) If elasticities ϵ and η are such that the non-monotonicity condition [4.3] *holds then there exist $\underline{\lambda}(\epsilon, \eta)$ and $\overline{\lambda}(\epsilon, \eta)$ such that $0 < \underline{\lambda}(\epsilon, \eta) < \overline{\lambda}(\epsilon, \eta) < 1$, and if $\lambda \in [\underline{\lambda}(\epsilon,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

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

Proof See appendix A.3 ■

This result indicates that two conditions need to be satisfied for two prices to be an equilibrium. First, marginal revenue must be non-monotonic, as has been discussed above and analysed in Proposition 2. Second, there must not be too many loyal consumers, or too many bargain hunters,

Figure 3: *Total revenue and total cost functions*



Notes: Schematic representation of total revenue function $\mathscr{R}(q; P_B, \mathcal{E})$ from [3.11] and total cost function $\mathscr{C}(Q; W)$ from [3.8], shown for parameters ϵ and η satisfying [4.3].

but instead a sufficient mixture of the two. This justifies having a high price designed for the loyal customers, and a low one for the bargain hunters at other moments, even though no actual price discrimination can be practised as it is not possible for firms to distinguish the two types prior to the moment of purchase.

Note that there is no reference to the degree of convexity of the cost function in Theorem 1. What guarantees the existence and uniqueness of the two-price equilibrium for a wide range of parameters, even if marginal cost were constant, is that the actions of other firms affect the total revenue function, in particular the slope of the chord in Figure 3. This interdependence of firms plays a key role throughout the paper and is discussed in full later.

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

Given the stylized facts discussed in the introduction, this paper focuses on the two-price equilibrium. The total physical quantity of output sold by firms is $Q = sq_S + (1-s)q_N$. Let $X \equiv \mathscr{C}'(Q; W)$ denote the associated marginal cost in money terms. Using the relationship between price and marginal revenue in [4.2], the marginal revenue equals marginal cost conditions for each price can be expressed in terms of desired markups on marginal cost:

$$p_i = \mu(p_i; P_B)X \ , \quad \text{with} \ \ \mu(p; P_B) = \frac{\lambda\epsilon + (1-\lambda)\eta v(p; P_B)}{\lambda(\epsilon - 1) + (1-\lambda)(\eta - 1)v(p; P_B)} \ . \qquad [4.5]$$

The desired markups depend on the parameters ϵ, η and λ, and the purchase multiplier $v(p; P_B)$ from [3.10]. Let $v_S \equiv v(p_S; P_B)$ and $v_N \equiv v(p_N; P_B)$ denote the two purchase multipliers, and the $\mu_S \equiv \mu(p_S; P_B)$ and $\mu_N \equiv \mu(p_N; P_B)$ the associated markups. Hence,

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

By using the demand function from [3.10], the first-order condition in [4.4] linking average extra revenue to marginal cost can be expressed as:

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

Given that a fraction s of all prices are set at p_S and the remaining $1-s$ are set to p_N, equation [3.4] implies the price index P_B for the bargain hunters is

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

In finding the equilibrium, the model has a convenient block-recursive structure, or in other words, the microeconomic aspects can be characterized independently of the macroeconomic side, which is then determined afterwards. The key micro variables are the sales fraction s, the *relative* markup $\mu \equiv \mu_S/\mu_N$, and the *relative* quantity sold at the sale price compared to at the normal price, denoted by $\chi \equiv q_S/q_N$. Using the demand function [3.10], first-order condition [4.7] gives the relationship between quantity ratio χ and markups μ_S and μ_N.

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

The following proposition verifies the block-recursive structure and derives some comparative statics

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

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

(ii) For given values of ϵ and η:

$$\frac{\partial \mu}{\partial \lambda} = 0 \ , \quad \frac{\partial \mu_S}{\partial \lambda} = 0 \ , \quad \frac{\partial \mu_N}{\partial \lambda} = 0 \ , \quad \frac{\partial \chi}{\partial \lambda} = 0 \ , \quad \frac{\partial s}{\partial \lambda} < 0 \ .$$

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

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

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

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

where $\eta^*(\epsilon) \equiv (3\epsilon - 1) + 2\sqrt{2\epsilon(\epsilon - 1)}$ is the lower bound for η that ensures non-monotonicity, according to [4.3] in Proposition 2.

PROOF See appendix A.4. ■

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

Given the purchase multipliers and markups, finding the equilibrium values of the other endogenous variables is straightforward. The general price level P is obtained by combining [3.3] and demand functions [3.4], and making use of the definition of the purchase multipliers in [3.10]:

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

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

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

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

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

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

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

Using the production function [3.7], the cost function [3.8], and the labour supply function [3.1], a relationship is obtained between real marginal cost x and output Y:

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

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

5 Monetary shocks in a model of sales

The benchmark model of section 2 analysed the effect of a monetary shock with predetermined prices p_S and p_N and wage W, but crucially, the reason why firms started with two prices rather than just one was left unexplained. The sales model introduced in section 3 provides precisely such a reason, and this section performs a similar experiment when sales are flexible.[6]

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

The freedom to adjust sales, but not the normal price p_N, means that of the first-order conditions in [4.4], only the second and third equalities holds:

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

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

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

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

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

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

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

$$\mathsf{X} = \mathsf{P}_B \; ,$$

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

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

Proof See appendix A.6. ■

The first part of the theorem makes sales strategic substitutes. As other firms cut back on sales either by reducing s or increasing p_S, the bargain hunters' price index P_B increases. Theorem 1 shows this leads a given firm optimally to raise total quantity sold to the point where marginal cost X has risen one-for-one with P_B. As the normal price is not adjusted, the increase in quantity sold is brought about by an increase in sales.

The problem of choosing the profit-maximizing sales adjustment is essentially one of a firm deciding how much to target its loyal customers versus the bargain hunters. Because competition for the bargain hunters is more intense than for loyal customers, the incentive to target them 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.

The option of adjusting the fraction of sales s was also open to firms in the benchmark model, but here the use of this margin has important implications for the competition among firms. This can be seen algebraically by substituting the demand function and purchase multipliers from [3.10], and the cost function [3.8] into the first part of [5.1]:

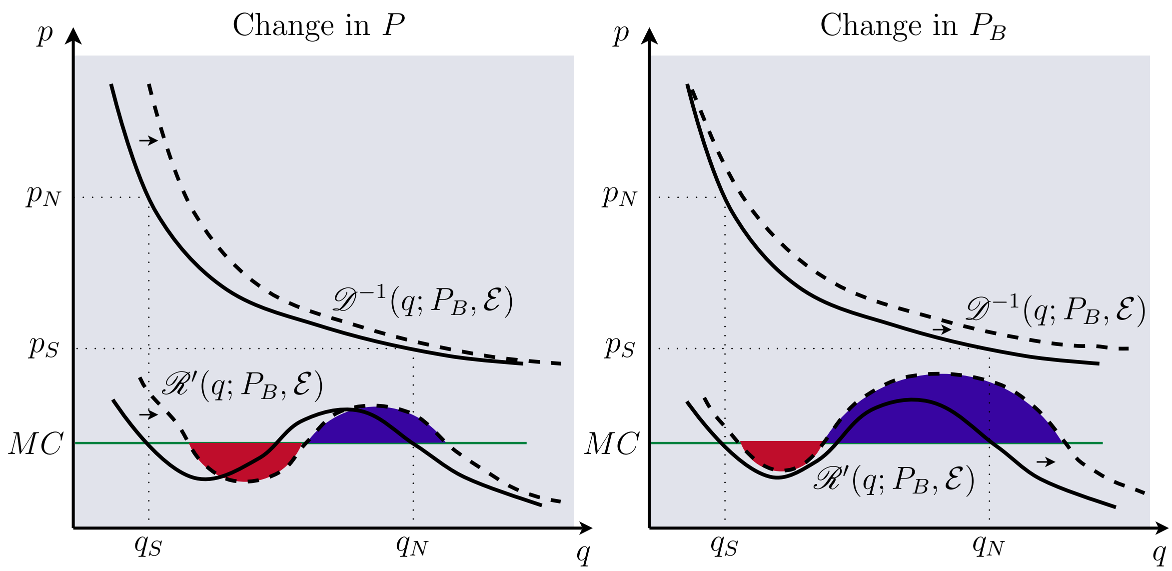
$$\frac{1}{\mathcal{F}'\left(\mathcal{F}^{-1}(Q)\right)} = \frac{1}{W}\frac{p_S^{1-\epsilon}\left(\lambda + (1-\lambda)\left(\frac{p_S}{P_B}\right)^{-(\eta-\epsilon)}\right) - p_N^{1-\epsilon}\left(\lambda + (1-\lambda)\left(\frac{p_N}{P_B}\right)^{-(\eta-\epsilon)}\right)}{p_S^{-\epsilon}\left(\lambda + (1-\lambda)\left(\frac{p_S}{P_B}\right)^{-(\eta-\epsilon)}\right) - p_N^{-\epsilon}\left(\lambda + (1-\lambda)\left(\frac{p_N}{P_B}\right)^{-(\eta-\epsilon)}\right)} \; . \qquad [5.2]$$

This difference between the models can be understood by looking at the respective first-order conditions [2.6] and [5.2] for the sales fraction s. The key difference is the presence of P_B in the model with sales. The terms involving P_B reflect the different degrees of competition for loyal customers and bargain hunters.

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

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

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



Notes: Schematic representation of shifts of demand and marginal revenue functions [3.10] and [4.2]. The price level P affects demand through $\mathcal{E} = P^{\epsilon}Y$ according to [3.9].

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

The previous discussion explains why there must be a positive relationship between P_B and X, but the result of Theorem 2 is stronger: the elasticity must be unity. This follows from some elementary properties of the profit function. Define $\mathscr{P}(p; P_B, X, P, Y)$ be the level of profits at the margin from selling at price p at one shopping moment:

$$\mathscr{P}(p; P_B, X, P, Y) = (p - X)\mathscr{D}(p; P_B, P^{\epsilon}Y) , \qquad [5.3]$$

where $\mathcal{E} = P^{\epsilon}Y$ has been used, in accordance with [3.9]. The first-order condition for the optimal sales fraction is $\wp(p_S, p_N, P_B, X, P, Y) = 1$, where

$$\wp(p_S, p_N, P_B, X, P, Y) \equiv \frac{\mathscr{P}(p_S; P_B, X, P, Y)}{\mathscr{P}(p_N; P_B, X, P, Y)} \,, \tag{5.4}$$

is the ratio of profits from selling at the sale price to profits from the normal price. The demand function is homogeneous of degree zero in all prices, and so the profit function [5.3] must be homogeneous of degree one in p, P_B, P and X, and therefore the profit ratio $\wp(p_S, p_N, P_B, X, P, Y)$ is homogeneous of degree zero in p_S, p_N, P_B, P and X. The form of the demand function [3.10] implies that P and Y proportionately affect profits at both prices and thus have no influence on relative profits, so $\wp(p_S, p_N, P_B, X, P, Y) = \wp(p_S, p_N, P_B, X, 1, 1)$ for all P and Y. Consequently, relative profits [5.4] must be homogeneous of degree zero in p_S, p_N, P_B and X alone. Since p_S and the predetermined value of p_N are chosen optimally, neither p_S nor p_N has a first-order effect on either profits or relative profits. Therefore, relative profits $\wp(p_S, p_N, P_B, X, 1, 1)$ must be locally homogeneous of degree zero in just P_B and X. Hence to ensure relative profits remain equal to one, P_B and X must change by the same proportion.

