EFCE
7/9/08
9:00 a.m.

Information-Constrained State-Dependent Pricing*

Michael Woodford
Columbia University

June 13, 2008

Abstract

I present a generalization of the standard (full-information) model of state-dependent pricing in which decisions about when to review a firm's existing price must be made on the basis of imprecise awareness of current market conditions. The imperfect information is endogenized using a variant of the theory of "rational inattention" proposed by Sims (1998, 2003, 2006). This results in a one-parameter family of models, indexed by the cost of information, which nests both the standard state-dependent pricing model and the Calvo model of price adjustment as limiting cases (corresponding to a zero information cost and an unboundedly large information cost respectively). For intermediate levels of the information cost, the model is equivalent to a "generalized *Ss* model" with a continuous "adjustment hazard" of the kind proposed by Caballero and Engel (1993a, 1993b), but provides an economic motivation for the hazard function and very specific predictions about its form. For moderate levels of the information cost, the Calvo model of price-setting is found to be a fairly accurate approximation to the exact equilibrium dynamics, except in the case of (infrequent) large shocks.

PRELIMINARY

*I would like to thank Marco Bonomo, Ricardo Caballero, Eduardo Engel, John Leahy, Filip Matejka, Giuseppe Moscarini, Emi Nakamura, Tony Smith, Jon Steinsson and Alex Wolman for helpful discussions; Maxim Pinkovskiy and Luminita Stevens for outstanding research assistance; the NSF for research support through a grant to the NBER; and the Arthur Okun and Kumho Visiting Professorship, Yale University, for providing the time to write this paper.

Models of state-dependent pricing [SDP], in which not only the size of price changes but also their timing is modeled as a profit-maximizing decision on the part of firms, have been the subject of an extensive literature.[1] For the most part, the literature dealing with empirical models of inflation dynamics and the evaluation of alternative monetary policies have been based on models of a simpler sort, in which the size of price changes is modeled as an outcome of optimization, but the timing of price changes is taken as given, and hence neither explained nor assumed to be affected by policy. The popularity of models with exogenous timing [ET] for such purposes stems from their greater tractability, allowing greater realism and complexity on other dimensions. But there has always been general agreement that an analysis in which the timing of price changes is also endogenized would be superior in principle.

This raises an obvious question: how much is endogeneity of the timing of price changes likely to change the conclusions that one obtains about aggregate dynamics? Results available in special cases have suggested that it may matter a great deal. In a dramatic early result, Caplin and Spulber (1987) constructed a tractable example of aggregate dynamics under SDP in which nominal disturbances have no effect whatsoever on aggregate output, despite the fact that individual prices remain constant for substantial intervals of time; and this result depends crucially on variation in the *number* of firms that change their prices in response to a shock, depending on the size of the shock. The Caplin-Spulber example is obviously extremely special; but Golosov and Lucas (2007) find, in numerical analysis of an SDP model calibrated to account for various facts about the probability distribution of individual price changes in U.S. data, that the predicted aggregate real effects of nominal disturbances are quite small, relative to what one might expect based on the average interval of time between price changes. And more recently, Caballero and Engel (2007) consider the real effects of variation in aggregate nominal expenditure in a fairly general class of "generalized Ss models," and show that quite generally, variation in the "extensive margin" of price adjustment (*i.e.*, variation in the number of prices that adjust, as opposed to variation in the amount by which each of these prices changes) implies a smaller real effect of nominal disturbances than would be predicted in an ET model (and hence variation only on the "intensive margin"); they argue that the degree of immediate adjustment of the overall level of prices can easily be several times as large as would

[1] See, for example, Burstein and Hellwig (2007), Dotsey and King (2005), Gertler and Leahy (2007), Golosov and Lucas (2007), Midrigan (2006), and Nakamura and Steinsson (2006) for some recent additions.

be predicted by an ET model.[2]

These results suggest that it is of some urgency to incorporate variation in the extensive margin of price adjustment into models of the real effects of monetary policy, if one hopes to obtain results of any quantitative realism. Yet there is one respect in which one may doubt that the results of standard SDP models are themselves realistic. Such models commonly assume that at each point in time, each supplier has completely precise information about current demand and cost conditions relating to its product, and constantly re-calculates the currently optimal price and the precise gains that would be obtained by changing its price, in order to compare these to the "menu cost" that must be paid to actually change the price. Most of the time no price change is justified; but on the first occasion on which the benefit of changing price becomes as large as the menu cost, a price change will occur. Such an account assumes that it is *only* costs associated with actually changing one's price that are economized on by firms that change prices only infrequently. Instead, studies such as Zbaracki *et al.* (2004) indicate that there are substantial costs associated with information gathering and decisionmaking that are also reduced by a policy of reviewing prices only infrequently. If this is true, the canonical SDP model (or "Ss model"), according to which a price adjustment occurs in any period if and only if a certain adjustment threshold has been reached, should not yield realistic conclusions. In fact, a model that takes account of the costs of gathering and processing information is likely to behave in at least some respects like ET models.[3] The question is to what extent a more realistic model of this kind would yield conclusions about aggregate price adjustment and the real effects of nominal disturbances that are similar to those of ET models, similar to those of canonical SDP models, or different from both.

The present paper addresses this question by considering a model in which the timing of price reviews is determined by optimization subject to an information constraint. The model generalizes the canonical SDP model (which appears as a limiting

[2] An earlier draft of their paper (Caballero and Engel, 2006) proposed as a reasonable "benchmark" that the degree of flexibility of the aggregate price level should be expected to be about three times as great as would be predicted by an ET model calibrated to match the observed average frequency of price changes.

[3] Phelps (1990, pp. 61-63) suggests that ET models may be more realistic than SDP models on this ground. Caballero (1989) presents an early analysis of a way in which costs of information acquisition can justify "time-dependent" behavior, which is further developed by Bonomo and Carvalho (2004) and Reis (2006).

case of the more general model, the case of zero information cost) to allow for costs of obtaining and/or processing more precise information about the current state of the economy, between the intermittent occasions on which full reviews of pricing policy are undertaken. For the sake of simplicity, and to increase the continuity of the present contribution with prior literature, it is assumed that when a firm decides to pay the discrete cost required for a full review of its pricing policy, it obtains *full information* about the economy's state at that moment; hence when price changes occur, they are based on full information, as in canonical SDP models (as well as canonical ET models).[4] However, *between* the occasions on which such reviews occur, the firm's information about current economic conditions is assumed to be much fuzzier; and in particular, the decision *whether to conduct a full review* must be made on the basis of much less precise information than will be available after the review is conducted. As a consequence, prices do not necessarily adjust at precisely the moment at which they first become far enough out of line for the profit increase from a review of pricing policy to justify the cost of such a review.

There are obviously many ways in which one might assume that information is incomplete, each of which would yield somewhat different conclusions. Here I adopt a parsimonious specification based on the concept of "rational inattention" proposed by Sims (1998, 2003, 2006). It is assumed that *all* information about the state of the world is equally available to the decisionmaker — one does not assume that some facts are more easily or more precisely observable than others — but that there is a limit on the decisionmaker's ability to process information of *any* kind, so that the decision is made on the basis of rather little information. The information that the decisionmaker obtains and uses in the decision is, however, assumed to be the information that is *most valuable* to her, given the decision problem that she faces, and subject to a constraint on the *overall rate of information flow* to the decisionmaker. This requires a quantitative measure of the information content of any given indicator that the decisionmaker may observe; the one that I use (following Sims) is based on the *information-theoretic* measure (entropy measure) proposed by

[4] The assumption that full information about current conditions can be obtained by paying a fixed cost also follows the previous contributions of Caballero (1989), Bonomo and Carvalho (2004), and Reis (2006); I depart from these authors in assuming that *partial* information about current conditions is also available between the occasions when the fixed cost is paid. The analysis here also differs from theirs in assuming that access to memory is costly, as discussed further in section 2.1.

Claude Shannon (1948).[5] The degree of information constraint in the model is then indexed by a single parameter, the cost per unit of information (or alternatively, the shadow price associated with the constraint on the rate of information flow). I can consider the optimal scheduling of price reviews under tighter and looser information constraints, obtaining both a canonical SDP model and a canonical ET model as limiting cases; but the more general model treated here introduces only a single additional free parameter (the information cost) relative to a canonical SDP model, allowing relatively sharp predictions.

The generalization of the canonical SDP model obtained here has many similarities with the "generalized Ss model" of pricing proposed by Caballero and Engel (1993a, 2007) and the SDP model with random menu costs of Dotsey, King and Wolman (1999). Caballero and Engel generalize a canonical Ss model of pricing by assuming that the probability of price change is a continuous function of the signed gap between the current log price and the current optimal log price (*i.e.*, the one that would maximize profits in the absence of any costs of price adjustment), and estimate the "adjustment hazard function" that best fits US inflation dynamics with few *a priori* assumptions about what the function may be like. The model of price-adjustment dynamics presented in section 2 below is of exactly the form that they assume. However, the "hazard function" is given an economic interpretation here: the randomness of the decision whether to review one's price in a given period is a property of the *optimal* information-constrained policy. Moreover, the model here makes quite specific predictions about the form of the optimal hazard function: given the specification of preferences, technology and the cost of a review of pricing policy, there is only a one-parameter family of possible optimal hazard functions, corresponding to alternative values of the information cost. For example, Caballero and Engel assume that the hazard function may or may not be symmetric and might equally well be asymmetric in either direction; this is treated as a matter to be determined empirically. In the model developed here, the hazard function is predicted to be asymmetric in a particular way, for any assumed value of the information cost.

