

## Sales and Monetary Policy

By BERNARDO GUIMARAES AND KEVIN D. SHEEDY\*

A striking fact about pricing is that many price changes are “sales”: large temporary cuts followed by prices returning exactly to their former levels. Figure 1 shows a typical price path for a six-pack of Corona beer at an outlet of Dominick’s Finer Foods, a US supermarket. Sales are frequent; other types of price change are rare. This pattern is an archetype of retail pricing.<sup>1</sup>

Monetary policy’s real effects on the economy depend crucially on the stickiness of prices. So Figure 1 poses a conundrum: viewed from different perspectives, the price path exhibits great flexibility on the one hand, but substantial stickiness on the other. While changes between some “normal” price and a temporary “sale” price are frequent, the normal price itself changes far less often.<sup>2</sup> Consequently, empirical estimates of price stickiness widely diverge when sales are treated differently. Mark 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 four months; by disregarding sales, Nakamura and Steinsson (2008) find a median duration of around nine months.<sup>3</sup> Quantitative models deliver radically different estimates of the real effects of monetary policy depending on which of these two numbers is used. Hence, an understanding of sales is needed to answer the question of how large those real effects should be.

In the industrial organization (IO) and marketing literatures, the most prominent theories of sales are based on customer heterogeneity, together with incomplete information. Leading examples include Steven Salop and Joseph E. Stiglitz (1977, 1982), Hal R. Varian (1980), Joel Sobel (1984), and Chakravarthi Narasimhan (1988). This paper builds a general-equilibrium macroeconomic model with sales that draws upon the rationale proposed in these literatures. Despite substantial

\* Guimaraes: Department of Economics, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK (e-mail: b.guimaraes@lse.ac.uk); Sheedy: Department of Economics, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK (e-mail: k.d.sheedy@lse.ac.uk). We thank Ruediger Bachmann, Hafedh Bouakez, Ariel Burstein, Andy Levin, Ananth Ramanarayanan, Attila Rátfai, and especially Rachel Ngai for helpful discussions and suggestions. The paper has also benefited from suggestions made by three anonymous referees. We are grateful for comments from seminar and conference participants at: Amsterdam, Austrian National Bank, Bocconi, Bristol, Cergy-Pontoise, Essex, ECB, FGV–Rio, Humboldt, London School of Economics, Paris School of Economics, Tilburg, and the Toulouse School of Economics; Banque de France, CEP annual conference, CEP/ESRC Monetary Policy conference, CEPR–MNB conference 2009, COOL 2008, ESSIM 2009, LACEA 2008, LBS, MMFRG conference 2008, RES 2009, Midwest Macro Meetings 2009, NBER Monetary Economics Fall meeting 2008, SED 2009, and the Texas Monetary Conference 2008.

<sup>1</sup>See Daniel Hosken and David Reiffen (2004), Peter J. Klenow and Oleksiy Kryvtsov (2008), Emi Nakamura and Jon Steinsson (2008), Patrick J. Kehoe and Virgiliu Midrigan (2008), Pinelopi K. Goldberg and Rebecca Hellerstein (2007), and Martin Eichenbaum, Nir Jaimovich, and Sergio Rebelo (2008) for recent studies.

<sup>2</sup>It is harder to make generalizations about sale prices. Some products feature a relatively stable sale discount; others display sizable variation over time.

<sup>3</sup>Comparisons across euro area countries also reveal that the treatment of sales has a significant bearing on the measured frequency of price adjustment, as discussed in Emmanuel Dhyne et al. (2006).



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

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

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

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

Even though the recent empirical literature on price adjustment has highlighted the importance of sales, macroeconomic models have largely sidestepped the issue. The one exception is Kehoe and Midrigan (2008). In their model, firms face different menu costs depending on whether they make permanent or temporary price changes. Coupled with large but transitory idiosyncratic shocks, this mechanism gives rise to sales in equilibrium.

The plan of the paper is as follows. The model of sales is introduced in Section I, and the equilibrium of the model is characterized in Section II. Section III embeds sales into a DSGE model and analyses the real effects of monetary policy. Section IV presents the two-sector extension of the model. Section V draws some conclusions.

## I. The Model

### A. Households

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

$$(1) \quad \mathcal{U}_\iota(t) = \sum_{\ell=0}^{\infty} \beta^\ell \mathbb{E}_t[\nu(C_{t+\ell}(t)) - \nu(H_{t+\ell}(t))],$$

where  $C(t)$  is consumption of household  $\iota$ 's specific basket of goods (defined below),  $H(t)$  is hours of labor supplied, and  $\beta$  is the subjective discount factor ( $0 < \beta < 1$ ). The function  $\nu(C)$  is strictly increasing and strictly concave in  $C$ ;  $\nu(H)$  is strictly increasing and convex in  $H$ .

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

$$(2a) \quad P_t(t)C_t(t) + M_t(t) + \mathbb{E}_t[\mathcal{A}_{t+1|\iota}A_{t+1}(t)] \\ = W_t H_t(t) + D_t + T_t + M_{t-1}(t) + A_t(t),$$

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

$$(2b) \quad P_t(t)C_t(t) \leq M_{t-1}(t) + T_t.$$

Maximizing lifetime utility (1) subject to the sequence of budget constraints (2a) implies the following first-order conditions for consumption  $C(t)$  and hours  $H(t)$ :

$$(3) \quad \beta \frac{v_c(C_{t+1}(t))}{v_c(C_t(t))} = \mathcal{A}_{t+1|\iota} \frac{P_{t+1}(t)}{P_t(t)}, \quad \text{and} \quad \frac{v_h(H_t(t))}{v_c(C_t(t))} = \frac{W_t}{P_t(t)}.$$

<sup>4</sup>These assumptions on asset markets are standard and play no important role in the model.

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

$$(4) \quad 1 + i_t = (\mathbb{E}_t \mathcal{A}_{t+1|t})^{-1}.$$

The net transfer  $T_t$  is equal to the change in the money supply  $\Delta M_t \equiv M_t - M_{t-1}$ . The cash-in-advance constraint (2b) binds when the nominal interest rate  $i_t$  is positive.

### B. Composite Goods

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

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

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

$$(5) \quad C \equiv \left( \int_{\Lambda} c(\tau, \mathcal{B}(\tau))^{\frac{\epsilon-1}{\epsilon}} d\tau + \int_{\mathcal{T} \setminus \Lambda} \left( \int_{\mathcal{B}} c(\tau, b)^{\frac{\eta-1}{\eta}} db \right)^{\frac{\eta(\epsilon-1)}{\epsilon(\epsilon-1)}} d\tau \right)^{\frac{\epsilon}{\epsilon-1}},$$

where  $c(\tau, b)$  is the household's consumption of brand  $b$  of product type  $\tau$ .<sup>5</sup> It is assumed that  $\eta > \epsilon$ , so bargain hunters are more willing to substitute between different brands of a specific product type than households are to substitute between different product types. Households have a zero elasticity of substitution between brands of a product type for which they are loyal to a particular brand.

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

The elasticities  $\epsilon$  and  $\eta$  are common to all households, as is the form of the consumption aggregator (5). Furthermore, the measure of the set  $\Lambda$  of product types for which a household is loyal to a brand is the same across all households. This measure is denoted by  $\lambda$ , and it is assumed that  $0 < \lambda < 1$ . Hence, each household's preferences feature some mixture of loyal and bargain-hunting behavior for different product types. The particular product types for which a household is loyal, and the particular brands receiving its loyalty, are randomly and independently assigned once and for all with equal probability. For example, one household may be loyal to Corona beer and a bargain hunter for desserts, while another may be loyal to Ben & Jerry's ice cream but a bargain hunter for beer. After aggregation, such idiosyncrasies of households' preferences are irrelevant; all that matters is households' common distribution of loyal and bargain-hunting behavior over the whole set of goods.

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

Let  $p(\tau, b)$  be the price of brand  $b$  of product type  $\tau$  prevailing at a given household's shopping moment. The minimum expenditure required to purchase one unit of the composite good (5) is

$$(6) \quad P = \left( \int_{\Lambda} p(\tau, \mathcal{B}(\tau))^{1-\epsilon} d\tau + \int_{\mathcal{T} \setminus \Lambda} \left( \int_{\mathcal{B}} p(\tau, b)^{1-\eta} db \right)^{\frac{1-\epsilon}{1-\eta}} d\tau \right)^{\frac{1}{1-\epsilon}}.$$

The expenditure-minimizing demand functions are

$$(7) \quad c(\tau, b) = \begin{cases} \left( \frac{p(\tau, b)}{p_B(\tau)} \right)^{-\eta} \left( \frac{p_B(\tau)}{P} \right)^{-\epsilon} & \text{if } \tau \in \mathcal{T} \setminus \Lambda, \text{ where } p_B(\tau) \equiv \left( \int_{\mathcal{B}} p(\tau, b)^{1-\eta} db \right)^{\frac{1}{1-\eta}}, \\ \left( \frac{p(\tau, b)}{P} \right)^{-\epsilon} C & \text{if } \tau \in \Lambda \text{ and } b = \mathcal{B}(\tau), \\ 0 & \text{if } \tau \in \Lambda \text{ and } b \neq \mathcal{B}(\tau), \end{cases}$$

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

As shown later, a firm will not charge the same price for its good at all shopping moments in a given time period. At each moment, it will randomly draw a price from some desired price distribution. When this distribution is common to all firms, the price index for bargain hunters is the same for all product types and at all shopping

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

moments, that is,  $P_B = p_B(\tau)$ , and all household price levels are the same and equal across all shopping moments, that is,  $P(t) = P$ . Thus there is a general price level  $P$  in spite of households' individual consumption baskets all differing.<sup>7</sup>

Given that households share a common price level, have the same preferences (1) over their composite goods and hours, and have the same initial wealth and expected lifetime income, all households choose the same levels of composite consumption and hours, hence,  $C(t) = C$  and  $H(t) = H$  for all  $t$ . Since consumption is the only source of demand in the economy, goods market equilibrium requires  $C = Y$ , where  $Y$  is aggregate output.