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

5.1 Calibration

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

There is a growing empirical literature examining price adjustment patterns at the microeconomic level. This literature provides information about the markup ratio and the sales fraction. The benchmark values of these variables are taken from Nakamura and Steinsson (2007). Their study draws on individual price data from the BLS CPI research database, which covers approximately 70% of U.S. consumer expenditure. They report that the fraction of price quotes that are sales (weighted by expenditure) is 7.4%. They also report that the median difference between $\log(p_S)$ and $\log(p_N)$ is 0.295, which yields $\mu = 0.745$.

In the retail and marketing literature, there has for a long time been a discussion of the effects of price promotions on demand. This literature provides information about the quantity ratio.

However, papers in this literature report a range of estimates conditional on factors other than price that affect the impact of the price promotion, for example, advertising. The benchmark value of the quantity ratio is obtained from the study by Chakravarthi, Neslin and Sen (1996). Their results are based on scanner data from a large number of U.S. supermarkets. According to the elasticities they report, a price cut of the size consistent with the markup ratio taken from Nakamura and Steinsson (2007) implies a quantity ratio of between approximately 4 and 6 if the retailer draws the price cut to the attention of customers. The benchmark number used here is the simple average of the two, so $\chi = 5$.

The three facts about sales, summarized in Table 1, are then used to find matching values of the unknown parameters. This exercise first requires finding the equilibrium of the economy for the variables μ, χ and s. Proposition 3 shows that these depend only on the parameters ϵ, η and λ. Lemma 3 in the appendix shows how μ and χ are determined as functions of ϵ and η. Then equation [A.3.6] in the proof of Theorem 1 determines s as a function of all three parameters.

Table 1: *Stylized facts about sales*

Description	Parameter	Value
Ratio of sale markup to normal markup (μ_S/μ_N)	μ	$0.745^{\dagger}$
Ratio of quantity sold at sale price to normal price (q_S/q_N)	χ	$5^{\ddagger}$
Fraction of goods sold at sale price	s	$0.074^{\dagger}$

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

Given this solution method, parameters matching the stylized facts were found using the Nelder-Mead simplex algorithm. An extensive grid search over ϵ and η was used to verify that these are the only parameters matching μ and χ. Proposition 3 demonstrates that given ϵ and η, there is always one and only one λ value matching the sales fraction s. The results of this exercise are shown in Table 2.

Table 2: *Parameters matching stylized facts about sales*

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

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

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

$$\mathcal{F}(H) = AH^{\alpha} \,, \tag{5.5}$$

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

5.2 Results

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

The results for the benchmark calibration are examined first. Using the parameters from Table 2 and $\alpha = 2/3$ the elasticities are:

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

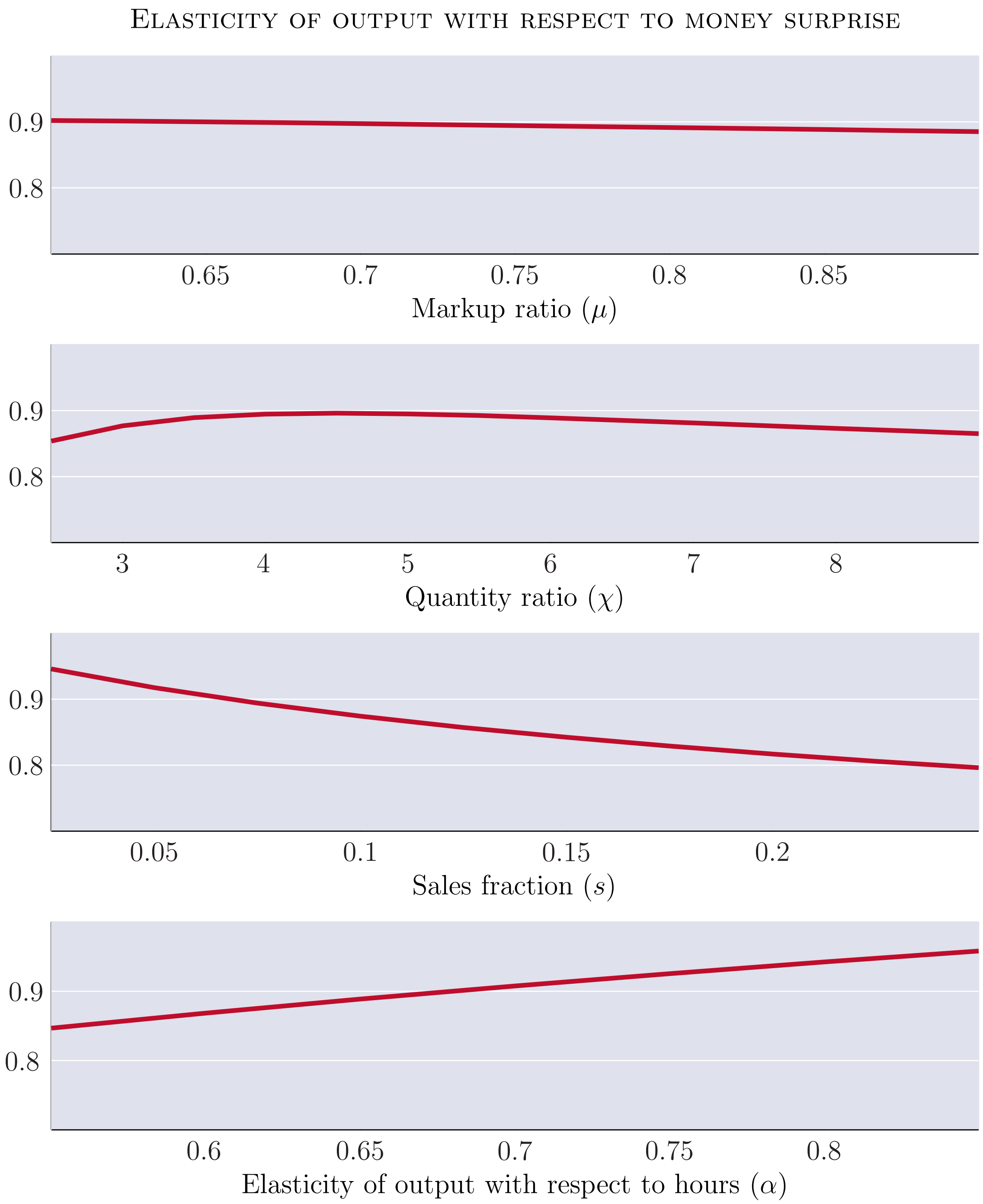
For a 1% surprise increase in the money supply, output rises by 0.895%. The results are not very sensitive to the stylized facts about sales used to calibrate the model. A sensitivity analysis is shown in Figure 5. Of the three targets, the effects of monetary policy are most sensitive to the sales fraction s. In the range of empirically plausible s values (5% – 15%), monetary policy has substantial real effects: the elasticity varies between 0.84 and 0.92.

The quantity ratio χ is the target for which the literature yields widest range of estimates. But nonetheless, varying χ from 3 to 8 implies that the elasticity lies between 0.87 and 0.90. Finally, the target value of the markup ratio μ makes essentially no difference to the results.

These findings are in sharp contrast to the results of the experiment performed using the benchmark model of section 2, where there was no rationale for having a two-price distribution. In the new model, consumer preferences are such that sales are an equilibrium phenomenon. In both cases, firms have an incentive to adjust the fraction of sales following a monetary shock. But the consumer preferences introduced to explain sales also give rise to strategic substitutability in the sales decision. Strategic substitutability is so strong that flexibility in sales brings very little flexibility to the aggregate price level.

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

Figure 5: *Sensitivity analysis*



Notes: The results are obtained by fixing two of the three targets at their benchmark values as given in Table 1 and choosing matching values of the parameters ϵ, η and λ as explained in section 5.1.

5.3 Justification for the "sticky" normal price

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

Recent micro evidence on price setting has highlighted the relative stickiness of so-called "reference" prices (Eichenbaum, Jaimovich and Rebelo, 2008), which correspond to "normal" prices in this paper. The model developed in this paper is consistent with the finding of sticky reference prices, and moreover, in the setting of the model, it makes sense for the normal price to be relatively sticky. This section develops three arguments in support of this claim: (i) in the context of the model, the extra gains from adjusting the normal price after firms have optimally chosen s and p_S are only 14% of the corresponding gains in a standard sticky price model; (ii) adjusting the sales fraction reaps most of the benefits of price adjustment; and (iii) after adjusting s, the gains from repeatedly adjusting the normal price (which is used 92.6% of the time in the baseline calibration) are actually very close to the gains obtained by changing the sale price only when the good is on sale (which occurs 7.4% of the time).

These results build on the following proposition:

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

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

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

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

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

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

Proof See appendix A.8. ■

Corollary *If p_S is optimally chosen, so $\mathsf{p}_S = \mathsf{p}_S^* = \mathsf{X}$ then the gain from adjusting p_N to p_N^* is:*

$$\text{Gain} = \frac{1}{2}(1-\bar{s})\frac{\bar{q}_N}{\bar{Q}}\bar{x}\left(\bar{\zeta}_N - \frac{(\eta-\epsilon)^2\lambda(1-\lambda)\bar{v}_N(\bar{\mu}_N-1)}{(\lambda+(1-\lambda)\bar{v}_N)^2}\right)(\mathsf{p}_N-\mathsf{X})^2 \quad [5.7]$$

□

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

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

To see how this compares with standard analyses of menu costs and sticky prices, the expression for the gain in profits can be contrasted with that which obtains in a model with entirely standard Dixit-Stiglitz preferences, and thus one price in equilibrium, but which is otherwise identical. As is demonstrated in appendix A.11, the gain in profits from price adjustment (again expressed as a fraction of steady-state total revenue) is:

$$\text{Gain} = \frac{1}{2}\varepsilon(1+\varepsilon\gamma)\bar{x}\left(\mathsf{p} - \left(\mathsf{P} + \frac{1}{1+\varepsilon\gamma}\mathsf{x}\right)\right)^2 , \qquad [5.8]$$

where ε is the constant price elasticity of demand and γ is the elasticity of marginal cost with respect to quantity produced. With the production function [5.5], $\gamma = (1-\alpha)/\alpha$.

When comparing the gain from adjusting only s with the gain from adjusting price in a standard one-price model, there are two crucial differences between [5.6] and [5.8]. Quantitatively, the most important difference between the profit gains corresponds to the term $1+\varepsilon\gamma$, which appears only in [5.8]. This represents the gains from selling the optimal quantity, which in a standard model can only be achieved through a price change. But as Proposition 4 shows, the gains from producing the optimal quantity automatically accrue when firms are free to choose their desired sales fraction.