Caballero and Engel (1999) propose a structural interpretation of generalized Ss

[5] See, *e.g.*, Cover and Thomas (2006) for further discussion. The appendix of Sims (1998) argues for the appropriateness of the Shannon entropy measure as a way of modeling limited attention. As is discussed further in section 2, the informational constraint assumed here differs from the one proposed by Sims in the way that memory is treated.

adjustment dynamics (in the context of a model of discrete adjustment of firms' capital stocks), in which the cost of adjustment by any given firm is drawn independently (both across firms and over time) from a continuous distribution of possible costs; Dotsey, King and Wolman (1999) [DKW] consider the implications for aggregate price adjustment and the real effects of nominal disturbances of embedding random menu costs of this kind in a DSGE model with monopolistically competitive pricing. The predicted dynamics of price adjustment in the model developed here are essentially the same as in a particular case of the DKW model; there exists a particular distribution for the menu cost under which the DKW model would imply the same hazard function for price changes as is derived here from optimization subject to an information constraint.[6]

However, the present model supplies an alternative interpretation of the randomness of adjustment at the microeconomic level that some may find more appealing than the idea of random menu costs. Moreover, the present model makes much sharper predictions than the DKW model; there is only a very specific one-parameter family of menu-cost distributions under which the DKW model makes predictions consistent with the information-constrained model. Assumptions that appear completely arbitrary under the random-menu-cost interpretation (why is it natural to assume that the menu cost should be i.i.d.?) are here derived as a consequence of optimization. At the same time, assumptions that might appear natural under the random-menu-cost interpretation (a positive lower bound on menu costs, or a distribution with no atoms) can here be theoretically excluded: the optimal hazard function in this model necessarily corresponds to a distribution of menu costs with *an atom at zero.* This has important implications: contrary to the typical prediction of parametric versions of the Caballero-Engel or DKW model, the present model implies that there is always (except in the limit of zero information cost) a positive adjustment hazard even when a firm's current price is *exactly optimal.* This makes the predicted dynamics of price adjustment under the present model more similar to those of the Calvo (1983) model than is true of these other well-known gener-

[6]Like the DKW model, the present model implies in general that the adjustment hazard should be a monotonic function of the amount by which the firm can increase the value of its continuation problem by changing its price. Only in special cases will this allow one to express the hazard as a function of the signed gap between the current log price and the optimal log price, as in the "generalized Ss" framework of Caballero and Engel (1993a, 1993b). Section 3, however, offers an example of explicit microfoundations for such a case.

alizations of the canonical SDP model. It also helps to explain the observation in microeconomic data sets of a large number of very small price changes, as stressed by Midrigan (2006),[7] and increases the predicted real effects of nominal disturbances (for a given overall frequency of price change), for reasons explained by Caballero and Engel (2006).

In fact, the results obtained here suggest that the predictions of ET models may be more reliable, for many purposes, than results from the study of SDP models have often suggested. The Calvo (1983) model of staggered price-setting is derived as a limiting case of the present model (the limit of an unboundedly large information cost); hence this model, often regarded as analytically convenient but lacking in any appealing behavioral foundations, can be given a fully explicit decision-theoretic justification — the quantitative realism of which, relative to other possible specifications, then becomes an empirical matter. Moreover, even in the more realistic case of a positive but finite information cost, the model's prediction about the effects of typical disturbances can be quite similar to those of the Calvo model, as is illustrated numerically below. The present model predicts that the Calvo model will be quite inaccurate in the case of *large* enough shocks — large shocks should trigger immediate adjustment by almost all firms, because even firms that allocate little attention to monitoring current market conditions between full-scale reviews of pricing policy should notice when something dramatic occurs — and in this respect it is surely more realistic than the simple Calvo model. Yet the shocks for which this correction is important may be so large as to occur only infrequently, in which case the predictions of the Calvo model can be quite accurate much of the time.

Section 1 analyzes the optimal price-review policy under an information constraint. I begin by characterizing optimal policy for a single-period problem, to show how the information constraint gives rise to a continuous hazard function in the simplest possible setting. In section 2, this problem is then embedded in an infinite-horizon dynamic setting. Section 3 illustrates the application of the general framework to a specific model of monopolistically competitive price-setting. Section 4 concludes.

[7] Midrigan (2006) proposes an alternative explanation to the one given here for a positive hazard function when the current price is nearly optimal. The present model achieves a similar effect, without the complication of assuming interdependence between price changes for different goods.

1 Rational Inattention and the Optimal Adjustment Hazard

In this section, I consider the decision problem of a firm that chooses when to review its pricing policy, subject to both a fixed cost of conducting such a review and a unit cost of information about market conditions during the intervals between full reviews. In order to show how "rational inattention" of the sort hypothesized by Sims (1998, 2003, 2006) gives rise to a continuous "adjustment hazard" of the kind postulated by Caballero and Engel (1993a, 1993b), it is useful to first consider the information-constrained price-review decision in a simple static context. The characterization given here of the optimal adjustment hazard will then apply directly to the dynamic setting considered in section 2 as well; in the eventual infinite-horizon model, the firm has a decision of this kind to make in *each period*.

1.1 Formulation of the Problem

Let the "normalized price" of a firm i be defined as $q(i) \equiv \log(p(i)/PY)$, where $p(i)$ is the price charged by firm i for its product, P is an aggregate price index, and Y is an index of aggregate output (or aggregate real expenditure), and suppose that the expected payoff[8] to the firm of charging normalized price q is given by a function $V(q)$, which achieves its maximum value at the optimal normalized price

$$q^* \equiv \arg\max_q V(q).$$

I shall assume that $V(q)$ is a smooth, strictly quasi-concave function. By *strict* quasi-concavity, I mean that not only are the sets $\{q|V(q) \geq v\}$ convex for all v, but in addition the sets $\{q|V(q) = v\}$ are of (Lebesgue) measure zero. Strict quasi-concavity implies that there exists a smooth, monotonic transformation $q = \phi(\hat{q})$ such that the function $\hat{V}(\hat{q}) \equiv V(\phi(\hat{q}))$ is not only a concave function, but a *strictly* concave function of $\hat{q}$. In this case, under the further assumption that $V(q)$ achieves a maximum, the maximum q^* must be unique. Moreover, q^* is the unique point at

[8] I need not be specific at this stage about the nature of this payoff. In the eventual dynamic problem considered below, it includes not only profits in the current period (when the price $p(i)$ is charged), but also the implications for expected discounted profits in later periods of having chosen a price $p(i)$ in the current period.

which $V'(q^*) = 0$; and one must have $V'(q) > 0$ for all $q < q^*$, while $V'(q) < 0$ for all $q > q^*$.

We can then define a "price gap" $x(i) \equiv q(i) - q^*$, as in Caballero and Engel, indicating the signed discrepancy between a firm's actual price and the price that it would be optimal for it to charge.[9] Under full information and in the absence of any cost of changing its price, a firm should choose to set $q(i) = q^*$. Let us suppose, though, that the firm must pay a fixed cost $\kappa > 0$ in order to conduct a review of its pricing policy. I shall suppose, as in canonical menu-cost models, that a firm that conducts such a review learns the precise value of the current optimal price, and therefore adjusts its price so that $q(i) = q^*$. A firm that chooses not to review its existing policy instead continues to charge the price that it chose on the occasion of its last review of its pricing policy. The loss from failing to review the policy (or alternatively, the gain from reviewing it, net of the fixed cost) is then given by

$$L(x) \equiv V(q^*) - V(q^* + x) - \kappa, \tag{1.1}$$

as a function of the price gap x that exists prior to the review.

If $V(q)$ is a smooth, strictly quasi-concave function, then $L(x)$ is a smooth, strictly quasi-convex function, with a unique minimum at $x = 0$. Then in the case of full information, the optimal price-review policy is to review the price if and only if the value of x prior to the review is in the range such that $L(x) \geq 0$.[10] The values of x such that a price review occurs will consist of all x *outside* a certain interval, the "zone of inaction," which necessarily includes a neighborhood of the point $x = 0$. The boundaries of this interval (one negative and one positive, in the case that the interval is bounded) constitute the two "Ss triggers" of an "Ss model" of price adjustment.

I wish now to consider instead the case in which the firm does *not* know the value of x prior to conducting the review of its policy. I shall suppose that the firm *does*

[9]It might appear simpler to directly define the normalized price as the price relative to the optimal price, rather than relative to aggregate nominal expenditure, so that the optimal normalized price would be zero, by definition. But the optimal value q^* is something that we need to determine, rather than something that we know at the time of introducing our notation. (Eventually, the function $V(q)$ must be endogenously determined, as discussed in section 2 below.)