### C. Firms

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

$$(8) \quad Q = \mathcal{F}(H),$$

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

$$(9) \quad \mathcal{C}(Q; W) = W \mathcal{F}^{-1}(Q).$$

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

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

At a particular shopping moment, the quantity sold by the producer of good  $(\tau, b)$  at price  $p$  is obtained by aggregating customers' demand functions from (7):

$$\int c(\tau, b) dt = (\lambda p^{-\epsilon} + (1 - \lambda) P_B^{\eta-\epsilon} p^{-\eta}) P^\epsilon C,$$

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

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

<sup>8</sup>It is assumed for simplicity that firms can only hold inventories within a time period.

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

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

$$(10) \quad \mathcal{D}(p; P_B, \mathcal{E}) = (\lambda + (1 - \lambda)v(p; P_B))p^{-\epsilon}\mathcal{E},$$

$$\text{where } v(p; P_B) \equiv \left(\frac{p}{P_B}\right)^{-(\eta-\epsilon)} \text{ and } \mathcal{E} \equiv P^\epsilon C.$$

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

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

$$(11) \quad \mathcal{R}(q; P_B, \mathcal{E}) \equiv q\mathcal{D}^{-1}(q; P_B, \mathcal{E}), \quad \text{with price } p = \mathcal{D}^{-1}(q; P_B, \mathcal{E}),$$

where  $\mathcal{D}^{-1}(q; P_B, \mathcal{E})$  is the inverse demand function corresponding to (10).

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

$$(12) \quad \mathcal{P} = \int_p \mathcal{R}(\mathcal{D}(p; P_B, \mathcal{E}); P_B, \mathcal{E})dF(p) - \mathcal{C}\left(\int_p \mathcal{D}(p; P_B, \mathcal{E})dF(p); W\right),$$

where the first integral aggregates revenue  $\mathcal{R}(q; P_B, \mathcal{E})$  over all shopping moments, and second term is the total cost  $\mathcal{C}(Q; W)$  of producing the whole batch of output  $Q$ , which is equal to demand aggregated over all moments.

Consider a discrete distribution of prices  $\{p_i\}$  with weights  $\{\omega_i\}$ .<sup>10</sup> The first-order conditions for maximizing profits (12) with respect to prices  $p_i$  and weights  $\omega_i$  are

$$(13a) \quad \mathcal{R}'(\mathcal{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}) = \mathcal{C}'\left(\sum_j \omega_j \mathcal{D}(p_j; P_B, \mathcal{E}); W\right) \text{ and}$$

$$(13b) \quad \mathcal{R}(\mathcal{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}) = \aleph + \mathcal{D}(p_i; P_B, \mathcal{E})\mathcal{C}'\left(\sum_j \omega_j \mathcal{D}(p_j; P_B, \mathcal{E}); W\right)$$

$$\text{if } \omega_i > 0; \quad \text{and}$$

<sup>10</sup>It is shown later that restricting attention to discrete distributions is without loss of generality.



$$(13c) \quad \mathcal{R}(\mathcal{D}(p_i; P_B, \mathcal{E}); P_B, \mathcal{E}) \leq \aleph + \mathcal{D}(p_i; P_B, \mathcal{E}) \mathcal{G}'\left(\sum_j \omega_j \mathcal{D}(p_j; P_B, \mathcal{E}); W\right)$$

if  $\omega_i = 0$ ,

where  $\aleph$  is the Lagrangian multiplier attached to the constraint  $\sum_j \omega_j = 1$ . Equation (13a) is the usual marginal revenue equals marginal cost condition, which must hold for any price that receives positive weight. As discussed later, (13b) requires a firm to be indifferent between any prices receiving positive weight, and (13c) requires any price not used to be weakly dominated by some price receiving positive weight.

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

## II. Equilibrium with Flexible Prices

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

### A. Profit-Maximizing Price Distributions

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

However, in the model presented here, firms may prefer to randomize across shopping moments, that is, choose a nondegenerate price distribution. The reason is that the model features a price elasticity that decreases with price, potentially leading to a nonmonotonic marginal revenue, in which case the profit function ceases to be globally concave. This can be seen from the following identity:

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

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

As demand in the model comes from two different sources, loyal customers and bargain hunters, and these groups have different price sensitivities, the price elasticity of demand changes with the composition of a firm's customers. High prices

mean that most bargain hunters have deserted its brand, and the residual mass of loyal customers has a low price elasticity. Low prices put the firm in contention to win over the bargain hunters, but fierce competition among brands for these customers means the price elasticity is high.<sup>11</sup>

The price elasticity  $\zeta(p; P_B)$  implied by the demand function  $\mathcal{D}(p_i; P_B, \mathcal{E})$  in (10) is

$$(14) \quad \zeta(p; P_B) = \frac{\lambda\epsilon + (1 - \lambda)\eta v(p; P_B)}{\lambda + (1 - \lambda)v(p; P_B)}.$$

This price elasticity is a weighted average of  $\epsilon$  and  $\eta$ , with the weight on the larger elasticity  $\eta$  increasing with the purchase multiplier  $v(p; P_B)$ , as defined in (10). The higher is the price  $p$ , the lower is the purchase multiplier, and the smaller is the price elasticity.<sup>12</sup>

Marginal revenue is nonmonotonic when  $\eta$  is sufficiently large relative to  $\epsilon$ . This case is depicted in Figure 2. For very low prices, the price elasticity is approximately constant and equal to  $\eta$  because the bargain hunters are preponderant; for very high prices, it is approximately constant and equal to  $\epsilon$  because only loyal customers remain. In an intermediate region there is a smooth transition between  $\epsilon$  and  $\eta$ , and this increase in price elasticity can be large enough to make marginal revenue positively sloped, although it has its usual negative slope outside this intermediate range.

For some parameters  $\epsilon$ ,  $\eta$ , and  $\lambda$ , firms find it optimal to choose a distribution with two prices: a *normal* high price, and a low *sale* price. Denote these two prices respectively by  $p_N$  and  $p_S$ , and let  $q_N = \mathcal{D}(p_N; P_B, \mathcal{E})$  and  $q_S = \mathcal{D}(p_S; P_B, \mathcal{E})$  be the quantities demanded at a single shopping moment at these prices. The frequency of sales (the fraction of shopping moments when a firm’s good is on sale) is denoted by  $s$ . If  $0 < s < 1$  then both prices must satisfy first-order conditions (13a)–(13b). By eliminating the Lagrangian multiplier  $\aleph$  from (13b), profit maximization requires

$$(15) \quad \mathcal{R}'(q_N; P_B, \mathcal{E}) = \mathcal{R}'(q_S; P_B, \mathcal{E}) = \frac{\mathcal{R}(q_S; P_B, \mathcal{E}) - \mathcal{R}(q_N; P_B, \mathcal{E})}{q_S - q_N} \\ = G'(sq_S + (1 - s)q_N; W).$$

There are three requirements for the optimality of this price distribution, represented graphically in Figure 2. First, marginal revenue must be equalized at both normal and sale prices.<sup>13</sup> Second, the extra revenue generated by having a sale at a particular shopping moment per extra unit sold must be equal to the common marginal revenue. This is represented in the figure by the equality of the two shaded areas bounded between the marginal revenue function and the equilibrium level of marginal cost (the horizontal line MC), and between the quantities  $q_N$  and  $q_S$ . Finally,

<sup>11</sup>This change in price elasticity along the demand function is a less extreme version of a “kinked” demand curve. The difference between the demand function in this paper and the “smoothed-kink” of Miles S. Kimball (1995) is that there, the elasticity increases with price, whereas here it decreases with price. The behavior of the price elasticity here is a consequence of aggregation, not a direct assumption.

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

<sup>13</sup>There is a third point between  $q_N$  and  $q_S$  also associated with the same marginal revenue, but including this point in a firm’s price distribution would violate the second-order conditions for profit maximization.