The second reason for a smaller gain relative to a standard model from adjusting the normal and sale prices is that with a demand function consistent with sales in equilibrium, the price elasticity is decreasing in price, thus if prices are too high the desired markup also increases, and vice versa if prices are too low. The bracketed terms in [5.6] multiplying the deviations of prices are smaller than the price elasticities of demand $\bar{\zeta}_N$ and $\bar{\zeta}_S$, since the terms being subtracted are unambiguously positive. In contrast, in [5.8], the deviation is multiplied simply by the price elasticity ε.

In the sales model, the sizes of desired adjustment of the normal price being contemplated by firms in response to monetary shocks are significantly smaller than the changes observed in individual prices, which mostly correspond to shifts between the normal and sale prices. Therefore, large price changes are observed, but full reoptimization of prices requires only small adjustments, and so only small losses are incurred if firms fail to make these desired changes. This means that reoptimization of the normal price falls exactly within the remit of the literature in macroeconomics which seeks to justify why firms do not always make small price changes, such as Mankiw (1985) and Akerlof and Yellen (1985).[8]

The gains from adjusting prices in the sales model are compared with those in a one-price model where firms are faced with the same shocks, even though a one-price model would require much larger shocks to generate the magnitude of observed price changes. The difference in the size of menu costs needed to rule out a flexible-price equilibrium can be computed using the calibration

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

from section 5.1 and the expressions [5.6] and [5.8]. In the latter, the constant price elasticity ε is chosen to imply a markup equal to the average markup found in the calibrated model with sales.[9] With an elasticity of output with respect to hours of 2/3, the implied γ is 0.5. In the sales model, the menu cost needed to dissuade a given firm from changing both its sale and normal prices is only 27% of the menu cost that justifies not changing price in the standard model.

The same exercise can be performed assuming that p_S and s are optimally chosen, which corresponds to comparing the gains implied by [5.6] and [5.7]. This exercise reveals that the menu cost needed to dissuade a firm from adjusting p_N is only 14% of the menu cost needed to deter price adjustment in a standard model. This constitutes approximately half of the total gains from changing both p_S and p_N, which shows that the coefficients of the deviations of p_S and p_N are approximately the same.

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

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

6 Flexible wages

This section considers the model of sales with fully flexible wages. In this case, the first-order condition of households for labour supply [3.1] holds at all times. Since all households face the same prices, this implies:

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

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

Obtaining the effects of a monetary shock now requires calibrating the utility function. The main issue is to avoid the counterfactual prediction that the real wage fluctuates by more than output. Thus a lower bound for the real effects of monetary policy is found by choosing a utility function that implies the real wage and output move one for one. This is done by adopting the conventional

[9]The bracketed terms are multiplied by $\bar{s}\bar{q}_S/\bar{Q}$ and $(1 - \bar{s})\bar{q}_N/\bar{Q}$, which weight them according to the relative quantities sold. In the sales model, $\mu_S = 1.09$, $\mu_N = 1.47$, and $\bar{s}\bar{q}_S/\bar{Q} = 0.28$, which yields an average markup of 1.36. With Dixit-Stiglitz preferences, the optimal markup is $\varepsilon/(\varepsilon - 1)$, so $\varepsilon = 3.77$.

specification of log utility in consumption and linear disutility in hours worked:[10]

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

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

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

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

7 Dynamics

This section extends the previous analysis to a dynamic environment, where the normal price is adjusted, but not continuously so. There is a tractable dynamic version of the sales model and this section derives the resulting Phillips curve, which is easily embedded into any DSGE framework. While the presence of sales in the model adds an extra effect when compared to the standard New Keynesian Phillips curve, quantitatively the difference turns out not to be large. The conclusions are thus in line with the findings of section 5.

7.1 Staggered adjustment of the normal price

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

The assumption of Calvo pricing for the normal price is made only for simplicity. Of course the choice of an alternative model of price stickiness, for example, state-dependent pricing, would affect the results in its own right. But there is no obvious reason to believe that the interaction of different models with firms' optimal choice of sales will significantly affect the results (unless those models yield the counterfactual prediction that the price p_N is continuously adjusted, thus making the sales margin redundant). This is because Proposition 4 implies that profit-maximizing prices are a function only of the aggregate state of the economy, and thus independent of the distribution

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

of prices. Furthermore, a firm's optimal sales decisions depend only on its own normal price and the aggregate state of the economy.

In each 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 optimal price it selects is referred to as its reset price, and is denoted by $R_{N,t}$. Asset markets are assumed to be complete: $\mathscr{A}_{t+\ell|t}$ denotes the asset-pricing kernel for state-contingent monetary payoffs (relative to the conditional probability of each state occurring). The optimal sales decisions will in principle depend on the firm's normal price, and so on its last adjustment time. Denote by $s_{\ell,t}$ and $p_{S,\ell,t}$ the optimal sales fraction and sale price for a firm at time t that last changed its normal price ℓ periods ago. The reset price $R_{N,t}$ is chosen to maximize:

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

The first-order condition for the optimal reset price is given by:

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

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

Note that the optimal reset price is identical for all firms that change their normal price at the same time. Hence the expression for the aggregate price index P_t is

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

and the bargain hunters' price index $P_{B,t}$ is defined accordingly.

The sales fraction $s_{\ell,t}$ and sale price $p_{S,\ell,t}$ are determined as in [5.1]:

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

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

7.2 A Phillips curve with sales

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

Theorem 3 *Suppose firms determine optimal reset price $R_{N,t}$ according to equation [7.2] and their* *optimal sales fractions and sale prices using* [7.4]*. Let* $\pi_t \equiv P_t/P_{t-1}$ *be the inflation rate for the aggregate price index* [7.3]*. Log-linearizing around the flexible-price equilibrium of* section 4 *with zero inflation yields an optimal reset price satisfying*

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

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

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

where the parameter ψ *is defined as follows:*

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

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

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

Proof See appendix A.9. ■

Notice first that the Phillips curve with sales [7.5] reduces to the standard New Keynesian Phillips curve in the case that $\psi = 0$, but ψ is always positive in the model with sales. When $\psi \to 1$ the economy converges to the case of price flexibility. The condition $\psi < 1$ is equivalent to:

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

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

The effect of a positive value of ψ is to increase the response of inflation to real marginal cost when compared to the standard New Keynesian Phillips curve. This is best seen by looking at the solved-forwards version of the Phillips curve in [7.6], where there are two differences. The first is scaling of the coefficient multiplying expected real marginal costs, which is isomorphic to an increase in the probability of price adjustment. The second is the term driven by changes in real marginal cost, which is linked to the possibility of varying sales each period. It is subsequently shown how this will affect the dynamics of output and prices.

7.3 A DSGE model with sales

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

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

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

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

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

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

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

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

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

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

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

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

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

As in Erceg, Henderson and Levin (2000), firms hire differentiated types of labour. So hours H in the production function [3.7] is now a composite labour input defined by the following Dixit-Stiglitz aggregator

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

where $H(\imath)$ is hours supplied by household $\imath \in \mathscr{H}$ to a given firm, and ς is the elasticity of substitution between labour types. It is assumed that $\varsigma > 1$, and that firms are price takers in the markets for labour inputs. The money wage received by labour input $\imath$ is $W(\imath)$. The minimum cost of hiring one unit of the composite labour input H is denoted by W, and this is the relevant wage index in firms' cost function [3.8]. This wage index is given by

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

and the cost-minimizing labour demand functions are

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

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

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

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

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

Given that all households who change their wage at the same time pick the same reset wage, the wage index W_t in [7.12] evolves according to:

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

7.4 Dynamic calibration

This section presents the calibration of the DSGE model described above.

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

Table 3: *Dynamic calibration*

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

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

The elasticity of output with respect to hours α is chosen to match a labour share of 2/3. With the specification [5.5] of the production function, this implies an elasticity of marginal cost with respect to output of $\gamma = (1-\alpha)/\alpha$. So $\alpha = 2/3$ yields $\gamma = 0.5$. The elasticity of substitution between labour inputs ς is taken from Christiano, Eichenbaum and Evans (2005). The probability ϕ_p of stickiness of the normal price is set to match a price-spell duration of 9 months, which is taken from Nakamura and Steinsson (2007). The same number is used for the probability of wage stickiness ϕ_w, as evidence shows that most, but not all, wages are adjusted annually.

All the calibrated parameters are listed in Table 3.

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

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

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

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

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

where φ_π is the interest-rate response to inflation, φ_y is the response to output (or the output gap), and φ_i is the interest-rate smoothing parameter. The Taylor rule parameters are taken from the baseline estimates of the Volcker–Greenspan period in Clarida, Galí and Gertler (1998), which is 1979:8–1996:12 (the same sample period as was used for the money-supply growth specification).

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

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

Notes: Monthly calibration.
† *Source*: Authors' calculations using data on M1 for the period 1979:8–1996:12. Series `M1SL` from Federal Reserve Economic Data (`http://research.stlouisfed.org/fred2`).
‡ *Source*: Clarida, Galí and Gertler (1998), converted from estimates based on quarterly data to a monthly calibration.