[10]The way in which we break ties in the case that $L(x) = 0$ exactly is arbitrary; here I suppose that in the case of indifference the firm reviews its price. In the equilibrium eventually characterized below for the full-information case, values of x for which $L(x) = 0$ exactly occur with probability zero, so this arbitrary choice is of no consequence.

know its existing price, so that it is possible for it to continue to charge that price in the absence of a review; but it does not know the current value of aggregate nominal expenditure PY, and so does not know its *normalized* price, or the gap between its existing price and the currently optimal price. I shall furthermore allow the firm to have *partial* information about the current value of x prior to conducting a review; this is what I wish to motivate as optimal subject to limits on the attention that the firm can afford to pay to market conditions between the occasions when the fixed cost κ is paid for a full review. It is on the basis of this partial information that the decision whether to conduct a review must be made.

Following Sims, I shall suppose that absolutely *any* information about current (or past) market conditions can be available to the firm, as long as the quantity of information obtained by the firm outside of a full review is within a certain finite limit, representing the scarcity of attention, or information-processing capacity, that is deployed for this purpose. The quantity of information obtained by the firm in a given period is defined as in the information theory of Claude Shannon (1948), used extensively by communications engineers. In this theory, the quantity of information contained in a given signal is measured by the reduction in entropy of the decisionmaker's posterior over the state space, relative to the prior distribution. Let us suppose that we are interested simply in information about the current value of the unknown (random) state x, and that the firm's *prior* is given by a density function $f(x)$ defined on the real line.[11] Let $\hat{f}(x|s)$ instead be the firm's posterior, conditional upon observing a particular signal s. The *entropy* associated with a given density function (a measure of the degree of uncertainty with a number of attractive

[11] In section 1.2, we consider what this prior should be, if the firm understands the process that generates the value of x, but has not yet obtained any information about current conditions. For now, the prior is arbitrarily specified as some pre-existing state of knowledge that does not precisely identify the state x.

properties) is equal to[12]

$$-\int f(x)\log f(x)dx,$$

and as a consequence the entropy reduction when signal s is received is given by

$$I(s) \equiv \int \hat{f}(x|s)\log \hat{f}(x|s)dx - \int f(x)\log f(x)dx.$$

The average information revealed by this kind of signal is therefore

$$I \equiv \mathrm{E}_s I(s) \tag{1.2}$$

where the expected value is taken over the set of possible signals that were possible *ex ante*, using the prior probabilities of that each of these signals would be observed.[13] It is this total quantity I that determines the bandwidth (in the case of radio signals, for example), or the *channel capacity* more generally (an engineering limit of any communication system), that must be allocated to the transmission of this signal if the transmission of a signal with a given average information content is to be possible.[14] Sims correspondingly proposes that the limited attention of decisionmakers be modeled by assuming a constraint on the possible size of the average information flow I.

I shall suppose, then, that the firm arranges to observe a signal s before deciding whether to pay the cost κ and conduct a review of its pricing policy. The theory of *rational inattention* posits that both the design of this signal (the set of possible values of s, and the probability that each will be observed conditional upon any given

[12]In information theory, it is conventional to define entropy using logarithms with a base of two, so that the quantity I defined in (1.2) measures information in "bits", or binary digits. (One bit is the amount of information that can be transmitted by the answer to one optimally chosen yes/no question, or by revealing whether a single binary digit is 0 or 1.) I shall instead interpret the logarithm in this and subsequent formulas as a natural logarithm, to allow the elimination of a constant in various expressions. This is an equivalent measure of information, but with a different size of unit: one unit of information under the measure used here (sometimes called a "nat") is equivalent to 1.44 bits of information.

[13]The prior over s is the one implied by the decisionmaker's prior over possible values of x, together with the known statistical relationship between the state x and the signal s that will be received.

[14]Shannon's theorems pertain to the relation between the properties of a given communication channel and the *average* rate at which information can be transmitted over time using that channel, not the amount of information that will be contained in the signal that is sent over any given short time interval.

state x) and the decision about whether to conduct a price review conditional upon the signal observed will be optimal, in the sense of maximizing

$$\bar{L} \equiv \mathrm{E}[\delta(s)L(x)] - \theta I, \tag{1.3}$$

where $\delta(s)$ is a (possibly random) function of s indicating whether a price review is undertaken ($\delta = 1$ when a price review occurs, and $\delta = 0$ otherwise); the expectation operator integrates over possible states x, possible signals s, and possible price-review decisions, under the firm's prior; and $\theta > 0$ is a cost per unit of information of being more informed when making the price-review decision. (This design problem is solved from an *ex ante* perspective: one must decide how to allocate one's attention, which determines what kind of signal one will observe under various circumstances, before learning anything about the current state.)

I have here written the problem as if a firm can allocate an arbitrary amount of attention to tracking market conditions between full price reviews, and hence have an estimate of x of arbitrary precision prior to its decision about whether to conduct the review, if it is willing to pay for this superior information. One might alternatively consider the problem of choosing a partial information structure to maximize $\mathrm{E}[\delta(s)L(x)]$ subject to an upper bound on I. This will lead to exactly the same one-parameter family of informationally-efficient policies, indexed by the value of I rather than by the value of θ. (In the problem with an upper bound on the information used, there will be a unique value of θ associated with each informationally-efficient policy, corresponding to the Lagrange multiplier for the constraint on the value of I; there will be an inverse one-to-one relationship between the value of θ and the value of I.) I prefer to consider the version of the problem in which θ rather than I is given as part of the specification of the environment. This is because decisionmakers have much more attention to allocate than the attention allocated to any one task, and could certainly allocate more attention to aspects of market conditions relevant to the scheduling of reviews of pricing policy, were this of sufficient importance; it makes more sense to suppose that there is a given shadow price of additional attention, determined by the opportunity cost of reducing the attention paid to other matters, rather than a fixed bound on the attention that can be paid to the problem considered here, even if there *is* a global bound on the information-processing capacity of the decisionmaker.

1.2 Characterization of the Solution

I turn now to the solution of this problem, taking as given the prior $f(x)$, the loss function $L(x)$, and the information cost $\theta > 0$. A first observation is that an efficient signal will supply no information other than whether the firm should review its pricing policy.

Lemma 1 *Consider any signalling mechanism, described by a set of possible signals S and conditional probabilities $\pi(s|x)$ for each of the possible signals $s \in S$ in each of the possible states x in the support of the prior f, and any decision rule, indicating for each $s \in S$ the probability $p(s)$ with which a review occurs when signal s is observed. Let $\bar{L}$ be the value of the objective (1.3) implied by this policy on the part of the firm. Consider as well the alternative policy, under which the set of possible signals is $\{0, 1\}$, the conditional probability of receiving the signal 1 is*

$$\pi(1|x) = \int_{s \in S} p(s)\pi(s|x)ds$$

for each state x in the support of f, and the decision rule is to conduct a review with probability one if and only if the signal 1 is observed; and let $\bar{L}^*$ be the value of (1.3) implied by this alternative policy. Then $\bar{L}^* \geq \bar{L}$.

Moreover, the inequality is strict, except if the first policy is one under which either (i) $\pi(s|x)$ is independent of x (almost surely), so that the signals convey no information about the state x; or (ii) $p(s)$ is equal to either zero or one for all signals that occur with positive probability, and the conditional probabilities are of the form

$$\pi(s|x) = \pi(s|p(s)) \cdot \pi(p(s)|x),$$

where the conditional probability $\pi(s|p(s))$ of a given signal s being received, given that the signal will be one of those for which $p(s)$ takes a certain value, is independent of x (almost surely). That is, either the original signals are completely uninformative; or the original decision rule is deterministic (so that the signal includes a definite recommendation as to whether a price review should be undertaken) and any additional information contained in the signal, besides the implied recommendation regarding the price-review decision, is completely uninformative.

A proof is given in Appendix A. Note that this result implies that we may assume, without loss of generality, that an optimal policy involves only two possible signals, $\{0, 1\}$, and a decision rule under which a review is scheduled if and only if the signal 1 is received. That is, the only signal received is an indication whether it is time to review the firm's existing price or not. (If the firm arranges to receive any more information than this, it is wasting its scarce information-processing capacity.) A policy of this form is completely described by specifying the *hazard function* $\Lambda(x) \equiv \pi(1|x)$, indicating the probability that a price review occurs, in the case of any underlying state x in the support of f.

It follows from Lemma 1 that any randomization that is desired in the price-review decision should be achieved by arranging for the *signal* about market conditions to be random, rather than through any randomization by the firm after receiving the signal. This does not, however, imply in itself that the signal that determines the timing of price reviews should be random, as in the Calvo model (or the "generalized Ss model" of Caballero and Engel). But in fact one can show that it *is* optimal for the signal to be random, under extremely weak conditions.