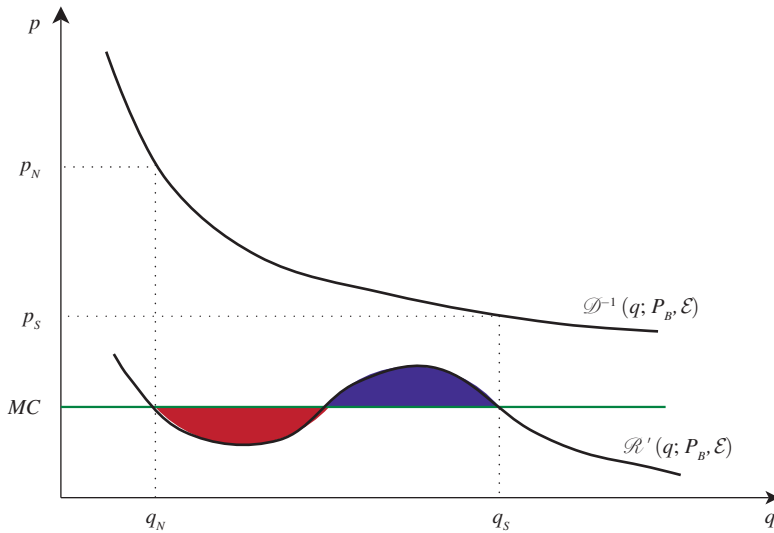


FIGURE 2. DEMAND FUNCTION AND NONMONOTONIC MARGINAL REVENUE FUNCTION

Notes: Schematic representation of demand function  $\mathcal{D}(p; P_B, \mathcal{E})$  from (10) and marginal revenue function  $\mathcal{R}'(p; P_B, \mathcal{E})$  from (11), when  $\eta$  is sufficiently large relative to  $\epsilon$ . The line labeled MC is the equilibrium level of marginal cost.

marginal revenue and average extra revenue must both be equal to the marginal cost of producing total output.

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

The full set of first-order conditions in (15) is depicted using the revenue and total cost functions in Figure 3. As firms can charge different prices at different shopping moments, the set of achievable total revenues is convexified. This raises attainable revenue in the range between  $q_N$  and  $q_S$ . The first two conditions for profit maximization in (15) require that the revenue function has a common tangent line at both quantities  $q_N$  and  $q_S$ , which is equivalent to the slope of the chord being the same as that of the common tangent itself. This slope is then associated with a total quantity sold  $Q = sq_S + (1 - s)q_N$  where marginal cost equals the common value of marginal revenue, which in turn corresponds to a value of the sale frequency  $s$ .

### B. Strategic Interaction

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

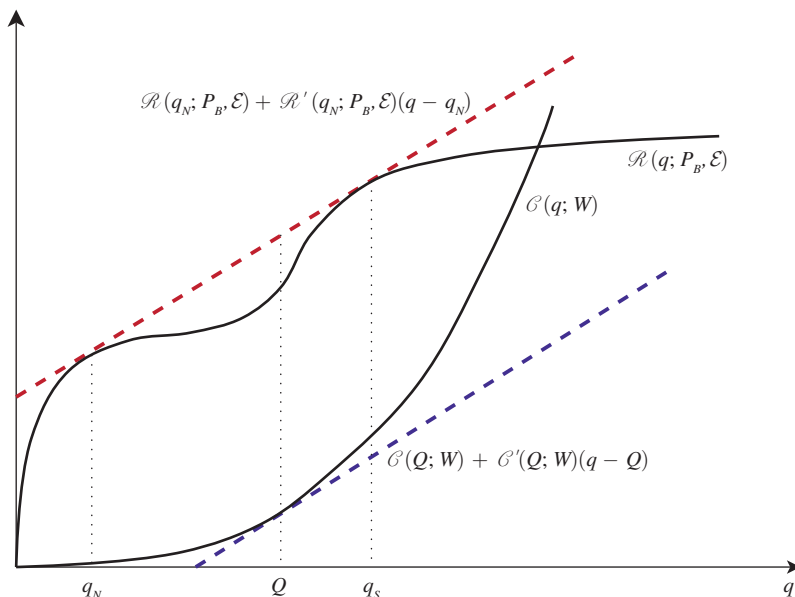


FIGURE 3. REVENUE AND TOTAL COST FUNCTIONS WITH FIRST-ORDER CONDITIONS

Note: Schematic representation of the revenue function  $\mathcal{R}(q; P_B, \mathcal{E})$  from (11) and total cost function  $\mathcal{C}(Q; W)$  from (9), when  $\eta$  is sufficiently large relative to  $\epsilon$ .

However, this reasoning completely neglects the impact of other firms’ actions, and the resulting strategic interaction among firms.

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

The reason why profits at one price relative to the other are affected by others’ sales decisions in the way shown in Figure 4 is apparent from looking at the demand function in (10). For high prices, the first term in  $\lambda$  corresponding to demand from loyal customers is dominant, while for low prices, the second term  $(1 - \lambda)v(p; P_B)$  corresponding to demand from bargain hunters is more important. This is because the purchase multiplier  $v(p; P_B)$  is decreasing in price  $p$ , as demand from bargain hunters is much more sensitive to price. The strategic dimension of this equation comes from

<sup>14</sup>The argument here is based on the case of constant marginal cost, but similar reasoning applies in the general case.

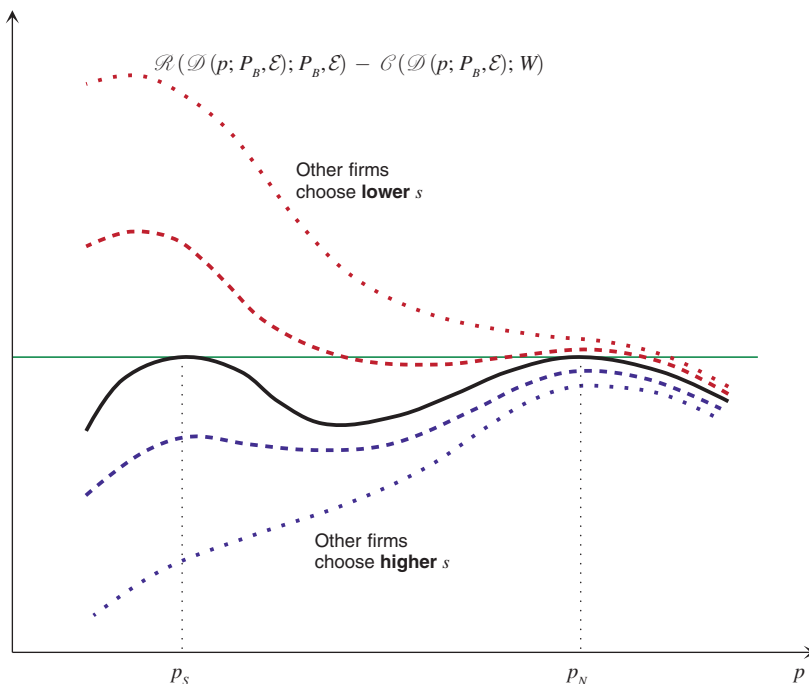


FIGURE 4. PROFITS AT A SINGLE MOMENT, AS AFFECTED BY OTHER FIRMS' SALE FREQUENCIES

Note: Schematic representation of profits as a function of the price charged at a single shopping moment, in the case where the total cost function  $G(Q; W)$  is linear,  $\eta$  is sufficiently large relative to  $\epsilon$ , and  $\lambda$  is not too close to zero or one.

the presence of  $P_B$ . As other firms increase  $s$ ,  $P_B$  falls, which has a negative impact on  $v(p; P_B)$  through demand from bargain hunters, but no effect on demand from loyal customers.<sup>15</sup> Therefore, other firms' sales decisions have a strong effect on profits from selling at low prices, but only a weak effect on profits at high prices.

Conditional on the marginal revenue function being nonmonotonic, this strategic argument for sales depends only on a sufficient mixture of the two types of customer (a value of  $\lambda$  not very close to zero or one). If there were very few of one type of customer, then the maximum attainable profits from a price aimed at the other type might always be larger irrespective of other firms' actions. This is because the value of  $\lambda$  influences the relative height of the two local maxima of profits in addition to the pricing strategies of other firms.

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

<sup>15</sup> Changing  $s$  also affects  $P$ , but this has a proportional effect on both groups' demand and hence on profits at all prices.

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

### C. Discussion

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

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

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

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

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

In addition to temporary sales, the phenomenon of clearance sales has also been analyzed. Understanding the implications of clearance sales requires developing a different model (perhaps along the lines suggested by Edward P. Lazear 1986). But

the typical price pattern shown in Figure 1, which is responsible for the bulk of the divergence between the estimated duration of a price spell in Bils and Klenow (2004) and Nakamura and Steinsson (2008), reflects temporary sales rather than clearance sales.

#### D. Characterizing the Equilibrium

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

**THEOREM 1:** *Marginal revenue  $\mathcal{R}'(q; P_B, \mathcal{E})$  is nonmonotonic (initially decreasing, then increasing on an interval, and then subsequently decreasing) if and only if*

$$(16) \quad \eta > (3\epsilon - 1) + 2\sqrt{2\epsilon(\epsilon - 1)}$$

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

- (i) *If  $\epsilon$  and  $\eta$  satisfy the nonmonotonicity condition (16) and  $\lambda \in (\underline{\lambda}(\epsilon, \eta), \bar{\lambda}(\epsilon, \eta))$  then there exists a two-price equilibrium, and no other equilibria exist.*
- (ii) *If  $\epsilon$  and  $\eta$  violate the nonmonotonicity condition (16) or  $\lambda \notin (\underline{\lambda}(\epsilon, \eta), \bar{\lambda}(\epsilon, \eta))$  then there exists a one-price equilibrium, and no other equilibria exist.*

**PROOF:**

See online Appendix.