7.5 Dynamic simulations

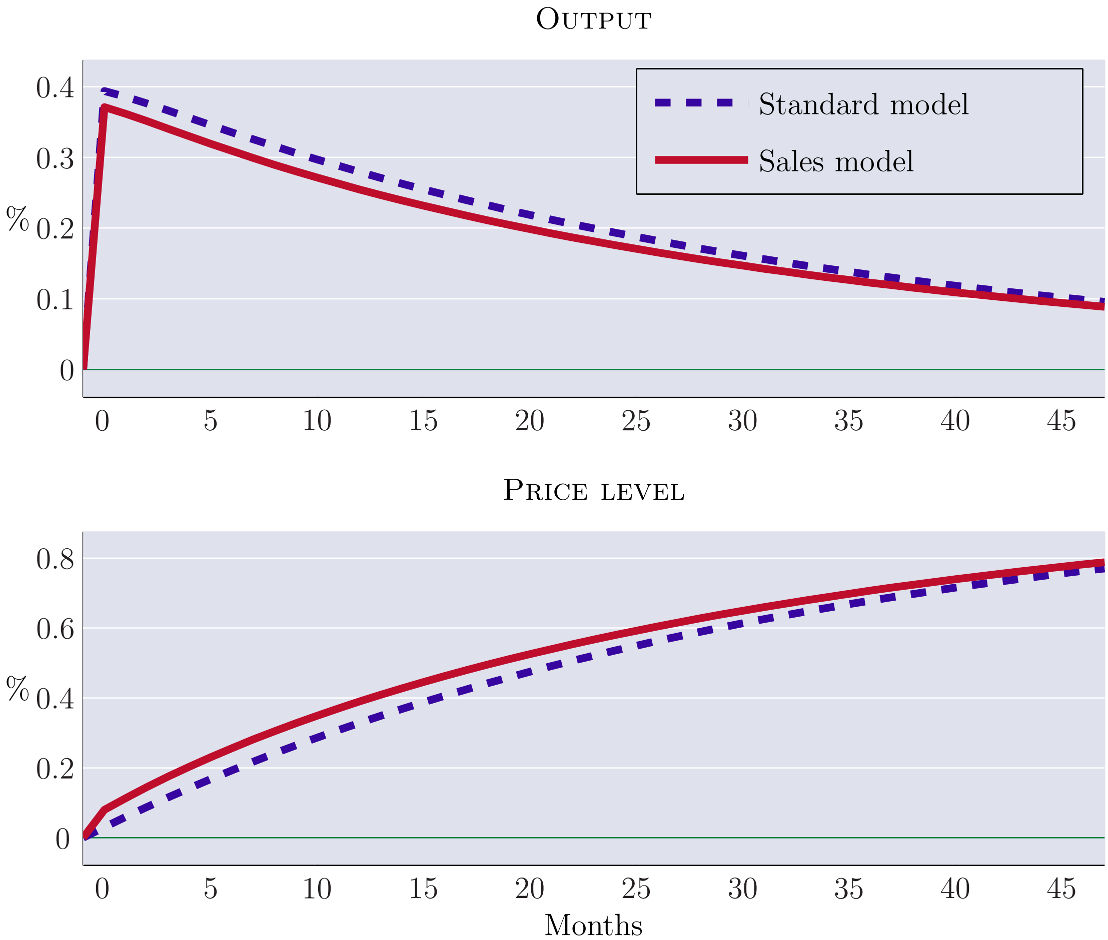
This section calculates the impulse responses of output and the price level to monetary policy shocks in the DSGE model with sales described in section 7.1 and section 7.3. These are compared to the corresponding impulse responses in a standard DSGE model, that is, one where consumers have regular Dixit-Stiglitz preferences and thus firms employ a one-price strategy, and price adjustment times are staggered according to the Calvo model. With Calvo pricing, a standard New Keynesian Phillips curve is obtained.[11] The latter model is otherwise identical to the DSGE model with sales.

[11]See appendix A.10 for details.

The calibrated parameters of the DSGE model with sales are given in Table 2 and Table 3. For the standard model, the same parameter values from Table 3 are used, with the probability of price stickiness applying to a firm's single price, rather than its normal price in the sales model. In place of parameters ϵ, η and λ, the standard model requires only a calibration of the constant price elasticity of demand ε. This is chosen to match the average markup found in the sales model.[12]

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

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



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

Figure 6 plots the impulse responses when money growth follows an AR(1) process in both the sales model and the standard model. As in the static analysis of section 5, 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 the full flexibility of sales. The ratio of the cumulative deviations of output in the two models is 0.929. The response of prices in the sales model shows a small jump immediately

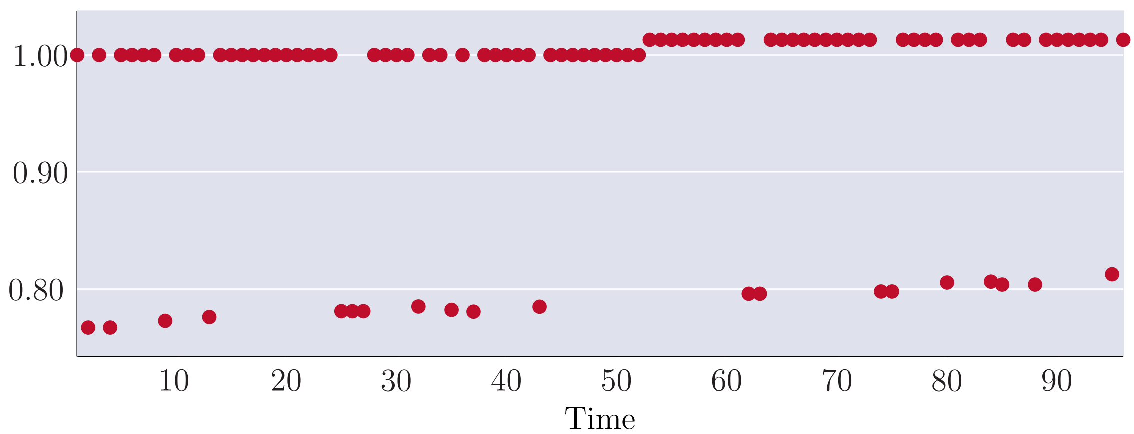
[12]See footnote 9 for details. The calculations lead to $\varepsilon = 3.77$.

after the shock. This corresponds to the term Δx_t in the Phillips curve [7.6].

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

Figure 7 shows an example of a price path in the model with sales using the baseline calibration. The underlying stochastic process for the money supply is a random walk with drift. The behaviour depicted is qualitatively and quantitatively consistent with real-world examples of prices, even without any idiosyncratic shocks in the model.[13]

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



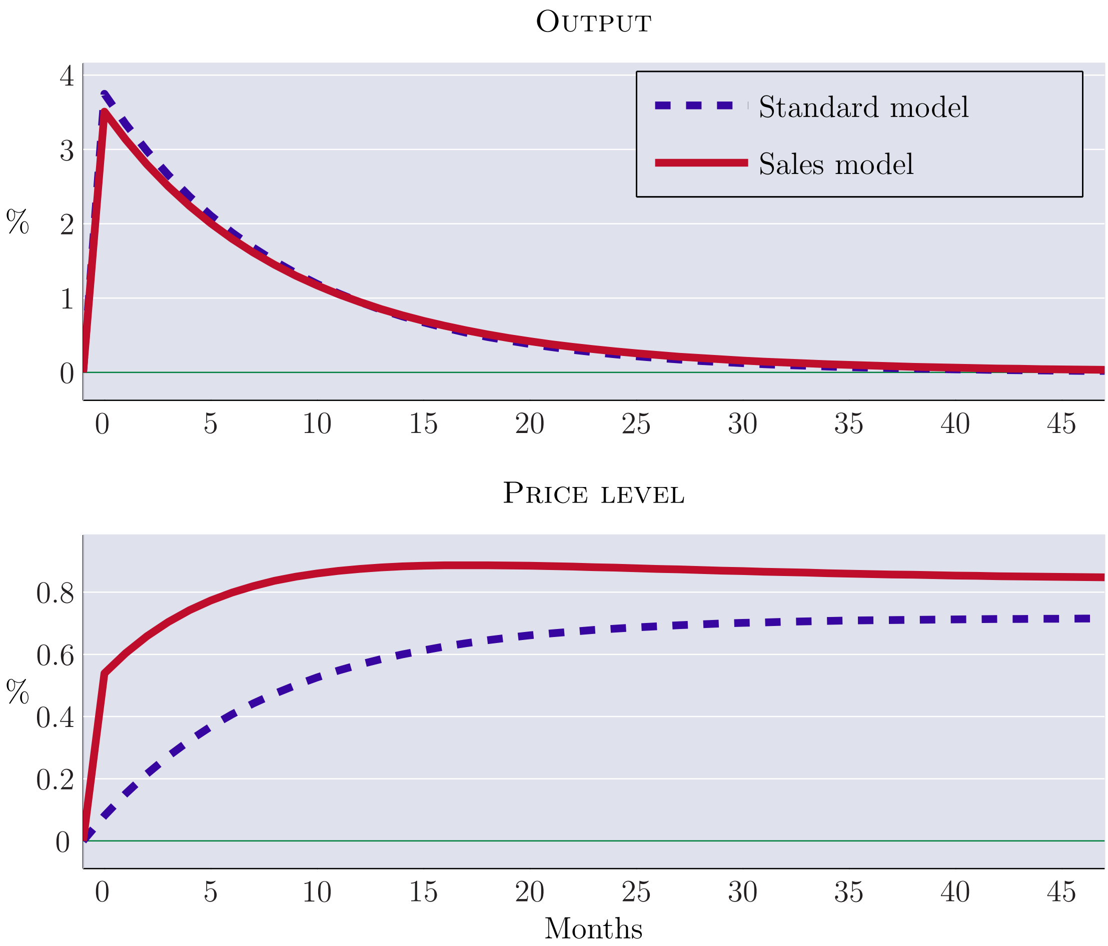
Notes: Generated using the baseline calibration of the DSGE model with sales when the money supply follows a random walk with drift.

When the central bank follows a Taylor rule, as in [7.17b], the reaction to shocks is somewhat different, as is seen in Figure 8. The responses of output in the sales model and in the standard model are now virtually identical. But the responses of prices are different. As before, the sales model features an initial jump in the price level. This is more marked than in the case of a shock to the money supply. The difference in price-level response diminishes over time, but does not vanish in the long run, and is found to be around 17% in the baseline calibration.

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

[13] Without any shocks at all, sales would still occur at a very similar frequency, but the price would alternate between two fixed levels.

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



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

8 Conclusions

For macroeconomists grappling with the welter of recent micro pricing evidence, one particularly puzzling feature is noteworthy: the large, frequent and short-lived price changes followed by prices returning exactly to their former levels. If price changes are driven purely by shocks then explaining this tendency requires a very special configuration of shocks. The model presented in this paper shows that this pricing behaviour arises in equilibrium if firms face consumers with sufficiently different price sensitivities. No idiosyncratic shocks are needed to generate sales.

The model explains why firms choose a two-price distribution with a normal price and a sale price, and thus want to switch frequently between the two points of the distribution. The two desired prices themselves are sensitive to shocks, but the magnitude of changes in the desired normal and sale prices is dwarfed by the gap between the two. So the apparent "puzzle" of why prices return to their former levels reduces to explaining why after a move from \$5.99 to \$4.49, a price returns to \$5.99 instead of \$6.02. But small costs of reoptimizing the normal price would explain firms' reluctance to make such small changes in accordance with a well-established literature in macroeconomics.