Let us consider the problem of choosing a measurable function $\Lambda(x)$, taking values on the interval $[0, 1]$, so as to maximize (1.3). One must first be able to evaluate (1.3) in the case of a given hazard function. This is trivial when $\Lambda(x)$ is (almost surely) equal to either 0 or 1 for all x, as in either case the information content of the signal is zero. Hence $\bar{L} = E[L(x)]$ if $\Lambda(x) = 1$ (a.s), and $\bar{L} = 0$ if $\Lambda(x) = 0$ (a.s.). After disposing of these trivial cases, we turn to the case in which the prior probability of a price review

$$\bar{\Lambda} \equiv \int \Lambda(x) f(x) dx \tag{1.4}$$

takes an interior value, $0 < \bar{\Lambda} < 1$. As there are only two possible signals, there are two possible posteriors, given by

$$\hat{f}(x|0) = \frac{f(x)(1 - \Lambda(x))}{1 - \bar{\Lambda}}, \qquad \hat{f}(x|1) = \frac{f(x)\Lambda(x)}{\bar{\Lambda}}$$

using Bayes' Law. The information measure I is then equal to

$$\begin{aligned} I &= \bar{\Lambda} I(1) + (1 - \bar{\Lambda}) I(0) \\ &= \bar{\Lambda} \int \hat{f}(x|1) \log \hat{f}(x|1) dx + (1 - \bar{\Lambda}) \int \hat{f}(x|0) \log \hat{f}(x|0) dx - \int f(x) \log f(x) dx \\ &= \int \varphi(\Lambda(x)) f(x) dx - \varphi(\bar{\Lambda}), \end{aligned} \tag{1.5}$$

where

$$\varphi(\Lambda) \equiv \Lambda \log \Lambda + (1-\Lambda)\log(1-\Lambda) \tag{1.6}$$

in the case of any $0 < \Lambda < 1$, and we furthermore define[15]

$$\varphi(0) = \varphi(1) = 0.$$

We can therefore rewrite the objective (1.3) in this case as

$$\bar{L} = \int [L(x)\Lambda(x) - \theta\varphi(\Lambda(x))] f(x)dx + \theta\varphi(\int \Lambda(x) f(x)dx). \tag{1.7}$$

Given the observation above about the trivial cases, the same formula applies as well when $\bar{\Lambda}$ is equal to 0 or 1. Hence (1.7) applies in the case of any measurable function $\Lambda(x)$ taking values in $[0,1]$, and our problem reduces to the choice of $\Lambda(x)$ to maximize (1.7).

This is a problem in the calculus of variations. Suppose that we start with a function $\Lambda(x)$ such that $0 < \bar{\Lambda} < 1$, and let us consider the effects of an infinitesimal variation in this function, replacing $\Lambda(x)$ by $\Lambda(x) + \delta\Lambda(x)$, where $\delta\Lambda(x)$ is a bounded, measurable function indicating the variation. We observe that

$$\delta\bar{L} = \int \partial(x) \cdot \delta\Lambda(x)\, f(x)dx$$

where

$$\partial(x) \equiv L(x) - \theta\varphi'(\Lambda(x)) + \theta\varphi'(\bar{\Lambda}).$$

A first-order condition for (local) optimality of the policy is then at each point x (almost surely[16]), one of the following conditions holds: either $\Lambda(x) = 0$ and $\partial(x) \leq 0$; $\Lambda(x) = 1$ and $\partial(x) \geq 0$; or $0 < \Lambda(x) < 1$ and $\partial(x) = 0$. We can furthermore observe from the behavior of the function $\varphi'(\Lambda) = \log(\Lambda/1-\Lambda)$ near the boundaries of the domain that

$$\lim_{\Lambda(x)\to 0} \partial(x) = +\infty, \qquad \lim_{\Lambda(x)\to 1} \partial(x) = -\infty,$$

[15] This definition follows Shannon (1948); our $\varphi(\Lambda)$ is the negative of his "binary entropy function." Note that under this extension of the definition of $\varphi(\Lambda)$ to the boundaries of its domain, the function is continuous on the entire interval. Moreover, under this definition, (1.5) is a correct measure of the information content of the signal (namely, zero) even in the case that one of the signals occurs with probability zero.

[16] Note that we can only expect to determine the optimal hazard function $\Lambda(x)$ up to arbitrary changes on a set of values of x that occur with probability zero under the prior, as such changes have no effect on any of the terms in the objective (1.7).

so that neither of the first two conditions can ever hold. Hence the first-order condition requires that

$$\partial(x) = 0 \tag{1.8}$$

almost surely.

This condition implies that

$$\frac{\Lambda(x)}{1 - \Lambda(x)} = \frac{\bar{\Lambda}}{1 - \bar{\Lambda}} \exp\left\{\frac{L(x)}{\theta}\right\} \tag{1.9}$$

for each x. Condition (1.9) implicitly defines a measurable function $\Lambda(x) = \Lambda^*(x; \bar{\Lambda})$ taking values in $(0, 1)$.[17] It is worth noting that in this solution, for a fixed value of $\bar{\Lambda}$, $\Lambda(x)$ is monotonically increasing in the value of $L(x)/\theta$, approaching the value 0 for large enough negative values of $L(x)/\theta$, and the value 1 for large enough positive values; and for given x, $\Lambda^*(x; \bar{\Lambda})$ is an increasing function of $\bar{\Lambda}$, approaching 0 for values of $\bar{\Lambda}$ close enough to 0, and 1 for values of $\bar{\Lambda}$ close enough to 1. We can extend the definition of this function to extreme values of $\bar{\Lambda}$ by defining

$$\Lambda^*(x; 0) = 0, \qquad \Lambda^*(x; 1) = 1$$

for all values of x; when we do so, $\Lambda^*(x; \bar{\Lambda})$ remains a function that is continuous in both arguments.

The above calculation implies that in the case of any (locally) optimal policy for which $0 < \bar{\Lambda} < 1$, the hazard function must be equal (almost surely) to a member of the one-parameter family of functions $\Lambda^*(x; \bar{\Lambda})$. It is also evident (from definition (1.4) and the bounds that $\Lambda(x)$ must satisfy) that if $\bar{\Lambda}$ takes either of the extreme values 0 or 1, the hazard function must satisfy $\Lambda(x) = \bar{\Lambda}$ almost surely; hence the hazard function would be equal (almost surely) to a member of the one-parameter family in these cases as well. We can therefore conclude that the optimal hazard function must belong to this family; it remains only to determine the optimal value of $\bar{\Lambda}$.

In this discussion, $\bar{\Lambda}$ has been used both to refer to the value defined in (1.4) and to index the members of the family of hazard functions defined by (1.9). In fact, the same numerical value of $\bar{\Lambda}$ must be both things. Hence we must have

$$J(\bar{\Lambda}) = \bar{\Lambda}, \tag{1.10}$$

[17] We can easily give a closed-form solution for this function: $\Lambda^*(x; \bar{\Lambda}) = R/1 + R$, where R is the right-hand side of (1.9).

where

$$J(\bar{\Lambda}) \equiv \int \Lambda^*(x;\bar{\Lambda}) f(x) dx. \tag{1.11}$$

Condition (1.10) necessarily holds in the case of a locally optimal policy, but it does not guarantee that $\Lambda^*(x;\bar{\Lambda})$ is even locally optimal. We observe from the definition that $J(0) = 0$ and $J(1) = 1$, so $\bar{\Lambda} = 0$ and $\bar{\Lambda} = 1$ are always at least two solutions to equation (1.10); yet these need not be even local optima.

We can see this by considering the function $\bar{L}(\bar{\Lambda})$, obtained by substituting the solution $\Lambda^*(x;\bar{\Lambda})$ defined by (1.9) into the definition (1.7). Since any locally optimal policy must belong to this one-parameter family, an optimal policy corresponds to a value of $\bar{\Lambda}$ that maximizes $\bar{L}(\bar{\Lambda})$. Differentiating this function, we obtain

$$\begin{aligned}
\bar{L}'(\bar{\Lambda}) &= \int [L(x) - \theta\varphi'(\Lambda^*(x))]\Lambda^*_{\bar{\Lambda}}(x) f(x) dx + \theta\varphi'(J(\bar{\Lambda})) \int \Lambda^*_{\bar{\Lambda}}(x) f(x) dx \\
&= \theta[\varphi'(J(\bar{\Lambda})) - \varphi'(\bar{\Lambda})] \int \Lambda^*_{\bar{\Lambda}}(x) f(x) dx,
\end{aligned}$$

at any point $0 < \bar{\Lambda} < 1$, where $\Lambda^*_{\bar{\Lambda}}(x) > 0$ denotes the partial derivative of $\Lambda^*(x;\bar{\Lambda})$ with respect to $\bar{\Lambda}$, and we have used the first-order condition $\partial(x) = 0$, satisfied by any hazard function in the family defined by (1.9), to obtain the second line from the first. Since

$$\int \Lambda^*_{\bar{\Lambda}}(x) f(x) dx > 0,$$

it follows that $\bar{L}'(\bar{\Lambda})$ has the same sign as $\varphi'(J(\bar{\Lambda})) - \varphi'(\bar{\Lambda})$, which (because of the monotonicity of $\varphi'(\Lambda)$), has the same sign as $J(\bar{\Lambda}) - \bar{\Lambda}$.