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

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

Now the two-price equilibrium is characterized. The total physical quantity of output sold by firms is  $Q = sq_S + (1 - s)q_N$  and the corresponding marginal cost is denoted by  $X \equiv \mathcal{C}'(Q; W)$ . Each of the markups on marginal cost associated with the two prices must satisfy the usual optimality condition in terms of the price elasticity of demand. What is new here is that two markups can satisfy this condition simultaneously. The optimal markup at price  $p$  is  $\mu(p; P_B) = \zeta(p; P_B) / (\zeta(p; P_B) - 1)$ . Using the price elasticity  $\zeta(p; P_B)$  from (14), the first-order conditions for  $p_S$  and  $p_N$  are

$$(17) \quad p_S = \mu(p_S; P_B)X, \quad \text{and} \quad p_N = \mu(p_N; P_B)X,$$

$$\text{with } \mu(p; P_B) = \frac{\lambda\epsilon + (1 - \lambda)\eta v(p; P_B)}{\lambda(\epsilon - 1) + (1 - \lambda)(\eta - 1)v(p; P_B)}.$$

The optimal markup function  $\mu(p; P_B)$  depends on the parameters  $\epsilon$ ,  $\eta$ , and  $\lambda$ , and the purchase multiplier  $v(p; P_B)$  from (10). Let  $v_S \equiv v(p_S; P_B)$  and  $v_N \equiv v(p_N; P_B)$  denote the purchase multipliers at the two prices, and  $\mu_S \equiv \mu(p_S; P_B)$  and  $\mu_N \equiv \mu(p_N; P_B)$  the associated optimal markups:

$$(18) \quad \mu_S = \frac{\lambda\epsilon + (1 - \lambda)\eta v_S}{\lambda(\epsilon - 1) + (1 - \lambda)(\eta - 1)v_S}, \quad \text{and}$$

$$\mu_N = \frac{\lambda\epsilon + (1 - \lambda)\eta v_N}{\lambda(\epsilon - 1) + (1 - \lambda)(\eta - 1)v_N}.$$

The first-order condition for the sale frequency  $s$  is

$$(19) \quad (\mu_S - 1)q_S = (\mu_N - 1)q_N.$$

Given that a fraction  $s$  of all prices are at  $p_S$  and the remaining  $1 - s$  are at  $p_N$  at any shopping moment, equation (7) implies the bargain hunters' price index is

$$(20) \quad P_B = (sp_S^{1-\eta} + (1 - s)p_N^{1-\eta})^{\frac{1}{1-\eta}},$$

which is used to calculate the purchase multipliers and determine the optimal markups  $\mu_S$  and  $\mu_N$ .

In finding the stationary equilibrium, the model has a convenient block-recursive structure, that is, the microeconomic aspects of the equilibrium can be characterized independently of the macroeconomic equilibrium, which is then determined afterward. The key micro variables are the sales frequency  $s$ , the markups  $\mu_S$  and  $\mu_N$ , the markup ratio  $\mu \equiv \mu_S / \mu_N$ , and the ratio of the quantities sold at the sale and normal prices, denoted by  $\chi \equiv q_S / q_N$ .

**PROPOSITION 1:** *Suppose parameters  $\epsilon$ ,  $\eta$ , and  $\lambda$  are such that there is a unique two-price equilibrium.*



- (i) *The first-order conditions in (18) and (19) are necessary and sufficient to characterize the equilibrium price distribution  $(\mu_S, \mu_N, s)$ .*
- (ii) *The equilibrium values of  $\mu$ ,  $\chi$ ,  $\mu_S$ , and  $\mu_N$  are functions only of the parameters  $\epsilon$  and  $\eta$ .*
- (iii) *The equilibrium values of  $s$ ,  $v_S$ , and  $v_N$  are functions only of the parameters  $\epsilon$ ,  $\eta$ , and  $\lambda$ .*
- (iv) *Let  $\underline{\lambda}(\epsilon, \eta)$  and  $\bar{\lambda}(\epsilon, \eta)$  be as defined in Theorem 1:*

$$\frac{\partial s}{\partial \lambda} < 0, \quad \lim_{\lambda \rightarrow \underline{\lambda}(\epsilon, \eta)^+} s = 1, \quad \text{and} \quad \lim_{\lambda \rightarrow \bar{\lambda}(\epsilon, \eta)^-} s = 0.$$

PROOF:

See online Appendix.

The first part of the proposition shows that even though firms are maximizing a nonconcave objective function, the first-order conditions are necessary and sufficient. The second and third parts establish the separation of the equilibrium for the microeconomic variables from the broader macroeconomic equilibrium, that is, the parameters  $\epsilon$ ,  $\eta$ , and  $\lambda$  alone determine  $\mu$ ,  $\chi$ , and  $s$ .<sup>16</sup> The final part shows that the equilibrium sales frequency  $s$  is strictly decreasing in  $\lambda$  and varies from one to zero as  $\lambda$  spans its interval of values consistent with a two-price equilibrium.

Proposition 1 also establishes that the purchase multipliers  $v_S$  and  $v_N$  and the markups  $\mu_S$  and  $\mu_N$  are determined by parameters  $\epsilon$ ,  $\eta$ , and  $\lambda$ ; hence, finding the macroeconomic equilibrium is straightforward. The aggregate price level  $P$  is obtained by combining equation (6) and the demand function (7), and making use of the definition of the purchase multiplier  $v(p; P_B)$  from (10):

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

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

$$(21) \quad x = (s(\lambda + (1 - \lambda)v_S)\mu_S^{1-\epsilon} + (1 - s)(\lambda + (1 - \lambda)v_N)\mu_N^{1-\epsilon})^{\frac{1}{\epsilon-1}}.$$

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

$$(22) \quad q_S = (\lambda + (1 - \lambda)v_S)\rho_S^{-\epsilon}Y, \quad \text{and} \quad q_N = (\lambda + (1 - \lambda)v_N)\rho_N^{-\epsilon}Y.$$

<sup>16</sup>A solution method for  $\mu$ ,  $\chi$ , and  $s$  is described in the online Appendix.

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

$$(23) \quad \Delta \equiv \frac{1}{s(\lambda + (1 - \lambda)v_S)\rho_S^{-\epsilon} + (1 - s)(\lambda + (1 - \lambda)v_N)\rho_N^{-\epsilon}},$$

which satisfies  $0 < \Delta < 1$ . The production function (8), cost function (9), and labor supply function (3) imply a positive relationship between real marginal cost  $x$  and aggregate output  $Y$ :

$$(24) \quad x = \frac{\nu_h(\mathcal{F}^{-1}(Y/\Delta))}{\nu_c(Y)\mathcal{F}'(\mathcal{F}^{-1}(Y/\Delta))}.$$

As the equilibrium real marginal cost  $x$  is already known from (21), the equation above uniquely determines output  $Y$ . Since the cash-in-advance constraint (2b) binds, the aggregate price level  $P$  is then given by  $P = M/Y$ . Finally, the interest rate is  $i = (1 - \beta)/\beta$ .

### III. Flexible Sales with Sticky Normal Prices

#### A. Staggered Adjustment of Normal Prices

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

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

In every time period, each firm has a probability  $1 - \phi_p$  of receiving an opportunity to adjust its normal price. Consider a firm that receives such an opportunity at time  $t$ . The new normal price it selects is referred to as its reset price, and is denoted by  $R_{N,t}$ . All firms that choose new normal prices at the same time choose the same reset price. In any time period, each firm’s optimal sales decisions will in principle depend on its current normal price, and so on its last adjustment time. Denote by  $s_{\ell,t}$  and  $p_{S,\ell,t}$  the optimal sales frequency and sale price for a firm at time  $t$  that last changed its normal price  $\ell$  periods ago (referred to as a vintage- $\ell$  firm). The reset price  $R_{N,t}$  is chosen to maximize the present value of a resetting firm calculated using the profit function (12) and stochastic discount factor  $\mathcal{A}_{t+\ell|t}$ :

$$(25) \quad \max_{R_{N,t}} \sum_{\ell=0}^{\infty} \phi_p^\ell \mathbb{E}_t \left[ \mathcal{A}_{t+\ell|t} \left\{ s_{\ell,t+\ell} p_{S,\ell,t+\ell} \mathcal{D}(p_{S,\ell,t+\ell}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) \right. \right. \\ \left. \left. + (1 - s_{\ell,t+\ell}) R_{N,t} \mathcal{D}(R_{N,t}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) \right. \right. \\ \left. \left. - G(s_{\ell,t+\ell} \mathcal{D}(p_{S,\ell,t+\ell}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}) \right. \right. \\ \left. \left. + (1 - s_{\ell,t+\ell}) \mathcal{D}(R_{N,t}; P_{B,t+\ell}, \mathcal{E}_{t+\ell}); W_{t+\ell} \right\} \right].$$

Using the demand function (10), the total quantity  $Q_{\ell,t}$  sold by a vintage- $\ell$  firm at time  $t$  is

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

$$q_{N,\ell,t} = \mathcal{D}(R_{N,t-\ell}; P_{B,t}, \mathcal{E}_t).$$

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

The first-order condition for the reset price  $R_{N,t}$  maximizing the firm value (25) is

$$(26) \quad \sum_{\ell=0}^{\infty} \phi_p^\ell \mathbb{E}_t \left[ (1 - s_{\ell,t+\ell}) D_{t+\ell|t} \left\{ \frac{R_{N,t}}{P_{t+\ell}} - \mu(R_{N,t}; P_{B,t+\ell}) \frac{X_{\ell,t+\ell}}{P_{t+\ell}} \right\} \right] = 0,$$