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

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

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

References

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

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

Calvo, G. A. (1983), "Staggered prices in a utility-maximizing framework", *Journal of Monetary Economics*, **12**:383–398. 28

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

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. 33

Clarida, R., Galí, J. and Gertler, M. (1998), "Monetary policy rules and macroeconomic stability: Evidence and some theory", *Quarterly Journal of Economics*, **115(1)**:147–180. 34

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. 1

Eichenbaum, M., Jaimovich, N. and Rebelo, S. (2008), "Reference prices and nominal rigidities", Working paper 13829, National Bureau of Economic Research. 1, 25

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**:281–313. 31, 32

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

Golosov, M. and Lucas, R. E., Jr. (2007), "Menu costs and Phillips curves", *Journal of Political Economy*, **115(2)**:171–199. 28

Hansen, G. D. (1985), "Indivisible labor and the business cycle", *Journal of Monetary Economics*, **16**:309–327. 28

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

Kehoe, P. and Midrigan, V. (2007), "Sales and the real effects of monetary policy", Working paper 652, Federal Reserve Bank of Minneapolis. 1, 3, 28

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

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

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

Nakamura, E. and Steinsson, J. (2007), "Five facts about prices: A reevaluation of menu cost models", Working paper, Harvard University. 1, 21, 22, 33

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. 2

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

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

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

Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, New Jersey. 33, 65

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

A Technical appendix

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

The structure of household consumption preferences introduced in section 3.2 implies that firms face a demand curve $q = \mathscr{D}(p; P_B, \mathcal{E})$ of the form given in equation [3.10] at each moment. It it 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

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

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

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

With these definitions, the original demand function [3.10] can be stated in terms of the relative demand function [A.1.1]:

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

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

$$\mathcal{R}(\mathfrak{q}) \equiv \mathfrak{q}\mathcal{D}^{-1}(\mathfrak{q}) . \qquad \text{[A.1.4]}$$

Using the inverse demand function $\rho = \mathcal{D}^{-1}(\mathfrak{q})$, total revenue can be equivalently expressed as $\mathcal{R}(\mathfrak{q}) = \mathcal{D}^{-1}(\mathfrak{q})\mathcal{D}\left(\mathcal{D}^{-1}(\mathfrak{q})\right)$, and by substituting the demand function from [A.1.1],

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

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

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

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

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

Just as [A.1.3] shows the original demand function $\mathscr{D}(p; P_B, \mathcal{E})$ in [3.10] is related to the relative demand function $\mathcal{D}(\rho)$ in [A.1.1], there are similar connections between the original inverse demand function, original total revenue $\mathscr{R}(q; P_B, \mathcal{E})$ and marginal revenue $\mathscr{R}'(q; P_B, \mathcal{E})$ functions and their equivalents defined in terms of the relative demand function. The relation between the inverse demand functions follows directly from [A.1.3]:

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

Equations [3.11], [A.1.4] and [A.1.6] justify the following links between the total revenue functions and their derivatives:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

As $\epsilon > 1$, 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 can be confirmed by deriving the discriminants and the sums and products of the roots). Let $\eta^*(\epsilon)$ be the larger of the two roots of the equation $\mathcal{G}_r(\eta;\epsilon) = 0$:

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

and observe that $\eta^*(\epsilon) > \epsilon$ for all $\epsilon > 1$. Since both quadratics $\mathcal{G}_p(\eta;\epsilon)$ and $\mathcal{G}_r(\eta;\epsilon)$ have a positive coefficient on η^2, it must be the case that they are negative for all η values lying strictly between their two roots.

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

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

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. Since $\epsilon > 1$, such a solution exists and is unique, and $\mathcal{G}_p(\eta;\epsilon) > \mathcal{G}_r(\eta;\epsilon)$ if and only if $\eta > \hat{\eta}(\epsilon)$. The difference between solution $\hat{\eta}(\epsilon)$ and ϵ is given by:

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

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

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

Therefore, for $\eta > \epsilon > 1$, if $\eta > \eta^*(\epsilon)$ then the quadratic equation $\mathcal{Q}(z) = 0$ from [A.1.11] has two distinct positive real roots $\underline{z}$ and $\overline{z}$ such that $\underline{z} < \overline{z}$, and $\mathcal{Q}(z) < 0$ must hold for all $z \in [\underline{z}, \overline{z}]$ since the coefficient of z^2 is positive. For $z \in [0, \underline{z})$ or $z \in (\overline{z}, \infty)$, the quadratic satisfies $\mathcal{Q}(z) > 0$. If $\eta \leq \eta^*(\epsilon)$ then $\mathcal{Q}(z) > 0$ for all z (except at just one isolated z value when $\eta = \eta^*(\epsilon)$ exactly). Therefore, in the case

$\eta \leq \eta^*(\epsilon)$, it follows from [A.1.10] and [A.1.11] that $\mathcal{R}'(\mathfrak{q})$ is strictly decreasing for all $\mathfrak{q} \geq 0$.

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

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

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

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

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

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

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

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

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

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

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

Proof (i) When $0 < \lambda < 1$, and condition [4.3] hold then Lemma 1 demonstrates that there exist $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$ such that $0 < \underline{\mathfrak{q}} < \overline{\mathfrak{q}} < \infty$ and $\mathcal{R}''(\underline{\mathfrak{q}}) = \mathcal{R}''(\overline{\mathfrak{q}}) = 0$. Define $\underline{\mathcal{R}'} \equiv \mathcal{R}'(\underline{\mathfrak{q}})$ and $\overline{\mathcal{R}'} \equiv \mathcal{R}'(\overline{\mathfrak{q}})$. Since Lemma 1 also shows that $\mathcal{R}'(\mathfrak{q})$ is strictly increasing between $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$, it must be the case 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 $[\underline{\mathfrak{q}}_1, \overline{\mathfrak{q}}_1]$, 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 the function 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 $\mathfrak{q}_1(\varkappa)$ to be

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

in other words, a convex combination of $\underline{\mathfrak{q}}_1$ and $\overline{\mathfrak{q}}_1$, which 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 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 a particular $\varkappa$. Hence define the functions $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$ to give these values on the two intervals for each specific $\varkappa \in [0, 1]$:

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

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

$$\mathfrak{q}_2(0) = \mathfrak{q}_3(0) = \bar{\mathfrak{q}} \ , \quad \mathfrak{q}_1(1) = \mathfrak{q}_2(1) = \underline{\mathfrak{q}} \ . \tag{A.1.21}$$

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

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

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

Existence

Now for each $\varkappa \in [0, 1]$, define the function $\digamma(\varkappa)$ in terms of the following integrals:

$$\digamma(\varkappa) \equiv \int_{\mathfrak{q}_2(\varkappa)}^{\mathfrak{q}_3(\varkappa)} \big(\mathcal{R}'(\mathfrak{q}) - \mathcal{R}'(\mathfrak{q}_2(\varkappa))\big)\, d\mathfrak{q} - \int_{\mathfrak{q}_1(\varkappa)}^{\mathfrak{q}_2(\varkappa)} \big(\mathcal{R}'(\mathfrak{q}_2(\varkappa)) - \mathcal{R}'(\mathfrak{q})\big)\, d\mathfrak{q} \ . \tag{A.1.22}$$

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

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

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 $\underline{\mathfrak{q}}_2 < \mathfrak{q} < \bar{\mathfrak{q}}_3$. By differentiating $\digamma(\varkappa)$ in [A.1.22] using Leibniz's rule and substituting the definitions from [A.1.20] leads to the following expression:

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

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

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

$$\int_{\mathfrak{q}_2(\varkappa^*)}^{\mathfrak{q}_S} \big(\mathcal{R}'(\mathfrak{q}) - \mathcal{R}'(\mathfrak{q}_2(\varkappa^*))\big)\, d\mathfrak{q} = \int_{\mathfrak{q}_N}^{\mathfrak{q}_2(\varkappa^*)} \big(\mathcal{R}'(\mathfrak{q}_2(\varkappa^*)) - \mathcal{R}'(\mathfrak{q})\big)\, d\mathfrak{q} \ , \tag{A.1.23}$$

which can be rearranged to deduce

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

And because [A.1.20] implies $\mathcal{R}'(\mathfrak{q}_2(\varkappa^*)) = \mathcal{R}'(\mathfrak{q}_N) = \mathcal{R}'(\mathfrak{q}_S)$, it is established that

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

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

Uniqueness

Now the uniqueness of this solution is demonstrated. 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 this function is $[\underline{\mathcal{R}'}, \overline{\mathcal{R}'}]$ on $[\underline{\mathfrak{q}}_1, \overline{\mathfrak{q}}_1]$, $[\underline{\mathfrak{q}}_2, \overline{\mathfrak{q}}_2]$ and $[\underline{\mathfrak{q}}_3, \overline{\mathfrak{q}}_3]$, it follows that no solution of [A.1.16] can be found in either $[0, \underline{\mathfrak{q}}_1)$ or $(\overline{\mathfrak{q}}_3, \infty)$ since marginal revenue $\mathcal{R}'(\mathfrak{q})$ needs to be equalized at two points. Furthermore, as the definition of the functions $\mathfrak{q}_1(\varkappa)$, $\mathfrak{q}_2(\varkappa)$ and $\mathfrak{q}_3(\varkappa)$ in [A.1.20] makes clear, for marginal revenue to be equalized at two quantities, it is necessary that those 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 $\varkappa \in [0, 1]$.

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

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

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

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

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

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

where the inequalities are strict for $\mathfrak{q} \neq \mathfrak{q}_N$ and $\mathfrak{q} \neq \mathfrak{q}_S$ respectively. Note that the equations in [A.1.16] characterizing $\mathfrak{q}_S$ and $\mathfrak{q}_N$ can be rearranged to show that:

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

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

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

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

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

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

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

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

using the fact that the coefficients of $\mathcal{R}(\underline{\mathfrak{q}})$ and $\mathcal{R}(\overline{\mathfrak{q}})$ in the above are strictly positive and sum to one. A weighted average of the two inequalities in [A.1.28] using as weights the coefficients from [A.1.29] yields

$\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_N) + \mathcal{R}'(\mathfrak{q}_N)(\mathfrak{q} - \mathfrak{q}_N)$ for all $\mathfrak{q} \in (\underline{\mathfrak{q}}, \overline{\mathfrak{q}})$. This finding, together with the inequalities in [A.1.26] and the equations [A.1.25] and [A.1.27], implies

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

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

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

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

Consider the functions:

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

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

$$\mathfrak{a}_2(\eta) \equiv \eta(\eta-1) \,, \tag{A.1.30c}$$

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

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

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

$$\mathfrak{b}_3(\eta) \equiv (\eta-1) \,, \tag{A.1.30g}$$

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

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

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

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

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

(ii) Given the value of μ satisfying $\mathfrak{R}(\mu;\epsilon,\eta) = 0$ and the function $\mathfrak{z}(\mu;\epsilon,\eta)$ from [A.1.32], the quantity ratio $\chi \equiv \mathfrak{q}_S/\mathfrak{q}_N$ is

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

and thus depends only on the parameters ϵ and η.