Hence a value of $\bar{\Lambda}$ that satisfies (1.10) corresponds to a critical point of $\bar{L}(\bar{\Lambda})$, but not necessarily to a local maximum. The complete set of necessary and sufficient conditions for a local maximum are instead that $\Lambda(x)$ be a member of the one-parameter family of hazard functions defined by (1.9), for a value of $\bar{\Lambda}$ satisfying (1.10), and such that in addition, (i) if $\bar{\Lambda} > 0$, then $J(\Lambda) > \Lambda$ for all Λ in a left neighborhood of $\bar{\Lambda}$; and (ii) if $\bar{\Lambda} < 1$, then $J(\Lambda) < \Lambda$ for all Λ in a right neighborhood of $\bar{\Lambda}$. The argument just given only implies that there must exist solutions with this property, and that they correspond to at least locally optimal policies. In fact, however, there is necessarily a unique solution of this form, and it corresponds to the global optimum, owing to the following result.

Lemma 2 *Let the loss function $L(x)$, the prior $f(x)$, and the information cost $\theta > 0$ be given, and suppose that $L(x)$ is not equal to zero almost surely [under the measure defined by f].*[18] *Then the function $J(\Lambda)$ has a graph of one of three possible kinds: (i) if*

$$\int \exp\left\{\frac{L(x)}{\theta}\right\} f(x)dx \leq 1, \qquad \int \exp\left\{-\frac{L(x)}{\theta}\right\} f(x)dx > 1,$$

then $J(\Lambda) < \Lambda$ for all $0 < \Lambda < 1$ [as in the first panel of Figure 1], and the optimal policy corresponds to $\bar{\Lambda} = 0$; (ii) if

$$\int \exp\left\{-\frac{L(x)}{\theta}\right\} f(x)dx \leq 1, \qquad \int \exp\left\{\frac{L(x)}{\theta}\right\} f(x)dx > 1,$$

then $J(\Lambda) > \Lambda$ for all $0 < \Lambda < 1$ [as in the second panel of Figure 1], and the optimal policy corresponds to $\bar{\Lambda} = 1$; and (iii) if

$$\int \exp\left\{\frac{L(x)}{\theta}\right\} f(x)dx > 1, \qquad \int \exp\left\{-\frac{L(x)}{\theta}\right\} f(x)dx > 1,$$

then there exists a unique interior value $0 < \bar{\bar{\Lambda}} < 1$ at which $J(\bar{\bar{\Lambda}}) = \bar{\bar{\Lambda}}$, while $J(\Lambda) > \Lambda$ for all $0 < \Lambda < \bar{\bar{\Lambda}}$, and $J(\Lambda) < \Lambda$ for all $\bar{\bar{\Lambda}} < \Lambda < 1$ [as in the third panel of Figure 1], and the optimal policy corresponds to $\bar{\Lambda} = \bar{\bar{\Lambda}}$.

The proof is again in Appendix A. Note that the three cases considered in the lemma exhaust all possibilities, as it is not possible for both of the integrals to simultaneously have a value no greater than 1 (in the case that $L(x)$ is not equal to zero almost surely), as a consequence of Jensen's Inequality. Thus we have given a complete characterization of the optimal policy.

Our results also provide a straightforward approach to computation of the optimal policy, once the loss function $L(x)$, the prior $f(x)$, and the value of θ are given. Given $L(x)$ and θ, (1.9) allows us to compute $\Lambda^*(x; \bar{\Lambda})$ for any value of $\bar{\Lambda}$; given $f(x)$, it is then straightforward to evaluate $J(\Lambda)$ for any $0 < \Lambda < 1$, using (1.11). Finally, once one plots the function $J(\Lambda)$, it is easy to determine the optimal value $\bar{\Lambda}$; Lemma 2 guarantees that a simple bisection algorithm will necessarily converge to the right fixed point, as discussed in Appendix B.

[18] This is a very weak assumption. Note that it would be required by the assumption invoked earlier, that $L(x)$ is strictly quasi-concave. But in fact, since $L(0) = -\kappa$, it suffices that the loss function be continuous at zero and that $f(x)$ be positive on a neighborhood of zero, though even these conditions are not necessary.

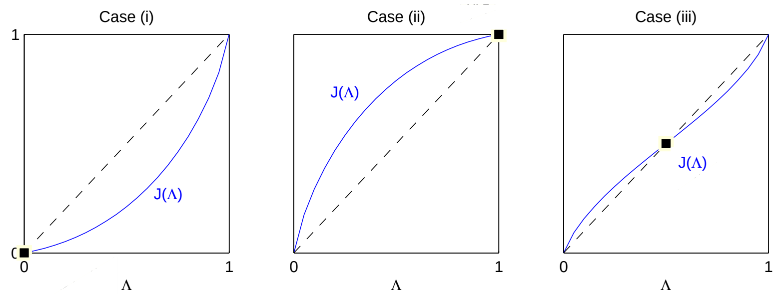

Figure 1: The three possible shapes of the function $J(\Lambda)$, as explained in Lemma 2. In each case, the optimal value of $\bar{\Lambda}$ is indicated by the black square.

1.3 Discussion

We can now see that the optimal signalling mechanism necessarily involves randomization, as remarked earlier. In any case in which it is optimal neither to *always* review one's price nor to *never* review one's price, so that the average frequency with which price reviews occur is some $0 < \bar{\Lambda} < 1$, the optimal hazard function satisfies $0 < \Lambda(x) < 1$, so that a price review may or may not occur, in the case of *any* current price gap x.[19] This is not simply an assumption. We have allowed for the possibility of a hazard function which takes the value 0 on some interval (the "zone of inaction") in which x falls with a probability $1 - \bar{\Lambda}$, and the value 1 everywhere outside that interval; but this can never be an optimal policy. Hence an optimal signalling mechanism never provides a signal that is a *deterministic* function of the true state.

One can also easily show that our assumption that the signal must be a random function of the current state x alone; that is, the randomness in the relation between the observed signal and the value of x must be purely uninformative about the state of the world — it must represent noise in the measurement process itself, rather than systematic dependence on some other aspect of the current (or past) state of the world. We could easily consider a mechanism in which the probability of receiving a given signal s may depend on both x and some other state y. (Statement of the problem then requires that the prior $f(x, y)$ over the joint distribution of the two states be specified.) The same argument as above implies that an optimal policy can be described by a hazard function $\Lambda(x, y)$, and that the optimal hazard function will again be of the form (1.9), where one simply replaces the argument x by (x, y)

[19]As usual, the qualification "almost surely" must be added.

everywhere. In the case that the value function depends only on the state x, as assumed above, the loss function will also be a function simply of x; hence (1.9) implies that the optimal hazard will depend only on x, and that it will be the function of x characterized above.

Among the consequences of this result is the fact that the random signals received by different firms, each of which has the same prior $f(x)$ about its current price gap, will be distributed *independently* of one another, as assumed in the Calvo model. If the signals received by firms were instead correlated (for example, if with probability $\bar{\Lambda}$ all firms receive a signal to review their prices, while with probability $1 - \bar{\Lambda}$ none of them do), then each firm's signal would convey information about *other firms' signals*, and also about their actions. Such signals would therefore convey more information (and, under our assumption about the cost of information, necessarily be more costly) than uncorrelated signals, without being any more useful to the firms in helping them to make profit-maximizing decisions; the correlated signals would therefore not represent an efficient signalling mechanism.[20] Hence the present model predicts that while the price-review decision is random at the level of an individual firm, the *fraction* of such firms that will review their prices in aggregate (assuming a large enough number of firms for the law of large numbers to apply) will be $\bar{\Lambda}$ with certainty.

The present model provides a decision-theoretic justification for the kind of "generalized Ss" behavior proposed by Caballero and Engel (1993a, 1993b) as an empirical specification. The interpretation is different from the hypothesis of a random menu cost in Caballero and Engel (1999) and Dotsey, King and Wolman (1999), but the present model is observationally equivalent to a random-menu-cost model, in the case that the distribution of menu costs belongs to a particular one-parameter family. Suppose that firm has perfect information, but that the menu cost $\tilde{\kappa}$ is drawn from a distribution with cumulative distribution function $G(\tilde{\kappa})$, rather than taking a certain positive value κ with certainty. Then a firm with price gap x should choose to revise

[20] Of course, this result depends on an assumption that, as in the setup assumed by Caballero and Engel (1993a, 2007), the payoff to a firm depends only on its *own* normalized price, and not also on the relation between its price and the prices of other imperfectly attentive firms; to the extent that information about others' actions is payoff-relevant, an optimal signalling mechanism *will* involve correlation.

its price if and only if

$$V(q^*) - \tilde{\kappa} \geq V(q^* + x),$$

which occurs with probability

$$\Lambda(x) = G(V(q^*) - V(q^* + x)) = G(L(x) + \kappa), \tag{1.12}$$

where once again $L(x)$ is the loss function (1.1) of a firm with constant menu cost κ. Thus (1.12) is the hazard function implied by a random-menu-cost model; the only restriction implied by the theory is that $\Lambda(x)$ must be a non-decreasing function of the loss $L(x)$. The present theory also implies that $\Lambda(x)$ should be a non-decreasing function of $L(x)$, as (1.9) has this property for each value of $\bar{\Lambda}$. In fact, the optimal hazard function under rational inattention is identical to the hazard function of a random-menu-cost model in which the distribution of possible menu costs is given by[21]

$$G(\tilde{\kappa}) = 1 - \left[1 + \left(\frac{\bar{\Lambda}}{1 - \bar{\Lambda}}\right) \exp\left\{\frac{\tilde{\kappa} - \kappa}{\theta}\right\}\right]^{-1}. \tag{1.13}$$

While the present model does not imply behavior inconsistent with a random-menu-cost model, it makes much sharper predictions. Moreover, not only does the present model correspond to a single very specific one-parameter family of possible distributions of menu costs, but these distributions are all fairly different from what is usually assumed in calibrations of random-menu-cost models. In particular, a distribution of the form (1.13) necessarily has an atom at zero, so that the hazard is bounded away from zero even for values of x near zero; it has instead been common in numerical analyses of generalized Ss models to assume that in a realistic specification there should be no atom at zero, so that $\Lambda(0) = 0$. The fact that the present model instead implies that $\Lambda(0)$ is necessarily positive (if price reviews occur with any positive frequency) — and indeed, may be a substantial fraction of the average frequency $\bar{\Lambda}$ — is an important difference; under the rule of thumb discussed by Caballero and Engel (2006), it reduces the importance of the "extensive margin" of price adjustment, and hence makes the predictions of a generalized Ss model more similar to those of the Calvo model.