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

This condition weights the sequence of one-period optimality conditions for the normal price over the expected lifetime of the price using a discount factor  $D_{t+\ell|t}$ . The profit-maximizing sales frequencies  $s_{\ell,t}$  and sale prices  $p_{S,\ell,t}$  are chosen to maximize profits (25) at all times, yielding first-order conditions:

$$(27) \quad \frac{p_{S,\ell,t} q_{S,\ell,t} - R_{N,t-\ell} q_{N,\ell,t}}{q_{S,\ell,t} - q_{N,\ell,t}} = X_{\ell,t}, \quad \text{and } p_{S,\ell,t} = \mu(p_{S,\ell,t}; P_{B,t}) X_{\ell,t}.$$

Firms' pricing behavior is aggregated as follows. Using equations (6), (7), and (10), an expression for the aggregate price level is

$$(28) \quad P_t = \left( (1 - \phi_p) \sum_{\ell=0}^{\infty} \phi_p^\ell \left\{ s_{\ell,t} (\lambda + (1 - \lambda) v(p_{S,\ell,t}; P_{B,t})) p_{S,\ell,t}^{1-\epsilon} \right. \right. \\ \left. \left. + (1 - s_{\ell,t}) (\lambda + (1 - \lambda) v(R_{N,t-\ell}; P_{B,t})) R_{N,t-\ell}^{1-\epsilon} \right\} \right)^{\frac{1}{1-\epsilon}},$$

and the bargain hunters' price index from (7) is given by

$$(29) \quad P_{B,t} = \left( (1 - \phi_p) \sum_{\ell=0}^{\infty} \phi_p^\ell \left\{ s_{\ell,t} p_{S,\ell,t}^{1-\eta} + (1 - s_{\ell,t}) R_{N,t-\ell}^{1-\eta} \right\} \right)^{\frac{1}{1-\eta}}.$$

Total labor demand from all firms is

$$(30) \quad H_t = \sum_{\ell=0}^{\infty} (1 - \phi_p)\phi_p^\ell H_{\ell,t},$$

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

### B. A Phillips Curve with Sales

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

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

**THEOREM 2:** *Consider parameter values  $\epsilon$ ,  $\eta$ , and  $\lambda$  for which the economy has a two-price equilibrium. Let  $\pi_t \equiv (P_t - P_{t-1})/P_{t-1}$  be the inflation rate for the aggregate price level (28).*

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

$$(31) \quad \rho_{S,\ell,t} = X_{\ell,t}, \quad \text{and } X_{\ell,t} = P_{B,t},$$

*which yield  $\rho_{S,\ell,t} = P_{S,t}$ ,  $X_{\ell,t} = X_t$ , and thus  $Q_{\ell,t} = Q_t$ . The first-order condition for the reset price implies*

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

(ii) *The Phillips curve linking inflation  $\pi_t = P_t - P_{t-1}$  and real marginal cost  $x_t = X_t - P_t$  is*

$$(32) \quad \pi_t = \beta \mathbb{E}_t \pi_{t+1} + \frac{1}{1 - \psi} (\kappa x_t + \psi(\Delta x_t - \beta \mathbb{E}_t \Delta x_{t+1})),$$

where  $\kappa \equiv ((1 - \phi_p)(1 - \beta\phi_p))/\phi_p$ , and the coefficient  $\psi$  is a function only of  $\epsilon$ ,  $\eta$ , and  $\lambda$ . By solving forward, inflation can also be expressed as

$$(33) \quad \pi_t = \frac{\kappa}{1 - \psi} \sum_{\ell=0}^{\infty} \beta^\ell \mathbb{E}_t x_{t+\ell} + \frac{\psi}{1 - \psi} \Delta x_t.$$

<sup>17</sup> All the log deviations of the special features of the sales equilibrium (sale discount, sales frequency, quantity ratio, price distortions) are proportional in equilibrium to the log deviation of real marginal cost. This feature makes the model particularly tractable. More details on the decomposition of aggregate inflation movements are provided by Lemma 4 in the online Appendix.

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

**PROOF:**

See online Appendix.

The first part of the theorem reflects the fact that sales are strategic substitutes. As other firms cut back on sales either by reducing  $s$  or increasing  $p_S$ , the bargain hunters' price index  $P_B$  in (29) increases. This leads a given firm optimally to increase its total quantity sold by holding more sales to the point where marginal cost  $X$  has risen one for one in percentage terms with  $P_B$ .

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

The optimal sale price features a constant markup on marginal cost, at least locally, and the equation determining the optimal reset price is the same as in any standard application of Calvo pricing. The optimal adjustment of sales has the consequence that all firms produce the same total quantity, and thus have the same level of marginal cost.

The Phillips curve with sales in equation (32) would reduce to the standard New Keynesian Phillips curve  $\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa X_t$  in the case that  $\psi = 0$ .<sup>19</sup> On the other hand, the case of a fully flexible price level (a vertical short-run Phillips curve) is equivalent to  $\psi = 1$ . With parameters consistent with sales in equilibrium,  $\psi$  always lies strictly between these extremes. While varying sale frequencies and discounts can always generate the same average price change as a given adjustment of normal prices, in equilibrium, firms never find these to be perfect substitutes and so flexibility in sales never replicates full price flexibility.

The effect of a positive value of  $\psi$  is to increase the response of inflation to real marginal cost to some extent when compared to the New Keynesian Phillips curve. This is best seen by looking at the solved-forward version of the Phillips curve with sales in (33), where there are two distinct differences relative to the solved-forward version of the standard New Keynesian Phillips curve:  $\pi_t = \kappa \sum_{\ell=0}^{\infty} \beta^\ell \mathbb{E}_t X_{t+\ell}$ . The first is a scaling of the coefficient multiplying expected real marginal costs, which is isomorphic to an increase in the probability of price adjustment  $1 - \phi_p$ . The second is the presence of a term in the growth rate of real marginal cost  $\Delta X_t$ . The growth rate appears in addition to the level because the extra margin of price adjustment operates through temporary sales rather than persistent changes to normal prices.

<sup>18</sup>The individual sale and normal prices themselves have only second-order effects on profits by the envelope theorem.

<sup>19</sup>See Michael Woodford (2003) for a derivation and discussion of the standard New Keynesian Phillips curve.

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

### C. A DSGE Model with Sales

This section embeds sales into a DSGE model with staggered adjustment of normal prices and wages.

As in Christopher J. Erceg, Dale W. Henderson, and Andrew T. Levin (2000), firms hire differentiated labor inputs. So hours  $H$  in the production function (8) is now the composite labor input

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

where  $H(t)$  is hours of type- $t$  labor supplied to a given firm, and  $\varsigma$  is the elasticity of substitution between labor types. It is assumed that  $\varsigma > 1$ , and firms are price takers in the markets for labor inputs. The minimum monetary cost of hiring one unit of the composite labor input  $H$  is denoted by  $W$ , and this is now the relevant wage index appearing in firms' cost function (9).

Each household (supplying a particular type of labor) has a probability  $1 - \phi_w$  of being able to adjust its money wage in any given time period. Since households have equal initial financial wealth and expected lifetime income, as asset markets are complete and utility (1) is additively separable between hours and consumption, households are fully insured and hence have equal consumption in equilibrium. Consumption is the only source of expenditure, so goods market equilibrium requires  $C_t = Y_t$ . Thus, by using (3) and (4), and by noting that (2b) is binding, the following intertemporal IS equation and money-market equilibrium condition are obtained:

$$(34) \quad \beta(1 + i_t)\mathbb{E}_t \left[ \frac{v_c(Y_{t+1})}{v_c(Y_t)} \frac{1}{1 + \pi_{t+1}} \right] = 1, \quad \text{and } Y_t = \frac{M_t}{P_t}.$$

The wage setting and wage index equations are as in Erceg, Henderson, and Levin (2000).<sup>21</sup>

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

$$(35) \quad \frac{M_t}{M_{t-1}} = \left( \frac{M_{t-1}}{M_{t-2}} \right)^{\mathfrak{p}} \exp \{ (1 - \mathfrak{p})\mathbf{e}_t \}, \quad \text{where } \mathbf{e}_t \sim \text{i.i.d.}(0, \Omega_m).$$

<sup>20</sup>This point is discussed further in an earlier working paper (Guimaraes and Sheedy 2008).

<sup>21</sup>See the online Appendix for details of these equations.

### D. Calibration

The distinguishing parameters of the sales model are the two elasticities  $\epsilon$  and  $\eta$  and the fraction  $\lambda$  of loyal customers. As shown in Proposition 1, these parameters are directly related to observable prices and quantities: the markup ratio  $\mu$ , which gives the size of the discount offered when a good is on sale; the quantity ratio  $\chi$ , which measures proportionately how much more is purchased when a good is on sale; and the frequency of sales  $s$ . Furthermore, the model has a convenient block recursive structure in that only  $\epsilon$ ,  $\eta$ , and  $\lambda$  need to be known to determine these observables. There are thus three parameters that can be matched to data on just these three variables.<sup>22</sup>

There is a growing empirical literature examining price adjustment patterns at the microeconomic level. This literature provides information about the markup ratio  $\mu$  and the sales frequency  $s$ . The baseline values of these variables are taken from Nakamura and Steinsson (2008). Their study draws on individual price data from the US Bureau of Labor Statistic (BLS) CPI research database, which covers approximately 70 percent of US consumer expenditure. They report that the fraction of price quotes that are sales (weighted by expenditure) is 7.4 percent, so  $s = 0.074$  is used here.<sup>23</sup> They also report that the median difference between the logarithms of the normal and sale prices is 0.295, which yields  $\mu = 0.745$ .