(iii) The equilibrium markups μ_S and μ_N depend only on ϵ and η and are given by:

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

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

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

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

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

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

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

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

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

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

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

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

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

Now note that the equations in [A.1.16] can be rearranged to deduce $\mathcal{R}(\mathfrak{q}_N) - \mathfrak{q}_N\mathcal{R}'(\mathfrak{q}_N) = \mathcal{R}(\mathfrak{q}_S) - \mathfrak{q}_S\mathcal{R}'(\mathfrak{q}_S)$. The definition of the total revenue function $\mathcal{R}(\mathfrak{q})$ in [A.1.4] shows that it can also be written as $\mathcal{R}(\mathcal{D}(\rho)) = \rho\mathcal{D}(\rho)$ 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 following equation is obtained:

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

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 equation [A.1.16]), [A.1.38] becomes

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

The formula for marginal revenue $\mathcal{R}'(\mathcal{D}(\rho))$ in [A.1.5] can be 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 can be substituted into [A.1.39] to obtain

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

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

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

Using the relative demand function from equation [A.1.1], the quantity ratio can also be written as $\chi = \mathcal{D}(\rho_S)/\mathcal{D}(\rho_N)$, thus

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

and by factorizing $\lambda\rho_S^{-\epsilon}$ and $\lambda\rho_N^{-\epsilon}$ from the numerator and denominator respectively, and using the definitions $\mu \equiv \rho_S/\rho_N$ and $z \equiv ((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, the expression becomes

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

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

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

Since the quantity ratio χ is finite, none of the terms in the denominators of [A.1.40] or [A.1.41] can be zero, so [A.1.42] may be rearranged 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 can be multiplied by $\mu^{2\eta-\epsilon}/(1-\mu^{\eta})$ to obtain an equivalent cubic equation:

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

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

These steps have demonstrated that starting from a solution $\mathfrak{q}_S$ and $\mathfrak{q}_N$ of [A.1.16], the quadratic and the cubic equations [A.1.37]–[A.1.43] in z must simultaneously hold for $z = ((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, with $\rho_N \equiv \mathcal{D}^{-1}(\mathfrak{q}_N)$, and where the coefficient functions [A.1.30] are evaluated at $\mu = \rho_S/\rho_N$, with $\rho_S \equiv \mathcal{D}^{-1}(\mathfrak{q}_S)$. If the quadratic equation $\mathfrak{Q}(z;\mu,\epsilon,\eta) = 0$ and cubic equation $\mathfrak{C}(z;\mu,\epsilon,\eta) = 0$ share a root then it is a standard result from the theory of polynomials that the *resultant* $\mathfrak{R}(\mu;\epsilon,\eta)$, as defined in [A.1.31], is zero. Since the coefficients of the polynomials $\mathfrak{Q}(z;\mu,\epsilon,\eta)$ and $\mathfrak{C}(z;\mu,\epsilon,\eta)$ are functions 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 can depend only on these parameters.

It is known that the solution of [A.1.16] for $\mathfrak{q}_S$ and $\mathfrak{q}_N$ is unique, and therefore so must the solution of $\mathfrak{R}(\mu;\epsilon,\eta)=0$ for μ, given the additional 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 also be a positive real number. As the product of the roots of the quadratic $\mathcal{Q}(z)=0$ is positive, the shared root z is positive and real if and only if either branch of the quadratic root function is positive and real. Thus this can be tested by checking whether $\mathfrak{z}(\mu;\epsilon,eta)$ is positive and real.

Note that the resultant is always zero at $\mu=0$ and $\mu=1$ for all values of ϵ and η. This can be seen by taking limits of the coefficients in [A.1.30] as $\mu\to 0$ and $\mu\to 1$, which yields

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

and 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 $z=((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$ finds the relative price ρ_N such that with $\rho_S=\mu\rho_N$, marginal revenue is the same at both ρ_S and ρ_N. Given the properties of marginal revenue derived in Lemma 1 under the conditions shown by Lemma 2, which are necessary for the solution $\mathfrak{q}_S$ and $\mathfrak{q}_N$ to exist, there are two candidate solutions for ρ_N that meet this criterion. However, Lemma 2 shows that both ρ_N and ρ_S are on the downward-sloping sections of the marginal revenue curve. To rule out a solution in the middle upward-sloping section of marginal revenue, it is necessary to discard the smaller of the two ρ_N candidates 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 $\mathfrak{a}_2(\eta)$ from [A.1.30] is positive, the smaller of the two roots of quadratic $\mathfrak{Q}(z;\mu,\epsilon,\eta)=0$ is found using the expression $\mathfrak{z}(\mu;\epsilon,\eta)$ in [A.1.32].

The equilibrium quantity ratio χ is obtained by substituting $z=\mathfrak{z}(\mu;\epsilon,\eta)$ into [A.1.41].

(iii) Since $\rho_S\equiv P_S/P_B$ and $\rho_N\equiv P_N/P_B$ according to [A.1.2], the formula for the purchase multipliers in [3.10] implies $v_N=\rho_N^{\epsilon-\eta}$ and $v_S=\mu^{\epsilon-\eta}v_N$. Using the fact that $z\equiv((1-\lambda)/\lambda)\rho_N^{\epsilon-\eta}$, and dividing numerator and denominator by λ in the expression [4.5] yields [A.1.34].

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

A.2 Proof of Proposition 2

Using the relationship between the total revenue function $\mathscr{R}(q;P_B,\mathcal{E})$ and its equivalent $\mathcal{R}(\mathfrak{q})$ defined in [A.1.4] using relative demand function $\mathcal{D}(\rho)$ from [A.1.1], the corresponding marginal revenue functions $\mathscr{R}'(q;P_B,\mathcal{E})$ and $\mathcal{R}'(\mathfrak{q})$ are proportional according to [A.1.7]. Lemma 1 demonstrates that $\mathcal{R}'(\mathfrak{q})$ is non-monotonic under the condition [4.3], which yields the result.

A.3 Proof of Theorem 1

Existence of two-price equilibrium

For a two-price equilibrium to exist it is necessary that first-order conditions [4.4] for profit-maximization are satisfied for 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 price index for a bargain hunter, and $\mathcal{E}=P^\epsilon Y$ is the measure of aggregate expenditure.

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

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

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

It is necessary to check the relative quantities $\mathfrak{q}_S$ and $\mathfrak{q}_N$ are well defined to confirm the solution is 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 price index equation for P_B in [4.8] implies

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

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

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

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

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

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

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

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

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

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

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

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

Recalling the equivalence of inequalities [A.3.3] and [A.3.5], it follows that $s \in (0,1)$ if and only if $\lambda \in (\underline{\lambda}(\epsilon,\eta), \overline{\lambda}(\epsilon,\eta))$. Therefore, for $\lambda \in [0, \underline{\lambda}(\epsilon,\eta)]$ or $\lambda \in [\overline{\lambda}(\epsilon,\eta), 1]$ there can be no two-price equilibrium.

Therefore, given elasticities ϵ and η satisfying with non-monotonicity condition [4.3] and loyal fraction $\lambda \in (\underline{\lambda}(\epsilon,\eta), \overline{\lambda}(\epsilon,\eta))$, by using the arguments above there exist two distinct relative prices $\rho_S \equiv p_S/P_B$ and $\rho_N \equiv p_N/P_B$ and a sales fraction $s \in (0,1)$ consistent with the first two equalities in [4.4]. Lemma 3 then demonstrates that the two purchase multipliers v_S and v_N and the two optimal markups μ_S and μ_N are also determined. Equations [4.1], [4.2] and [4.5], it follows that by using the optimal markups, the remaining first-order condition in [4.4] involving marginal cost is also satisfied. The other variables needed for the macroeconomic equilibrium can then be determined as discussed in section 4.

Finally, the remaining first-order condition [3.13c] is checked, and then it is argued that the first-order conditions collectively are sufficient as well as necessary for maximizing profits. Using the relationships in [A.1.7] and the results in [A.1.17] of Lemma 2 the following can be deduced:

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

Since $s \in (0,1)$, the Lagrangian multiplier $\aleph$ from first-order conditions [3.13b]–[3.13c] can be determined:

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

and thus from [A.3.7] it is known that $\aleph > 0$. By combining this expression for the Lagrangian multiplier with the first-order condition [3.13c]:

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

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

The assumptions made on the production function [3.7] ensure that the total cost function $\mathscr{C}(Q; W)$ in [3.8] is continuously differentiable and convex, so for all $q \geq 0$:

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

where $Q \equiv sq_S + (1-s)q_N$ is the specific total physical quantity sold using the two-price strategy constructed earlier. Now consider a general alternative pricing strategy for a given firm, assuming that all other firms continue to use the same two-price strategy. The new strategy is specified in terms of a distribution function $F(p)$ for prices. Let $G(q) \equiv 1 - F(\mathscr{D}(p; P_B, \mathcal{E}))$ be the implied distribution function for quantities sold. The level of profits $\mathscr{P}$ from the new strategy can be obtained by making a change of variable from prices to quantities in the integrals in [3.12]:

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

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

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

The first-order conditions [4.4] imply that the coefficient of the integral in the above 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 that $Q = sq_S + (1-s)q_N$, 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. There is thus no profit-improving deviation from the two-price strategy.

This establishes that a two-price equilibrium exists when [4.3] and $\lambda \in (\underline{\lambda}(\epsilon,\eta), \overline{\lambda}(\epsilon,\eta))$ hold, and that it is unique within the class of two-price equilibria.

Uniqueness of two-price equilibrium

Suppose that the parameters ϵ and η are such that the 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 corresponding to this single price must necessarily be equal to one. The relative prices ρ_S and ρ_N in the two-price equilibrium cannot be found 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 $\rho_N = \mathcal{D}^{-1}(\rho_N)$ for the relative quantities $\mathfrak{q}_S$ and $\mathfrak{q}_N$. Since [A.1.1] implies $\mathcal{D}(1) = 1$ and because demand $\mathcal{D}(\rho)$ is strictly decreasing in ρ, it must be the case that $\mathfrak{q}_N < 1 < \mathfrak{q}_S$.

It is known from Lemma 1 that $\mathcal{R}(\mathfrak{q})$ is strictly concave in the intervals $(0, \underline{\mathfrak{q}})$ and $(\overline{\mathfrak{q}}, \infty)$, strictly convex in $(\underline{\mathfrak{q}}, \overline{\mathfrak{q}})$, 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 $\mathfrak{q} = 1$ for all firms in the one-price equilibrium, the actual common quantity being produced is $q_1 = \mathcal{E}/P_B^{\epsilon}$ using [A.1.2], where P_B and $\mathcal{E}$ are the values of these variables associated with the putative one-price equilibrium. Since $\mathcal{R}''(1) > 0$, equation [A.1.7] implies $\mathscr{R}''(q_1; P_B, \mathcal{E}) > 0$. Therefore, for sufficiently small $\xi > 0$, the profits $\mathscr{P}$ from offering quantity $q_1 - \xi$ to one half of moments and $q_1 + \xi$ to the other half exceed profits from offering one price and quantity to all moments:

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

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

Now consider the case where $\mathfrak{q}_N < 1 < \underline{\mathfrak{q}}$. Let $p_1 = P_B$ denote the price that it is claimed all firms will use 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 sales *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. Consider an alternative strategy that sells at price $\rho_S p_1$ at a fraction ξ of moments and at price p_1 at the remaining fraction $1 - \xi$. Profits from the hybrid strategy are denoted by $\mathscr{P}$ and are given by

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

Since the cost function $\mathscr{C}(q; W)$ is differentiable in q, the above equation can be expressed as

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

where $\mathscr{O}\left(\xi^2\right)$ denotes second- and higher-order terms in ξ. A necessary condition for a one-price equilibrium is that the single price p_1 is chosen optimally, which reduces to the usual marginal revenue equals marginal cost condition $\mathscr{R}'(q_1; P_B, \mathcal{E}) = \mathscr{C}'(q_1; W)$, so:

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

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

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

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

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

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

$$\frac{\mathcal{R}(\mathfrak{q}_S)-\mathcal{R}(1)}{\mathfrak{q}_S-1} > \mathcal{R}'(\mathfrak{q}_N) > \mathcal{R}'(1) \; , \qquad \text{[A.3.14]}$$

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

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

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

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

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

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

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

Given $q_1 > q_N$, for sufficiently small $\xi > 0$, [A.3.16] and [A.3.17] show that there is a hybrid strategy that 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 parameter values.