The random-menu-cost model also provides no good reason why, in a dynamic extension of the model, the adjustment hazard should depend only on the current

[21]Note that in this formula, κ is a parameter of the distribution, not the size of the menu cost.

price gap x, and not also on the time elapsed since the last price review. This case is *possible*, of course, if one assumes that the menu cost $\tilde{\kappa}$ is drawn *independently* each period from the distribution G. But there is no reason to assume such independence, and the specification does not seem an especially realistic one (though obviously convenient from the point of view of empirical tractability), if the model is genuinely about exogenous time variation in the cost of changing one's price. The theory of rational inattention instead *requires* that the hazard rate depend only on the current state x, as long as the dynamic decision problem is one in which both the prior and the value function are stationary over time (rather than being duration-dependent), as in the dynamic model developed in the next section.

2 A Dynamic Model of the Timing of Discrete Adjustments

In the static analysis of section 1, both the prior $f(x)$ and the value function $V(q)$ are taken as given. In fact, in a dynamic model of price adjustment, a similar decision about whether to undertake a price review or not must be made each period. But the prior in a given period will not be independent of the firm's price-review policies in previous periods; a past policy of frequently reviewing one's price would make it less likely that the current price gap is large. Nor will the value of continuing with a given normalized price be independent of the firm's expected future policies; allowing a larger price gap to remain in the current period will reduce expected discounted profits less if the firm expects to review its price in the following period regardless of the size of the price gap at that time than if the current price gap is likely to persist for many more periods. In the present section, both the prior and the value function are accordingly endogenized, in an explicit model of the optimal timing of discrete adjustments over an infinite horizon. The primitives of the model are instead the function $\pi(q)$ indicating the firm's profits in a single period as a function of that period's normalized price, and the probability distribution $g(\nu)$ of innovations in aggregate nominal expenditure, which indicate the amount by which each firm's normalized price will change if it does not review its price in that period.

The model is one with a countably infinite sequence of discrete dates (indexed by integers t) at which the firm's price may be adjusted (and at which sales occur).[22] I shall suppose that each firm seeks to maximize the expected value of a discounted objective function of the form

$$\sum_{t=0}^{\infty} \beta^t \pi(q_t), \tag{2.1}$$

where single-period profits are assumed to be given by a smooth, quasi-concave function $\pi(q)$ that reaches its unique maximum at an interior value that can be normalized as $q = 0$. Here the assumption that (real) profits depend on a firm's normalized price rather than upon its price and nominal expenditure separately follows from the usual irrelevance of absolute prices; but the assumption that profits depend only on the firm's *own* normalized price and not the normalized prices of other firms is instead a highly restrictive one.[23] However, the definition given here of a stationary optimal policy can be extended in a relatively straightforward way to the case in which profits also depend on aggregate state variables. The notation is simplified in this presentation by abstracting from such additional state variables, and it allows us to obtain a model in which the adjustment hazard is a function solely of a "price gap," as in the models of Caballero and Engel (1993a, 2007). An application to which the present simple framework applies is given in section 3.

Uncertainty about the firm's normalized price results from the random evolution of aggregate nominal expenditure $\mathcal{Y}_t \equiv P_t Y_t$. Again in order to reduce the size of the state space required to characterize equilibrium dynamics (and again following Caballero and Engel), I shall assume for the sake of simplicity that this evolves according to an exogenously given random walk,

$$\mathcal{Y}_t = \mathcal{Y}_{t-1} + \nu_t, \tag{2.2}$$

[22] The model could be extended in a reasonably straightforward way to the scheduling of reviews of pricing policy in continuous time, as in Reis (2006). But discrete time is mathematically simpler and allows more direct comparison with much of the prior literature on state-dependent pricing.

[23] As is explained further in section 3, even in cases where there are no strategic complementarities in the firms' optimal pricing decisions — in the sense that each firm's optimal price depends only on aggregate nominal expenditure, regardless of the prices of other firms — it is not generally true in a model of monopolistic competition that single-period real profits are independent of other firms' prices. Even when the optimal normalized price is independent of other firms' prices, the level of real profit associated with that optimum generally depends on the level of real aggregate demand, which is to say, on the *aggregate* normalized price.

where the innovation ν_t is drawn independently each period from a probability distribution with density function $g(\nu)$. The innovations ν_t are assumed to represent purely monetary disturbances; it is for this reason that they do not affect the profit function.

It remains to specify the nature of the plan chosen by a firm that reviews its pricing policy. In section 1, because the model is static, there is simply a decision whether to change to a different price in the single period; the existing policy is simply a price, and the new policy in the event of a review will be some other price. In a dynamic model, in which policy is not reviewed every period, it might seem reasonable to allow firms to choose pricing plans (to be followed until the next review) that are more complex than a constant price for the firm's product; for example, in the model of Reis (2006), a firm chooses a (potentially different) price for *each date* until the next review. Here I simplify the analysis by assuming that in addition to the cost of price reviews and the cost of interim information (both discussed in the previous section), there is a restriction on the *complexity* of the pricing policy that a firm can implement following a review. Some such restriction is clearly in the spirit of the assumption of costly information already made; for if there were no restriction on the informational complexity of the pricing policy that can be followed between "reviews" of the policy, the fact that it is costly to review policy more frequently would have no consequences. And, as is made clear below, I assume that a policy that depends on time (of the kind allowed by Reis) uses costly information, just as a policy that depends on the current state of aggregate demand does. Nonetheless, an analysis of the kind of pricing policy that would be chosen by a firm that must pay a higher cost in order to implement a policy that uses more information is left for future work; here, to simplify, I assume that only a pricing policy of *minimal* informational complexity — one that is *neither* time-dependent nor state-dependent — is feasible for the firm.

This means that a firm can change its price only when it pays the fixed cost κ and reviews its policy. Effectively, the model is one in which there is *both* a fixed cost of obtaining full information about current market conditions and a fixed cost of changing one's price, which must always be paid *together,* as in the model of Bonomo and Carvalho (2004).[24] In fact, in the limiting case in which $\theta = 0$, the fixed cost

[24] I depart from Bonomo and Carvalho, as noted earlier, by assuming that memory is costly — including memory of the length of time since the last price review, which plays a crucial role in their

κ corresponds precisely to the "menu cost" of changing prices in a standard SDP model. Nonetheless, κ also plays the role of the fixed cost of information collection assumed by Reis (2006); in the limiting case in which θ is made unboundedly large, the present model model can be viewed as a variant of Reis's model in which additional information restrictions are imposed, both on the complexity of the pricing policies that can be adopted and on the complexity of the policies with regard to the timing of reviews of pricing policy. I prefer to think of κ as essentially a cost of information collection and decisionmaking, while the costs of adopting a more complex pricing policy on the occasion of a price review are of a different sort (not a simple menu cost).[25] This affects the way in which it is reasonable to calibrate the size of the cost in a model intended to be empirically realistic.[26]

2.1 Costly Memory and the Invariant Prior

It follows from the assumptions just summarized that the normalized price of firm i evolves according to

$$q_{t+1}(i) = q_t(i) - \nu_{t+1}$$

if there is no review of the firm's policy in period t, while

$$q_{t+1}(i) = q_t^* - \nu_{t+1}$$

if firm i reviews its policy in period t. Here $q_t(i)$ is the normalized price of firm i in period t, *after* realization of the period t change in aggregate nominal expenditure, but *before* the decision about whether to review the firm's policy in period t, and q_t^* is the normalized price (*after* the review) that is chosen by a firm that reviews its policy in period t. The value of q_t^* is the same for all firms i because (as is shown below) the optimal choice for a firm that reviews its policy is independent of the normalized

derivation of a "time-dependent" rule for the timing of price changes — and by allowing partial information to be obtained between full price reviews by paying the variable cost θ.

[25] In fact, many retail prices do change between occasions on which pricing *policy* is reconsidered, as in the case of temporary "sales" following which the price returns again to an unchanged "regular" price. This may well reflect the execution of a pricing policy that involves little or no response to current market conditions. I believe that the form of such policies reflects a cost of complexity, but it is clearly one that is quite different from a "menu cost".