In the retail and marketing literature, there has long been a discussion of the effects of price promotions on demand. This research provides information about the quantity ratio. Papers typically report a range of estimates conditional on factors other than price that affect the impact of a price promotion, for example, advertising. The baseline value of the quantity ratio is obtained from the study by Narasimhan, Neslin, and Subrata Sen (1996). Their results are based on scanner data from a large number of US supermarkets. According to the elasticities they report, a temporary price cut of the size consistent with the sale discount taken from Nakamura and Steinsson (2008) implies a quantity ratio of between approximately four and six if retailers draw their sale to the attention of customers. The baseline number used here is the midpoint of this range, so  $\chi = 5$ .<sup>24</sup>

The three facts about sales are then used to find matching values of the three unknown parameters.<sup>25</sup> The results are shown in Table 1.

The remainder of the calibration is standard, drawing on conventional values from the DSGE literature. The parameter values selected are shown in Table 2. One time period corresponds to one month. The discount factor  $\beta$  is chosen to yield a 3 percent

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

<sup>23</sup>The sales frequency  $s$  is for the whole economy. Certain sectors have higher frequencies of sales and some sectors have none. The implications of such heterogeneity are considered in Section IV.

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

<sup>25</sup>A procedure for calculating the equilibrium values of  $\mu$ ,  $\chi$ , and  $s$  is described in the online Appendix. Given the mapping from parameters to the equilibrium of the model, parameters matching the three stylized facts were found by applying the Nelder-Mead simplex algorithm. An extensive grid search over the elasticities  $\epsilon$  and  $\eta$  was used to verify that no other values are consistent with the targets for  $\mu$  and  $\chi$ . Proposition 1 demonstrates that given  $\epsilon$  and  $\eta$ , there is always one and only one value of  $\lambda$  matching the target sales frequency  $s$ .

TABLE 1—CALIBRATION OF THE MODEL OF SALES

Description	Notation	Value
<i>Stylized facts</i>		
Ratio of the sale price markup to the normal price markup ( $\mu_S/\mu_N$ )	$\mu$	0.075*
Ratio of quantities sold at the sale price and the normal price ( $q_S/q_N$ )	$\chi$	5 <sup>†</sup>
Frequency of sales	$s$	0.074*
<i>Parameters</i>		
Elasticity of substitution between product types	$\epsilon$	3.01
Elasticity of substitution between brands for a bargain hunter	$\eta$	19.7
Fraction of product types for which a household is loyal to a brand	$\lambda$	0.901

*Note:* These parameters are the only values exactly consistent with the three stylized facts about sales.

\**Source:* Nakamura and Steinsson (2008).

†*Source:* Narasimhan, Neslin, and Sen (1996).

annual real interest rate, the intertemporal elasticity of substitution in consumption  $\theta_c$  is chosen to match a coefficient of relative risk aversion of three, and the Frisch elasticity of labor supply  $\theta_h$  is set to 0.7, which lies in the range of estimates found in the literature (Robert E. Hall 2009). The production function is  $\mathcal{F}(H) = AH^\alpha$ , where  $\alpha$  is the elasticity of output with respect to hours. The value of  $\alpha$  is chosen to match a labor share of 0.667. This production function implies that the elasticity  $\gamma$  of marginal cost with respect to output is given by  $\gamma = (1 - \alpha)/\alpha$ . So  $\alpha = 0.667$  yields  $\gamma = 0.5$ . The elasticity of substitution between labor inputs  $\zeta$  is taken from Lawrence J. Christiano, Eichenbaum, and Charles L. Evans (2005). The probability  $\phi_p$  of stickiness of the normal price is set to match an average price-spell duration of nine months, which is taken from Nakamura and Steinsson (2008). The same number is used for the probability of wage stickiness  $\phi_w$ , as evidence shows that most, but not all, wages are adjusted annually. The persistence parameter of money-supply growth  $\mathfrak{p}$  is chosen to match the first-order autocorrelation coefficient of M1 growth in the United States from 1960:1 to 1999:12.

### E. Dynamic Simulations

This section calculates the impulse responses of output and the price level to monetary policy shocks in the calibrated DSGE model with sales. These are compared to the corresponding impulse responses in a standard DSGE model, that is, one where consumers have regular Dixit-Stiglitz preferences and thus firms employ a one-price strategy, and where price-adjustment times are staggered according to the Calvo (1983) model. With Calvo pricing, a standard New Keynesian Phillips curve is obtained. The latter model is set up so that it is otherwise identical to the DSGE model with sales.

The calibrated parameters of the DSGE model with sales are given in Table 1 and Table 2. For the standard DSGE model without sales, the same parameter values from Table 2 are used, with the probability of price stickiness  $\phi_p$  applying to a firm's single price, rather than to its normal price in the sales model. In place of the parameters  $\epsilon$ ,  $\eta$ , and  $\lambda$ , the standard model requires only a calibration of its constant price elasticity of demand  $\xi$  (the elasticity of substitution in the usual Dixit-Stiglitz



TABLE 2—CALIBRATION OF THE DSGE MODEL PARAMETERS

Description	Notation	Value
<i>Preference parameters</i>		
Subjective discount factor	$\beta$	0.9975
Intertemporal elasticity of substitution in consumption	$\theta_c$	0.333
Frisch elasticity of labor supply	$\theta_h$	0.7 <sup>†</sup>
<i>Technology parameters</i>		
Elasticity of output with respect to hours	$\alpha$	0.667
Elasticity of marginal cost with respect to output	$\gamma$	0.5
Elasticity of substitution between differentiated labor inputs	$\varsigma$	20 <sup>‡</sup>
<i>Nominal rigidities</i>		
Probability of stickiness of normal prices	$\phi_p$	0.889 <sup>§</sup>
Probability of wage stickiness	$\phi_w$	0.889
<i>Monetary policy</i>		
First-order serial correlation of the money supply growth rate	$\rho$	0.536 <sup>#</sup>

Note: Monthly calibration.

<sup>†</sup>Source: Hall (2009).

<sup>‡</sup>Source: Christiano, Eichenbaum, and Evans (2005).

<sup>§</sup>Source: Nakamura and Steinsson (2008).

<sup>#</sup>Source: Authors' calculations using data on M1 for the period 1960:1–1999:1. Series M1SL from Federal Reserve Economic Data (<http://research.stlouisfed.org/fred2>).

aggregator). This is chosen to match the average markup (in the sense of the reciprocal of real marginal cost) from the calibrated sales model. For the baseline calibration, this implies  $\xi = 3.77$ .

Figure 5 plots the impulse response functions of aggregate output and the price level to a serially correlated money growth shock in both the sales model and the standard model without sales. The real effects of monetary policy in the model with sales are large and very similar to those found in the standard model, in spite of firms' full freedom to react to the shock by varying sales without cost. The ratio of the cumulated responses of output between the two models is 0.89. Underlying this finding is the modest reaction of sale discounts and the negligible reaction of sales frequencies to the shock.<sup>26</sup>

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

Quantitatively, finding the right balance between targeting their two groups of customers turns out to be much more important to firms' profits than using sales

<sup>26</sup>The impulse responses of the average sale discount and the average sale frequency are both proportional to that of real marginal cost, as shown by Lemma 4 in the online Appendix.

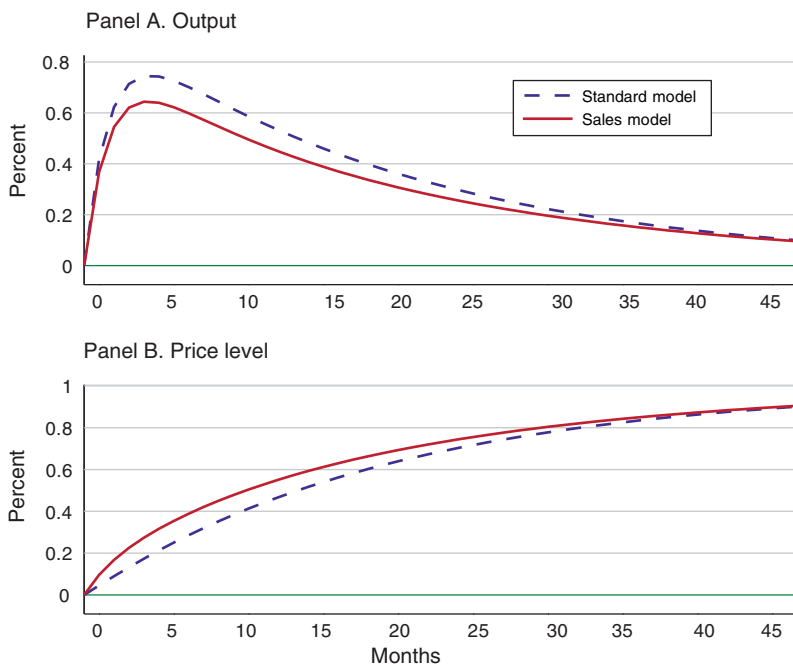


FIGURE 5. IMPULSE RESPONSES TO A PERSISTENT SHOCK TO MONEY GROWTH

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

as a means of changing their average prices. In the data, there is a substantial gap between sale and normal prices on average. So a relatively modest response of sales to a monetary shock would suffice to raise the price level in line with the money supply. However, the larger the gap between the two prices, the greater must be the difference between the two customer types in terms of their price sensitivities, which increases the incentives for a contrarian response to other firms’ pricing strategies. This strong strategic substitutability dissuades firms from adjusting sales because all firms would need to respond in the same way to the aggregate monetary shock.<sup>27</sup>

The role of strategic substitutability can be isolated by considering, instead, an idiosyncratic demand shock to one single firm. Since this one firm is negligible, no other firms react, so the bargain hunters’ price index  $P_B$  does not change. From the first-order condition in (31), the marginal cost of the affected firm must remain unchanged. Hence, the total quantity the firm produces is insulated from the demand shock through its adjustment of sales. This is in stark contrast to the small response of sales to aggregate demand shocks where strategic considerations dominate.