One-price equilibrium

The first thing to note is that when the two-price equilibrium fails to exist owing to the violation of the non-monotonicity condition [4.3], then marginal revenue $\mathscr{R}'(q; P_B, \mathcal{E})$ is strictly decreasing for all q. This makes total revenue $\mathscr{R}(q; P_B, \mathcal{E})$ a strictly concave function of quantity q. Since total cost $\mathscr{C}(q; W)$ is a convex function, it is easy to see that the profit function is globally concave, and thus a one-price equilibrium will always exist, and be the only possible equilibrium for this parameter range.

To see that a one-price equilibrium exists and is unique in the other case where a two-price equilibrium is not found, namely when marginal revenue is non-monotonic, but $\lambda \in [0, \underline{\lambda}(\epsilon, \eta)]$ or $\lambda \in [\overline{\lambda}(\epsilon, \eta), 1]$, note that λ lying in these intervals is equivalent to $1 > \mathfrak{q}_S$ or $1 < \mathfrak{q}_N$ respectively.

Taking the first of these cases, the concavity of $\mathcal{R}(\mathfrak{q})$ on $[\bar{\mathfrak{q}}, \infty)$ (which includes $\mathfrak{q}_S$), as shown in Lemma 1, establishes that $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(1) + \mathcal{R}'(1)(\mathfrak{q}-1)$ for all $\mathfrak{q} \in [\bar{\mathfrak{q}}, \infty)$. Lemma 2 shows that $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(\mathfrak{q}_S) + \mathcal{R}'(\mathfrak{q}_S)(\mathfrak{q} - \mathfrak{q}_S)$ for all $\mathfrak{q} \geq 0$. First note that the concavity of $\mathcal{R}(\mathfrak{q})$ 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 $\mathfrak{q} = \mathfrak{q}_S$ gives $\mathcal{R}(\mathfrak{q}_S) \leq \mathcal{R}(1) + \mathcal{R}'(1)(\mathfrak{q}_S - 1)$. By combining these results, $\mathcal{R}(\mathfrak{q}) \leq \mathcal{R}(1) + \mathcal{R}'(1)(\mathfrak{q}-1)$ for all $\mathfrak{q} \geq 0$ is obtained. Then using [A.1.2] and [A.1.7] to translate this into a property of the original total revenue function $\mathscr{R}(q; P_B, \mathcal{E})$ for all q:

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

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

A.4 Proof of Proposition 3

(i) It is shown in Lemma 3 that μ and χ can be uniquely determined as functions of ϵ and η when inequality [4.3] is satisfied, as is necessary for a two-price equilibrium. Lemma 3 also gives solutions for μ_S and μ_N, and implicitly v_S and v_N using [A.1.35] and the fact that $v_S = (p_S/P_B)^{-(\eta-\epsilon)}$ and $v_N = (p_N/P_B)^{-(\eta-\epsilon)}$. These only depend on ϵ, η and λ. In conjunction with equation [4.8], knowledge of v_S and v_N yield a linear equation for s after dividing both sides by P_B.

(ii) Lemma 3 shows that μ, μ_S, μ_N and χ are independent of λ, establishing the first four claims. Differentiating [A.3.6] yields the fifth claim.

(iii) Substituting the bounds for λ from [A.3.4] into equation [A.3.6] proves the claim.

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

Obtaining the roots $\underline{z}$ and $\overline{z}$ of the quadratic $\mathcal{Q}(z) = 0$ from [A.1.11] of Lemma 1, and taking the limit as $\epsilon \to 1^+$ yields $\underline{z} \to 0$ and $\overline{z} \to (\eta-2)/\eta$. Note that $\underline{\mathfrak{q}}$ and $\overline{\mathfrak{q}}$ from Lemma 1 are related to $\underline{z}$ and $\overline{z}$ by the transformation $\mathcal{Z}^{-1}(z)$ from [A.1.15], which is strictly increasing.

Let z_S and z_N be defined as follows in terms of the relative prices ρ_S and ρ_N:

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

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

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

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

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

A.5 Log linearizations

A.5.1 Sales model

The notational convention adopted here is that a bar above a variable 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 sales fraction s, where it denotes just the deviation).

Consider first the demand function faced by firms. In the following, p_S and p_N denote a particular firm's sale price and normal price respectively. P_S and P_N denote the common sale and normal prices chosen by other firms. Equation [4.10] gives the levels of demand q_S and q_N at sale and normal prices p_S

and p_N, which has the following log-linearized form:

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

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

given in terms of log deviations of the purchase multipliers:

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

Substituting the purchase multipliers into the demand functions:

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

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

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

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

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

Overall demand $Q = sq_S + (1-s)q_N$ can be log-linearized as follows:

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

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

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

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

where P_S and P_N are the average log-deviations of p_S and p_N, and s is the average deviation of the sales fractions. In the static model, $\mathsf{P}_S = \mathsf{p}_S = \mathsf{X}$ and $\mathsf{P}_N = \mathsf{p}_N = 0$. These averages are interpreted differently in the dynamic model.

Similarly, the log-linearized general price level P in [3.3] is:

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

where P_S, P_N and s are averages as discussed above, and the log-deviations of the purchase multipliers are evaluated at the average prices. Then using the expressions for the purchase multipliers and relative prices in the flexible-price equilibrium together with the log-deviation of v_S and v_N from [A.5.2], and the expression for P_B in [A.5.6], the aggregate price level is given by:

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

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

The production function [3.7] can be log-linearized to yield

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

The nominal marginal cost function [3.8] has the following log-linear form:

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

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

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

where P_N and s are the averages discussed above. Substituting $\mathsf{p}_S = \mathsf{P}_B = \mathsf{X}$ into [A.5.6] for P_B and rearranging yields:

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

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

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

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

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

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

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

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

A.5.2 Model with flexible wages

The log-linearized real wage $\mathsf{w} = \mathsf{W} - \mathsf{P}$ adjusts so that the log linearization of equation [6.1] is satisfied:

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

A.5.3 Aggregation in the dynamic sales model

The equivalent of aggregate price level P_t in [7.3] for $P_{B,t}$, the bargain hunters' price index, is obtained from the definition [4.8]:

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

Using the demand function [3.10], the total quantity sold for a vintage-ℓ firm is:

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

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

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

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

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

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

and also:

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

A.5.4 DSGE log-linearizations

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

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

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

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

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

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

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

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

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

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

Equation [7.15] for the utility-maximizing reset wage is log linearized as follows:

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

which has the following recursive form:

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

The log-linearized wage index [7.16] is:

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

which also has a recursive form:

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

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

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

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

A.6 Proof of Theorem 2

(i) Suppose all firms share the same fixed p_N consistent with the flexible-price equilibrium. The first-order condition for the optimal choice of the sales fraction s from the first part of [5.1] is log-linearized as follows:

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

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

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

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

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

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

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

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

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

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

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

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

so $\mathsf{p}_S = \mathsf{X}$ if the coefficient can be shown to be zero. Using the expressions for $\mathfrak{c}_S$ from [A.5.4] and $\bar{\mu}_S$ from [4.6]:

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

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

A.7 Log-linearized solution of the static model

A.7.1 Fixed wages

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

$$\mathsf{P} = \theta_P\mathsf{p}_S - \psi_P\mathsf{s} \,, \quad \text{[A.7.1a]}$$

$$\mathsf{P}_B = \theta_B\mathsf{p}_S - \psi_B\mathsf{s} \,, \quad \text{[A.7.1b]}$$

$$\mathsf{p}_S = \mathsf{X} \,, \quad \text{[A.7.1c]}$$

$$\mathsf{P}_B = \mathsf{X} \,, \quad \text{[A.7.1d]}$$

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

$$\mathsf{X} = \gamma\mathsf{Q} \,, \quad \text{[A.7.1f]}$$

$$\mathsf{Y} = \mathsf{M} - \mathsf{P} \,. \quad \text{[A.7.1g]}$$

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

A.7.2 Flexible wages

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

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

$$\mathsf{X} = \mathsf{W} + \gamma\mathsf{Q} \,. \qquad \text{[A.7.2b]}$$

A.8 Proof of Proposition 4

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

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

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

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

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

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

(iii) If both prices are optimally readjusted then [4.5] implies $p_S^* = \mu_S X$ and $p_N^* = \mu_N X$, which in log-linear terms becomes:

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

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

(iv) Let p_S and p_N be given prices for a particular firm, and let s be the optimal sales fraction implied by the first part of [5.1]. Profits [3.12] are denoted by $\mathscr{P}$:

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

where $\mathscr{C}(Q;W)$ is the cost function [3.8]. Taking a second-order Taylor expansion of profits around the flexible-price equilibrium yields:

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

The first-order approximation [A.5.3] of the demand functions [4.10] is extended to include second-order terms:

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

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

where the expressions for the price elasticities in [4.1] have been used, and the following are defined:

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

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

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

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

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

where:

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

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

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

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

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

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

Substituting the second-order expansions of demand from [A.8.3a] and using the expressions for $\bar{\mu}_S$ and $\bar{\mu}_N$ from [4.6], and $\bar{q}_S(\bar{\mu}_S - 1) = \bar{q}_N(\bar{\mu}_N - 1)$ to demonstrate that the first-order terms have zero coefficients:

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

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

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

where $\bar{\mu}_S - 1 = 1/(\bar{\zeta}_S - 1)$ has been used. A similar expression holds for p_N and q_N. Substituting this result into [A.8.4b] yields:

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

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

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

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

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

To simplify the expression for $\mathfrak{P}_3$, note that:

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

using $(\bar{\mu}_S - 1)\bar{q}_S = (\bar{\mu}_N - 1)\bar{q}_N$ repeatedly and the expressions for $\bar{\mu}_S$ and $\bar{\mu}_N$ from [4.6]. Substituting the result into [A.8.8]:

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

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

Appealing to Proposition 4, the log deviations of nominal marginal cost X and total quantity sold Q are independent of an individual firm's choice of p_S and p_N. Therefore all the terms in $\mathfrak{P}_4$ are independent of pricing policy. Furthermore, in the expression for $\mathfrak{P}_3$, the product of Q and P_B is also independent. Therefore:

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

By adding $\mathfrak{P}_2$ and $\mathfrak{P}_3$ from [A.8.6] and [A.8.9] and substituting the first-order expansion of the demands q_S and q_N into the latter, the following is obtained:

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

By completing the square and noting that the remainder is independent of pricing policy:

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

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

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

which completes the proof.