[26] Zbaracki *et al.* (2004) report that information and decision costs are several times as large as actual costs of announcing a new price, in the case of the firm that they study.

price that it has at the time of the review; hence if firms differ only in the periods in which they happen to have reviewed their prices in the past (despite having followed identical policies), q_t^* will be the same for all i. The aggregate shock ν_t is similarly the same for all i.

Given these dynamics for the normalized price of a firm, we can derive the dynamics of the firm's prior regarding its current normalized price. The dynamics of the prior depend, however, on what we assume about the firm's memory. In the theory of rational inattention proposed by Sims (1998, 2003, 2006), memory of the entire history of past signals is assumed to be perfectly precise (and costless); the information-flow constraint applies only to the degree of informativeness of *new* observations of external reality. Instead, I shall assume that access to one's own memory is *as costly* as access to any other source of information, during the intervals between full-scale price reviews. For example, one may allow firms to condition their price-review decision on the number of periods n that have elapsed since the last price review. In this case, the firm has a prior $f(q,n)$ over the joint distribution of its current normalized price q and the current value of n for that firm. The firm can learn the value of n and condition its decision on that value, but this would have an information cost of

$$-\theta \sum_n f_n \log f_n,$$

where $f_n \equiv \int f(q,n)dq$ is the marginal prior distribution over values of n. Assuming that the cost per nat θ of this kind of information is identical to the cost of information about the value of q, the firm will optimally choose not to learn the current value of n; since learning the value of n would be of use to the firm *only* because this information would allow it to estimate the current value of q with greater precision, it would always be more efficient to use any information capacity that it devotes to this problem to observe the current value of q with greater precision, rather than bothering to observe the value of n.

In assuming that the cost of information about the firm's memory of its own past signals is exactly the same as the cost of information about conditions external to the firm, I am making an assumption that is fully in the spirit of Sims' theory of rational inattention: rather than assuming that some kinds of information are easily observable while others are hidden, the cost of *any* kind of information is assumed to be the same as any other, because the relevant bottleneck is limited *attention* on the part of the decisionmaker, rather than anything about the structure of the world that obscures the values of certain state variables. This is admittedly a special case, but it is the assumption that makes Sims' theory such a parsimonious one. It is also a convenient case to analyze first, owing to its simplicity.[27]

In this case, any firm i begins any period t with a prior $f_t(q)$ over the possible values of $q_t(i)$. This prior indicates the *ex ante* distribution of possible values of the firm's normalized price in period t, given the policy followed in previous periods, but not conditioning on any of the signals received in previous periods, or on the timing of previous price reviews. By "the policy" followed in previous periods, I mean the design of the signalling mechanism, determining the probabilistic relation between the state and the signal received each period, and the firm's intended action in the event that any given signal is received, but not the history of the signals that were actually received or the actions that were taken. Following the analysis in section 1, the policy followed in period t can be summarized by a hazard function $\Lambda_t(q)$, indicating the probability of a price review in period t as a function of the normalized price in that period, and a reset value q_t^*, indicating the normalized price that the firm chooses if it reviews its pricing policy in period t. As a result of this policy, $q_{t+1}(i)$, the normalized price in period $t+1$ (after realization of the period $t+1$ innovative in aggregate nominal expenditure) will be equal to $q_t^* - \nu_t$ with probability $\Lambda(q_t(i))$ and equal to $q_t(i) - \nu_t$ with probability $1 - \Lambda(q_t(i))$, conditional on the value of $q_t(i)$. Integrating over the distribution of possible values of $q_t(i)$, one obtains a prior distribution for period $t+1$ equal to

$$f_{t+1}(q) = g(q_t^* - q) \int \Lambda_t(\tilde{q}) f_t(\tilde{q}) d\tilde{q} + \int g(\tilde{q} - q)(1 - \Lambda_t(\tilde{q})) f_t(\tilde{q}) d\tilde{q}. \tag{2.3}$$

This is the prior at the beginning of period $t+1$, regardless of the signal received in period t (*i.e.*, regardless of whether a price review occurs in period t), because the

[27] Interestingly, the literature on informational complexity constraints in game theory has more often made the *opposite* choice to that of Sims: it is considered more natural to limit the information content of a decisionmaker's memory than the information content of her perception of her current environment. For example, in Rubinstein (1986) and many subsequent papers, it is assumed that a strategy (in a repeated game) is preferred if it can be implemented by a finite-state automaton with a smaller number of states; this means, if it requires the decisionmaker to discriminate among a smaller number of different possible histories of previous play. But while memory is in this sense assumed to be costly, there is assumed to be no similar advantage of a strategy that reduces the number of different possible observations of current play among which the decisionmaker must discriminate.

firm has no costless memory.

The right-hand side of (2.3) defines a linear functional $T_{\Pi_t}[f_t]$ that maps any probability density f_t into another probability density f_{t+1}; the subscript indicates that the mapping depends on the policy $\Pi_t \equiv (\Lambda_t, q_t^*)$. Given an initial prior f_0 and policies $\{\Pi_t\}$ for each of the periods $t \geq 0$, the law of motion (2.3) implies a sequence of priors $\{f_t\}$ for all periods $t \geq 1$. Note that if for any policy Π, the prior f is such that

$$T_{\Pi}[f] = f, \tag{2.4}$$

it follows that if a firm starts with the prior $f_0 = f$ and implements policy Π each period, the dynamics (2.3) imply that the firm will have prior $f_t = f$ in every period. Thus f is an *invariant distribution* for the Markov process describing the dynamics of q under this policy. In such a situation, we can say that the firm's prior each period corresponds to the long-run frequency with which different values of its normalized price occur, under its constant policy Π. When the firm's prior is unchanging over time in this way, the constant prior makes it optimal for the firm to choose the same policy each period, which in turn makes it possible for the prior to remain constant. In the numerical analysis below, I shall be interested in computing statistics (for example, the frequency of price changes) for a stationary optimal plan of this kind.

The assumption that memory is (at least) as costly as information about current conditions external to the firm implies that under an optimal policy, the timing of price reviews is (stochastically) state-dependent, but *not* time-dependent, just as in full-information menu-cost models. When the cost θ of interim information is sufficiently large, the dependence of the optimal hazard on the current state is also attenuated, so that in the limit as θ becomes unboundedly large, the model approaches one with a constant hazard rate as assumed by Calvo (1983). If, instead, memory were costless, the optimal hazard under a stationary optimal plan would also depend on the number of periods since the last price review: there would be a different hazard function $\Lambda_n(q)$ for each value of n. In this case, in the limit of unboundedly large θ, each of the functions $\Lambda_n(q)$ would become a constant (there would cease to be dependence on q); but the constants would depend on n, and in the generic case, one would have Λ_n equal to zero for all n below some critical time, and Λ_n equal to 1 for all n above it. Thus the model would approach one in which prices would be reviewed at deterministic intervals, as in the analyses of Caballero (1989), Bonomo and Carvalho (2004), and Reis (2006). The analysis of this alternative case under the assumption of a finite positive cost of interim information is left for future work.

2.2 Stationary Optimal Plans

We can now state the firm's dynamic optimization problem. Its dynamic price-review scheduling strategy is a sequence of policies $\{\Pi_t\}$ for each of the periods $t \geq 0$; given the initial prior f_0, each such strategy implies a particular sequence of priors $\{f_t\}$ consistent with (2.3). The strategy is a *deterministic* sequence, insofar as in each period, the intended values of $\Lambda_t(q)$ and q_t^* depend only on t, and not on the signals received by the firm in any periods prior to t, on the timing of its price reviews prior to t, or on any information collected in the course of those reviews. This is because of the assumption that memory is costly; even if we imagine that the firm designs the signalling mechanism for period t and chooses its intended responses to signals in period t only when that period is reached, it must solve this design problem — which allows it to *choose* how much memory to access in period t in making its price-review decision — without making use of any memory.[28]

The firm's objective when choosing this strategy has three terms: the expected value of discounted profits (2.1), the expected discounted value of the costs of price reviews, and the discounted value of the costs of interim information used each period in that period's price-review decision. The cost of a price review is assumed to be κ in each period t in which such a review occurs; the cost of interim information is assumed to be θI_t in each period t (regardless of the signal received in that period), where I_t is defined as in (1.2) given the prior f_t for that period. In each case, the information costs are assumed to be in the same units as $\pi(q_t)$, and all costs in period t are discounted by the discount factor β^t.

A firm's *ex ante* expected profits in any period t can be written as $\bar{\pi}(\Pi_t; f_t)$, where $\Pi_t = (\Lambda_t(q), q_t^*)$ is the policy followed in period t, f_t is the firm's prior in period t

[28] I assume here that a firm can implement a sequence of policies $\{\Pi_t\}$ which need not specify the same policy Π for each period t, without using "memory" of the kind that is costly. I assume that a firm has no difficulty remembering the strategy that it chose *ex ante*; what is costly is memory of *things that happen* during the execution of the strategy, that were not certain to happen *ex ante*. Note also that the firm's price-review policy fails to be time-dependent, not because it lacks a "clock" to tell it the current value of t, but because it cannot costlessly remember whether it reviewed its pricing policy in any given previous period; it knows the value of t but not the value of n.