<sup>27</sup> Although firms’ sales are reacting only slightly to monetary shocks, the losses from failing to adjust the normal price more frequently are considerably smaller than they would otherwise be in a model without sales. The possibility of adjusting sales implies that the quantity produced by an individual firm would be exactly the same had this firm the option of adjusting its normal price in addition to adjusting its sales, as is shown in Theorem 2. Hence, there are no undesirable fluctuations in marginal cost, and so the further gains from adjusting the normal price are smaller. This point is discussed further in an earlier working paper (Guimaraes and Sheedy 2008).

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

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

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

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

The intuition for this finding is that when the cost function is linear, marginal cost does not depend on the quantity of output produced. So a rise in aggregate demand, which if accommodated would increase the quantity sold, no longer provides firms with a reason to reduce sales. Hence, all that matters for sales decisions is striking the right balance between targeting loyal customers and bargain hunters.

Figure 7 shows an example of an individual price path in the model, with sales generated using the baseline calibration. Price observations are sampled at a weekly frequency. The underlying stochastic process for the money supply is a random walk with drift. The behavior depicted is qualitatively and quantitatively consistent with real-world examples of prices without needing to assume any idiosyncratic shocks are present.<sup>28</sup>

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

Behind the findings of this section lies the fact that the equilibrium distribution of prices reacts little to monetary shocks. So while the occurrence of sales means that there is much more nominal flexibility of individual prices, the rationale for sales

<sup>28</sup>The model can be reinterpreted in terms of producers choosing a distribution of prices across a continuum of retail outlets, with each outlet maintaining the same price during each discrete time period. In this case, Figure 7 corresponds to the price path at a randomly chosen outlet.

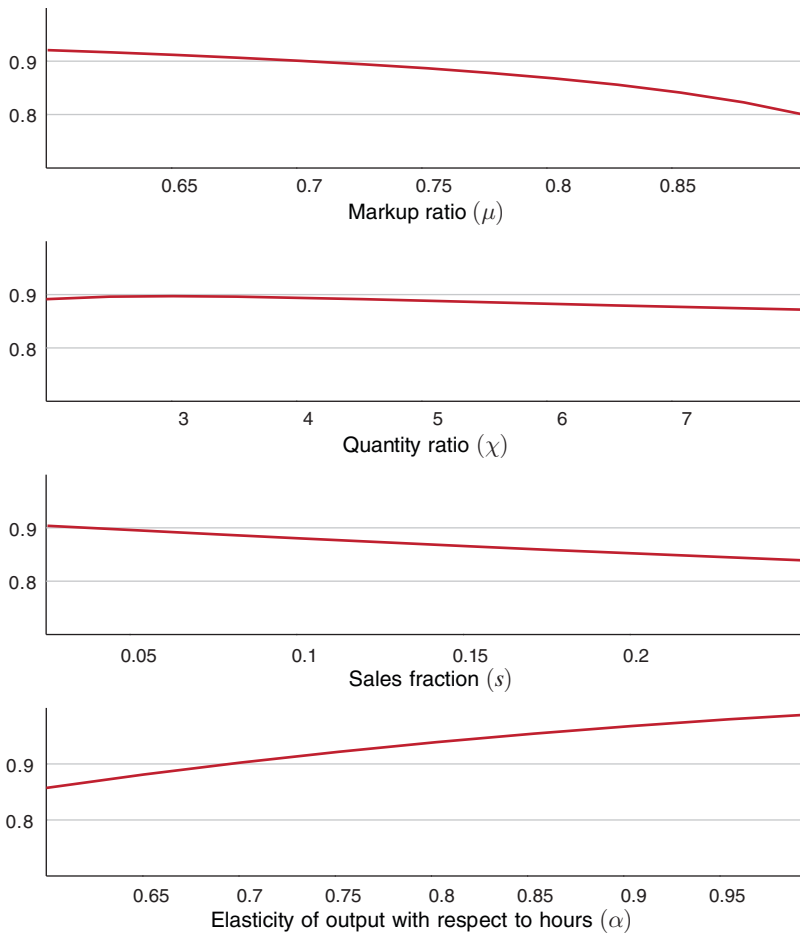


FIGURE 6. SENSITIVITY ANALYSIS FOR THE REAL EFFECTS OF MONETARY SHOCKS:  
RATIO OF CUMULATIVE OUTPUT RESPONSES

*Notes:* For each graph, the results are obtained by fixing the other targets at their baseline values as given in Table 1 (together with  $\alpha = 2/3$ ) and choosing matching values of the parameters  $\epsilon$ ,  $\eta$ , and  $\lambda$  as explained in Section IIID.

implies that there is an endogenous real rigidity constraining the adjustment of the relative prices in firms' price distributions.

#### IV. Sectoral Heterogeneity in Sales

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

The model of Section I is extended to include two sectors. In one sector, households have homogeneous Dixit-Stiglitz preferences over brands of product types, so

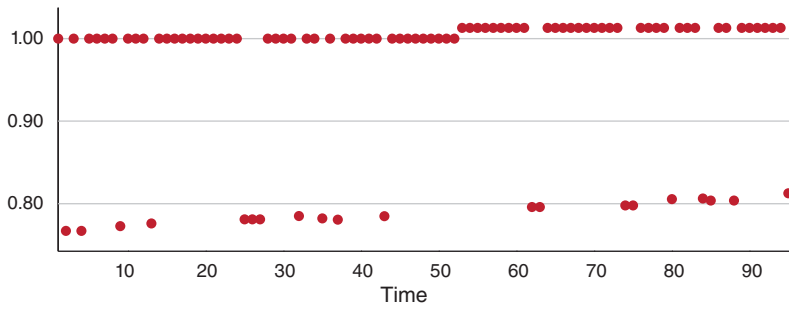


FIGURE 7. A TYPICAL INDIVIDUAL PRICE PATH GENERATED BY THE MODEL

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

no sales will occur in equilibrium. In the other sector, household preferences over brands are heterogeneous, with some mixture of loyal and bargain hunting behavior, which will give rise to sales in equilibrium. This extension is simple and tractable.

A measure  $\sigma$  of product types is in the sale sector, with  $\mathcal{S}$  now denoting the set of just these product types, and where household preferences are as described in Section I. The remaining set of product types with measure  $1 - \sigma$  in the nonsale sector is denoted by  $\mathcal{N}$ . The new composite good  $C$  replacing that in equation (5) is

$$C \equiv \left( \int_{\Lambda} c(\tau, \mathcal{B}(\tau))^{\frac{\epsilon-1}{\epsilon}} d\tau + \int_{\mathcal{S} \setminus \Lambda} \left( \int_{\mathcal{B}} c(\tau, b)^{\frac{\eta-1}{\eta}} db \right)^{\frac{\eta(\epsilon-1)}{\epsilon(\eta-1)}} d\tau \right. \\ \left. + \int_{\mathcal{N}} \left( \int_{\mathcal{B}} c(\tau, b)^{\frac{\xi-1}{\xi}} db \right)^{\frac{\xi(\epsilon-1)}{\epsilon(\xi-1)}} d\tau \right)^{\frac{\epsilon}{\epsilon-1}},$$

where  $\xi$  is the homogeneous elasticity of substitution between brands in the nonsale sector for all households.

Two restrictions are imposed. First, the elasticity  $\xi$  is chosen to ensure the markup in the nonsale sector is equal to the economy's average markup (in the sense of the reciprocal of real marginal cost). This entails choosing  $\xi = 1/(1 - x)$ , where  $x$  is calculated for the sale sector as in (21), as if it encompassed the whole economy. Second, the relative contributions of the sale and nonsale sectors to GDP must be proportional to  $\sigma$  and  $1 - \sigma$ . Given that the sale sector features price distortions, it is not possible to satisfy these two restrictions when the production function is the same in both sectors. Consequently, a slight adjustment is made to the nonsale sector production function  $\mathcal{F}(H)$ , but one which ensures it has the same elasticity of output with respect to hours  $\alpha$  and elasticity of marginal cost with respect to output  $\gamma$  as the production function  $\mathcal{F}(H)$  in the sale sector. These conditions are satisfied only when  $\mathcal{F}(H) = \Delta \mathcal{F}(\Delta^{-1}H)$ , where  $\Delta$  is the stationary equilibrium price distortion (ratio of  $Y$  to  $Q$ ) in the sale sector from (23) (again, calculated as if this sector encompassed the whole economy). Since  $\Delta$  is close to one in practice, the difference between the production functions is very small.