A.9 Proof of Theorem 3

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

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

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 purchase multiplier, itself log-linearized in [A.5.2]. Putting together those results, it follows that $\mu_{N,\ell,t+\ell} = (\eta - \epsilon)\mathfrak{c}_N \left(\mathsf{R}_{N,t} - \mathsf{P}_{B,t+\ell}\right)$. Proposition 4 shows that marginal cost is equalized across all price vintages and thus $\mathsf{X}_{\ell,t} = \mathsf{X}_t$. Furthermore it shows that $\mathsf{X}_t = \mathsf{P}_{B,t}$. Substituting these findings into [A.9.1]:

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

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

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

which can be expressed in an equivalent recursive form:

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

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

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

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

$$\mathsf{P}_{N,t} = \phi_p \mathsf{P}_{N,t-1} + (1 - \phi_p)\mathsf{R}_{N,t} \; . \qquad \text{[A.9.4]}$$

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

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

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

$$\psi_B \mathsf{s}_t = (1 - \theta_B)(\mathsf{P}_{N,t} - \mathsf{X}_t) \; . \qquad \text{[A.9.6]}$$

Equation [A.9.5] can be written as

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

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

$$\mathsf{X}_t - \mathsf{P}_{N,t} = \frac{1}{1 - \psi}\mathsf{x}_t \; , \qquad \text{[A.9.7]}$$

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

By multiplying both sides of [A.9.2] by $(1 - \phi_p)$ and by substituting the recursive equation [A.9.4] for

$\mathsf{P}_{N,t}$,

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

which can be expressed in terms of $\pi_{N,t} \equiv \mathsf{P}_{N,t} - \mathsf{P}_{N,t-1}$:

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

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

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

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

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

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

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

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

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

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

By using this equation and [A.9.7] together with [A.9.8] leads to:

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

which can be rearranged to yield the result [7.5].

By recursive forward substitution of the Phillips curve [7.5]:

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

Notice that all $\Delta\mathsf{x}_{t+\ell}$ terms apart from $\Delta\mathsf{x}_t$ cancel out because they occur twice with opposite signs and thus equation [7.6] is obtained. This completes the proof.

A.10 Log-linearized solution of the DSGE model

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

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

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

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

$$\pi_{W,t} = \beta\mathbb{E}_t\pi_{W,t+1} + \frac{(1-\phi_w)(1-\beta\phi_w)}{\phi_w}\frac{1}{1+\varsigma\sigma_h^{-1}}\left(\frac{\sigma_h^{-1}}{\alpha}\mathsf{Q}_t + \sigma_c^{-1}\left(\mathsf{Y}_t - \vartheta_m\mathsf{m}_t\right) - \mathsf{w}_t\right) , \qquad [A.10.1b]$$

$$\Delta\mathsf{w}_t = \pi_{W,t} - \pi_t , \qquad [A.10.1c]$$

$$\mathsf{Y}_t = \mathsf{Q}_t + \delta_x\mathsf{x}_t , \qquad [A.10.1d]$$

$$\mathsf{x}_t = \mathsf{w}_t + \gamma\mathsf{Q}_t , \qquad [A.10.1e]$$

$$\mathsf{Y}_t = \mathbb{E}_t\mathsf{Y}_{t+1} + \vartheta_m\left(\mathsf{m}_t - \mathbb{E}_t\mathsf{m}_{t+1}\right) - \sigma_c\left(\mathsf{i}_t - \mathbb{E}_t\pi_{t+1}\right) , \qquad [A.10.1f]$$

$$\mathsf{m}_t = \vartheta_y\mathsf{Y}_t - \vartheta_i\mathsf{i}_t . \qquad [A.10.1g]$$

Equation [A.10.1a] is the Phillips curve derived in Theorem 3. Equation [A.10.1b] is the Phillips curve for wage inflation from [A.5.22], and [A.10.1c] follows from the definition of the real wage. Equations [A.10.1d] and [A.10.1e] are taken from [A.7.1e] and [A.7.2b], which continue to hold in the dynamic model. The IS equation [A.10.1f] and money demand [A.10.1g] come from [A.5.17] and [A.5.18].

There are two specifications of monetary policy considered: exogenous money growth [7.17a],

$$\Delta\mathsf{M}_t = \varphi_m\Delta\mathsf{M}_{t-1} + \mathsf{e}_t , \qquad [A.10.1h]$$

and the Taylor rule [7.17b],

$$\mathsf{i}_t = \varphi_i\mathsf{i}_{t-1} + (1-\varphi_i)\left(\varphi_\pi\pi_t + \varphi_y\mathsf{Y}_t\right) + \mathsf{e}_t . \qquad [A.10.1i]$$

The standard model with Dixit-Stiglitz preferences, a one-price equilibrium, and Calvo staggered adjustment times leads to the following New Keynesian Phillips curve:

$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \kappa\mathsf{x}_t ,$$

in place of [A.10.1a].[14] Equation [A.10.1d] is replaced by $\mathsf{Q}_t = \mathsf{Y}_t$.

A.11 Second-order approximation of profits in standard model

Suppose a given firm charges price p and the general price level is P and output is Y. Standard Dixit-Stiglitz preferences imply the following demand function:

$$q = \left(\frac{p}{P}\right)^{-\varepsilon} Y .$$

Assume the total cost function is $\mathscr{C}(q;W)$. Profits $\mathscr{P}$ are then given by

$$\mathscr{P} = \frac{p^{1-\varepsilon}}{P^{-\varepsilon}}Y - \mathscr{C}\left(\left(\frac{p}{P}\right)^{-\varepsilon}Y;W\right) .$$

Taking a second-order approximation of total revenue yields

$$\frac{p^{1-\varepsilon}}{P^{-\varepsilon}}Y = \bar{Y}\left(1 + (1-\varepsilon)\mathsf{p} - \varepsilon\mathsf{P} + \mathsf{Y} + \frac{1}{2}\left((1-\varepsilon)\mathsf{p} - \varepsilon\mathsf{P} + \mathsf{Y}\right)^2\right) + \mathscr{O}(3) ,$$

and of total cost yields:

$$\mathscr{C}(q;W) = \mathscr{C}(\bar{Y};\bar{W}) + \left(\frac{\varepsilon-1}{\varepsilon}\right)\bar{Y}\left(-\varepsilon(\mathsf{p}-\mathsf{P}) + \mathsf{Y} + \frac{1}{2}(1+\gamma)\left(-\varepsilon(\mathsf{p}-\mathsf{P}) + \mathsf{Y}\right)^2\right) + \mathscr{O}(3) ,$$

[14] See Woodford (2003) for a derivation of this equation.

where $\gamma \equiv \bar{Y}\mathscr{C}''(\bar{Y};\bar{W})/\mathscr{C}'(\bar{Y};\bar{W})$, and $\mathscr{C}'(\bar{Y};\bar{W}) = (\varepsilon - 1)/\varepsilon$ and $\mathsf{q} = -\varepsilon(\mathsf{p} - \mathsf{P}) + \mathsf{Y}$ have been used. Putting these expressions together and rearranging terms leads to the following expression for profits:

$$\mathscr{P} = -\frac{1}{2}\varepsilon(1+\varepsilon\gamma)\bar{x}\bar{P}\bar{Y}\left(\mathsf{p} - \left(\mathsf{P} + \frac{1}{1+\varepsilon\gamma}\mathsf{x}\right)\right)^2 + \text{t.i.p.} + \mathscr{O}(3) \ ,$$

where $\mathsf{x} = \gamma\mathsf{Y}$ is the real marginal cost of all other firms.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

886	Andrew E. Clark David Masclet Marie-Claire Villeval	Effort and Comparison Income Experimental and Survey Evidence
885	Alex Bryson Richard B. Freeman	How Does Shared Capitalism Affect Economic Performance in the UK?
884	Paul Willman Rafael Gomez Alex Bryson	Trading Places: Employers, Unions and the Manufacture of Voice
883	Jang Ping Thia	The Impact of Trade on Aggregate Productivity and Welfare with Heterogeneous Firms and Business Cycle Uncertainty
882	Richard B. Freeman	When Workers Share in Profits: Effort and Responses to Shirking
881	Alex Bryson Michael White	Organizational Commitment: Do Workplace Practices Matter?
880	Mariano Bosch Marco Manacorda	Minimum Wages and Earnings Inequality in Urban Mexico. Revisiting the Evidence
879	Alejandro Cuñat Christian Fons-Rosen	Relative Factor Endowments and International Portfolio Choice
878	Marco Manacorda	The Cost of Grade Retention
877	Ralph Ossa	A 'New Trade' Theory of GATT/WTO Negotiations
876	Monique Ebell Albrecht Ritschl	Real Origins of the Great Depression: Monopoly Power, Unions and the American Business Cycle in the 1920s
875	Jang Ping Thia	Evolution of Locations, Specialisation and Factor Returns with Two Distinct Waves of Globalisation
874	Monique Ebell Christian Haefke	Product Market Deregulation and the U.S. Employment Miracle
873	Monique Ebell	Resurrecting the Participation Margin
872	Giovanni Olivei Silvana Tenreyro	Wage Setting Patterns and Monetary Policy: International Evidence

871	Bernardo Guimaraes	Vulnerability of Currency Pegs: Evidence from Brazil
870	Nikolaus Wolf	Was Germany Ever United? Evidence from Intra- and International Trade 1885 - 1993
869	L. Rachel Ngai Roberto M. Samaniego	Mapping Prices into Productivity in Multisector Growth Models
868	Antoni Estevadeordal Caroline Freund Emanuel Ornelas	Does Regionalism Affect Trade Liberalization towards Non-Members?
867	Alex Bryson Harald Dale-Olsen	A Tale of Two Countries: Unions, Closures and Growth in Britain and Norway
866	Arunish Chawla	Multinational Firms, Monopolistic Competition and Foreign Investment Uncertainty
865	Niko Matouschek Paolo Ramezzana Frédéric Robert-Nicoud	Labor Market Reforms, Job Instability, and the Flexibility of the Employment Relationship
864	David G. Blanchflower Alex Bryson	Union Decline in Britain
863	Francesco Giavazzi Michael McMahon	Policy Uncertainty and Precautionary Savings
862	Stephen Hansen Michael F. McMahon	Delayed Doves: MPC Voting Behaviour of Externals
861	Alex Bryson Satu Nurmi	Private Sector Employment Growth, 1998-2004: A Panel Analysis of British Workplaces
860	Alejandro Cuñat Szabolks Deak Marco Maffezzoli	Tax Cuts in Open Economies
859	Bernd Fitzenberger Karsten Kohn Alexander Lembcke	Union Density and Varieties of Coverage: The Anatomy of Union Wage Effects in Germany
858	Dimitra Petropoulou	International Trade, Minimum Quality Standards and the Prisoners' Dilemma
857	Andreas Georgiadis	Efficiency Wages and the Economic Effects of the Minimum Wage: Evidence from a Low-Wage Labour Market
856	L. Rachel Ngai Christopher A. Pissarides	Employment Outcomes in the Welfare State

The Centre for Economic Performance Publications Unit
Tel 020 7955 7673 Fax 020 7955 7595 Email info@cep.lse.ac.uk
Web site http://cep.lse.ac.uk