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

**PROPOSITION 2:** *Let  $\Psi(s; \epsilon, \eta)$  be the Phillips curve coefficient  $\psi$  from Theorem 2 implied by a sale frequency  $s$ , with parameters  $\epsilon$  and  $\eta$  consistent with  $\mu$  and  $\chi$ , and with  $\lambda$  implicitly adjusted to match  $s$ , as if the sale sector encompassed the whole economy.*

- (i) *The function  $\Psi(s; \epsilon, \eta)$  is strictly increasing and strictly concave in  $s$ .*
- (ii) *In the case of constant marginal cost ( $\gamma = 0$ ), the Phillips curve for aggregate inflation in the two-sector model is of exactly the same form as that in Theorem 2, with  $\psi$  replaced by the weighted average of  $\Psi(s; \epsilon, \eta)$  (for the sale sector) and 0 (for the nonsale sector) using weights  $\sigma$  and  $1 - \sigma$ . This weighted average is less than  $\Psi(\sigma s; \epsilon, \eta)$  for all  $\sigma < 1$ .*

**PROOF:**

See online Appendix.

The first finding states that  $\Psi(s; \epsilon, \eta)$ , which can be interpreted as the amount of price-level flexibility resulting from adjustment of sales, is increasing in the frequency with which sales occur, but at a diminishing rate. In a multisector context, what matters for aggregate price flexibility is mainly the weighted average of the value of  $\psi$  across sectors. Therefore, by Jensen's inequality, an economy with an unequal distribution of sales across sectors implies a lower average value of  $\psi$ , and thus a flatter aggregate Phillips curve where monetary policy has larger real effects, than an economy with just one sector, but the same average sales frequency. The second finding makes this intuition precise when marginal cost is constant.<sup>29</sup>

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

Eichenbaum, Jaimovich, and Rebelo (2008) study data from a major US retailer and find that prices are below their "reference" level 29 percent of the time on

<sup>29</sup>The equations of the two-sector model in the general case  $\gamma \neq 0$  are presented in the online Appendix. The analysis can easily be extended to an  $n$ -sector model.

TABLE 3—CALIBRATION OF THE TWO-SECTOR MODEL

Description	Notation	Value
<i>Stylized facts</i>		
Frequency of sales in the sale sector	$s$	0.29*
Aggregated frequency of sales	$\bar{s}$	0.074 <sup>†</sup>
<i>Parameters</i>		
Fraction of loyal customers for each brand in the sale sector	$\lambda$	0.735
Size of the sale sector	$\sigma$	0.255

Note: The stylized facts for  $\mu$  and  $\chi$  are as in Table 1 along with the matching parameters values for  $\epsilon$  and  $\eta$ .

\*Source: Eichenbaum, Jaimovich, and Rebelo (2008).

<sup>†</sup>Source: Nakamura and Steinsson (2008).

average. Hence, the target value for  $s$  is 0.29, which yields  $\sigma$ , the size of the sale sector, equal to 0.255 when the economy-wide sale frequency must be the same as the one-sector calibration (Table 1). The calibration exercise is summarized in Table 3.

Figure 8 shows the impulse responses to the same monetary policy experiment described in Section III for the two-sector model with sales and the standard model without any sales. The difference between the impulse responses is even smaller than before. The ratio of the cumulated responses of output is now 0.96, in contrast to 0.89 in the one-sector model. This shows that sales are essentially irrelevant for monetary policy analysis in the two-sector model.

## V. Conclusions

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

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

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

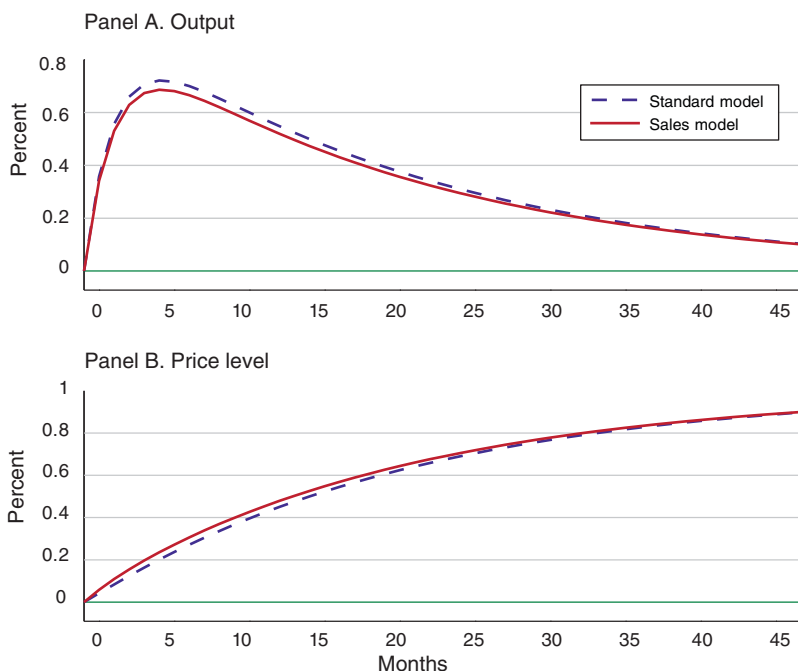


FIGURE 8. IMPULSE RESPONSES TO A PERSISTENT SHOCK TO MONEY GROWTH IN THE TWO-SECTOR MODEL

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

## REFERENCES

- Bils, Mark, and Peter J. Klenow.** 2004. "Some Evidence on the Importance of Sticky Prices." *Journal of Political Economy*, 112(5): 947–85.
- Calvo, Guillermo A.** 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics*, 12(3): 383–98.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans.** 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy*, 113(1): 1–45.
- Dhyne, Emmanuel, Luis J. Álvarez, Hervé Le Bihan, Giovanni Veronese, Daniel Dias, Johannes Hoffmann, Nicole Jonker, Patrick Lunnemann, Fabio Rumler, and Jouko Vilmunen.** 2006. "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data." *Journal of Economic Perspectives*, 20(2): 171–92.
- Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo.** 2008. "Reference Prices and Nominal Rigidities." National Bureau of Economic Research Working Paper 13829.
- Erceg, Christopher J., Dale W. Henderson, and Andrew T. Levin.** 2000. "Optimal Monetary Policy with Staggered Wage and Price Contracts." *Journal of Monetary Economics*, 46(2): 281–313.
- Goldberg, Pinelopi K., and Rebecca Hellerstein.** 2007. "A Framework for Identifying the Sources of Local-Currency Price Stability with an Empirical Application." National Bureau of Economic Research Working Paper 13183.
- Guimaraes, Bernardo, and Kevin D. Sheedy.** 2008. "Sales and Monetary Policy." Center for Economic and Policy Research Discussion Paper 6940.
- Hall, Robert E.** 2009. "Reconciling Cyclical Movements in the Marginal Value of Time and the Marginal Product of Labor." *Journal of Political Economy*, 117(2): 281–323.
- Hosken, Daniel, and David Reiffen.** 2004. "Patterns of Retail Price Variation." *RAND Journal of Economics*, 35(1): 128–46.
- Kehoe, Patrick J., and Virgiliu Midrigan.** 2008. "Temporary Price Changes and the Real Effects of Monetary Policy." National Bureau of Economic Research Working Paper 14392.



- Kimball, Miles S.** 1995. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit, and Banking*, 27(4): 1241–77.
- Klenow, Peter J., and Oleksiy Kryvtsov.** 2008. "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?" *Quarterly Journal of Economics*, 123(3): 863–904.
- Lazear, Edward P.** 1986. "Retail Pricing and Clearance Sales." *American Economic Review*, 76(1): 14–32.
- Levin, Andrew, and Tack Yun.** 2009. "Reconsidering the Microeconomic Foundations of Price-Setting Behavior." Unpublished.
- Nakamura, Emi.** 2008. "Pass-Through in Retail and Wholesale." *American Economic Review*, 98(2): 430–37.
- Nakamura, Emi, and Jon Steinsson.** 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics*, 123(4): 1415–64.
- Nakamura, Emi, and Jon Steinsson.** 2009. "Price Setting in Forward-Looking Customer Markets." Unpublished.
- Narasimhan, Chakravarthi.** 1988. "Competitive Promotional Strategies." *Journal of Business*, 61(4): 427–49.
- Narasimhan, Chakravarthi, Scott A. Neslin, and Subrata K. Sen.** 1996. "Promotional Elasticities and Category Characteristics." *Journal of Marketing*, 60(2): 17–30.
- Neslin, Scott A.** 2002. "Sales Promotion." In *Handbook of Marketing*, ed. Barton Weitz and Robin Wensley, 310–38. Thousand Oaks, CA: SAGE Publications.
- Salop, Steven, and Joseph E. Stiglitz.** 1977. "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion." *Review of Economic Studies*, 44(3): 493–510.
- Salop, S., and J. E. Stiglitz.** 1982. "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents." *American Economic Review*, 72(5): 1121–30.
- Sobel, Joel.** 1984. "The Timing of Sales." *Review of Economic Studies*, 51(3): 353–68.
- Thaler, Richard.** 1985. "Mental Accounting and Consumer Choice." *Marketing Science*, 4(3): 199–214.
- Varian, Hal R.** 1980. "A Model of Sales." *American Economic Review*, 70(4): 651–59.
- Woodford, Michael.** 2003. *Interest and Prices; Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press.