(given its policies in periods prior to t), and

$$\bar{\pi}(\Pi; f) \equiv \int [\Lambda(q)\pi(q^*) + (1 - \Lambda(q))\pi(q)] f(q) dq.$$

The *ex ante* expected cost of price reviews in period t can be written as $\kappa\bar{\lambda}(\Pi_t; f_t)$, where

$$\bar{\lambda}(\Pi; f) \equiv \int \Lambda(q) f(q) dq$$

indicates the probability of a price review under a policy Π. Finally, the cost of interim information can be written as $\theta I_t = \theta \bar{I}(\Pi_t; f_t)$, where

$$\bar{I}(\Pi; f) \equiv \int \varphi(\Lambda(q)) f(q) dq - \varphi(\bar{\lambda}(\Pi; f)), \tag{2.5}$$

again defining $\varphi(\Lambda)$ as in (1.6).

The firm's problem is then to choose a sequence of policies $\{\Pi_t\}$ for $t \geq 0$ to maximize

$$\sum_{t=0}^{\infty} \beta^t [\bar{\pi}(\Pi_t; f_t) - \kappa\bar{\lambda}(\Pi_t; f_t) - \theta \bar{I}(\Pi_t; f_t)], \tag{2.6}$$

where the prior evolves according to

$$f_{t+1} = T_{\Pi_t}[f_t] \tag{2.7}$$

for each $t \geq 0$, starting from a given initial prior f_0. A *stationary optimal policy* is a pair (f, Π) such that if $f_0 = f$, the solution to the above dynamic problem is $\Pi_t = \Pi$ for all $t \geq 0$, and the implied dynamics of the prior are $f_t = f$ for all $t \geq 0$. Note that this definition implies that f satisfies the fixed-point relation (2.4), so that f is an invariant distribution under the stationary price-review policy Π.

2.3 A Recursive Formulation

The optimization problem stated above can be given a recursive formulation. This is useful for computational purposes, and also allows us to see how the problem involves a sequence of single-period price-review decisions of the kind treated in section 1. As a result, the characterization given there is both useful in computing the stationary optimal policy, and helpful in characterizing the random timing of price reviews of under such a policy.

For any initial prior f_0, let $J(f_0)$ denote the maximum attainable value of the objective (2.6) in the problem stated above. Then standard arguments imply that $J(f)$ must satisfy a *Bellman equation* of the form

$$J(f_t) = \max_{\Pi_t} \left\{ \bar{\pi}(\Pi_t; f_t) - \kappa\bar{\lambda}(\Pi_t; f_t) - \theta\bar{I}(\Pi_t; f_t) + \beta J(f_{t+1}) \right\}, \tag{2.8}$$

where f_{t+1} is given by (2.7). If we can find a functional $J(f)$ (defined on the space of probability measures f) that is a fixed point of the mapping defined in (2.8), then this is a *value function* for the optimization problem stated above. Moreover, the dynamic price-review scheduling problem can then be reduced to a sequence of single-period problems: in each period t, the policy Π_t is chosen to maximize the right-hand side of (2.8) subject to the constraint (2.7), given the prior f_t in the current period. The policy chosen each period then determines the prior in the next period through the law of motion (2.7). A stationary optimal policy is then a pair (f, Π) such that (i) if $f_t = f$, the solution to the problem (2.8) is $\Pi_t = \Pi$; and (ii) the distribution f is a fixed point (2.4) of the mapping defined by the policy Π.

This still does not make it easy to compute a stationary optimal policy, as one must first compute a functional $J(f)$ that is a fixed point of (2.8), and this is far from trivial, since (2.8) defines a mapping from a very high-dimensional function space into itself. Nor is the single-period policy problem defined in (2.8) as simple as the one considered in section 1. However, we can obtain an even simpler characterization by observing that $J(f_t)$ is necessarily a *concave* functional, that is furthermore differentiable at $f_t = f$ (the invariant distribution under the stationary optimal policy), so that for distributions f_t close enough to f, the value function can be approximated by a *linear* functional

$$J(f_t) \approx J(f) + \int j(q)[f_t(q) - f(q)]dq,$$

where $j(q)$ is an integrable function. (Note that the derivative function $j(q)$ is defined only up to an arbitrary constant, since $J(f_t)$ is not defined for perturbations of the set-valued function f_t that do not integrate to 1.) The concavity of $J(f_{t+1})$ then implies that $\Pi_t = \Pi$ solves the problem (2.8) when $f_t = f$ if and only if it solves the alternative problem

$$\max_{\Pi_t} \left\{ \bar{\pi}(\Pi_t; f) - \kappa\bar{\lambda}(\Pi_t; f) - \theta\bar{I}(\Pi_t; f) + \beta \int j(q)[f_{t+1}(q) - f(q)]dq \right\}, \tag{2.9}$$

where f_{t+1} is again given by (2.7).

Using (2.7) to substitute for f_{t+1}, the objective in (2.9) can alternatively be expressed as

$$(V(q_t^*) - \kappa) \int \Lambda_t(q) f(q) dq + \int V(q)(1 - \Lambda_t(q)) f(q) dq - \theta \bar{I}(\Lambda_t; f), \tag{2.10}$$

where

$$V(q) \equiv \pi(q) + \beta \int j(\tilde{q}) g(q - \tilde{q}) d\tilde{q}, \tag{2.11}$$

and I have now written simply $\bar{I}(\Lambda_t; f)$, to indicate that the function $\bar{I}$ defined in (2.5) does not depend on the choice of q^*. (Here the variable of integration q in (2.10) is the normalized price in period t *after* the period t disturbance to aggregate expenditure, but *before* the decision whether to conduct a price review, as in section 1. In (2.11), q is instead the normalized price that is charged, *after* any price review has occurred, while $\tilde{q}$ is the normalized price in the following period, after that period's disturbance to aggregate expenditure, but before the decision whether to conduct a price review in that period.) Maximization of (2.10) is in turn equivalent to maximizing

$$\int L(q; q_t^*) \Lambda_t(q) dq - \theta \bar{I}(\Lambda_t; f), \tag{2.12}$$

if we define

$$L(q; q^*) \equiv V(q^*) - V(q) - \kappa, \tag{2.13}$$

as in section 1. Hence $\Pi_t = \Pi$ solves the problem (2.8) when $f_t = f$ if and only if it maximizes (2.12).

This, in turn, is easily seen to be true if and only if (i) q^* is the value of q that maximizes $V(q)$, and (ii) given the value of q^*, the hazard function Λ maximizes (2.12), which is identical to the objective (1.3) or (1.7) considered in section 1. Thus in a stationary optimal plan, each period a policy Π is chosen that solves a single-period problem identical to the one considered in section 1. However, the definition of this problem involves the function $j(q)$; thus it may still seem necessary to solve the Bellman equation for the function $J(f)$.

In fact, though, we only need to know the derivative function $j(q)$. And an envelope-theorem calculation, differentiating (2.8) at $f_t = f$, yields

$$\begin{aligned}
j(q) &= \Lambda_t(q)\pi(q_t^*) + (1-\Lambda_t(q))\pi(q) - \theta\ \left[\varphi(\Lambda_t(q)) - \varphi'\left(\int \Lambda_t(\tilde{q})f(\tilde{q})d\tilde{q}\right)\right] \\
&\quad -\kappa\Lambda_t(q) + \beta\int j(\tilde{q})[\Lambda_t(q)g(q_t^* - \tilde{q}) + (1-\Lambda_t(q))g(q-\tilde{q})]d\tilde{q} \\
&= \Lambda_t(q)[V(q_t^*) - \kappa] + (1-\Lambda_t(q))V(q) - \theta\ \left[\varphi(\Lambda_t(q)) - \varphi'\left(\int \Lambda_t(\tilde{q})f(\tilde{q})d\tilde{q}\right)\right] \\
&= V(q) + \Lambda_t(q)L(q;q_t^*) - \theta\ \left[\varphi(\Lambda_t(q)) - \varphi'\left(\int \Lambda_t(\tilde{q})f(\tilde{q})d\tilde{q}\right)\right] \\
&= V(q) - \theta[\varphi(\Lambda_t(q)) - \varphi'(\Lambda_t(q))\Lambda_t(q)] \\
&= V(q) - \theta\log(1-\Lambda_t(q)).
\end{aligned}$$

Here the second line uses the definition (2.11) of $V(q)$; the third line uses the definition (2.13) of $L(q;q^*)$; the fourth line uses the fact that, as shown in section 1, a solution to the problem (2.9) — and accordingly, a solution to the problem (2.8) — must satisfy the first-order condition (1.8) to substitute for $L(q;q^*)$; and the final line uses the definition (1.6) of the binary entropy function $\varphi(\Lambda)$. Note also that on each line, I have suppressed an arbitrary constant term, since $j(q)$ is defined only up to a constant.

Substituting the above expression for $j(q)$ into the right-hand side of (2.11), we obtain

$$V(q) \equiv \pi(q) + \beta\int [V(\tilde{q}) - \theta\log(1-\Lambda(\tilde{q}))]\ g(q-\tilde{q})d\tilde{q}, \tag{2.14}$$

a fixed-point equation for the function $V(q)$ that makes no further reference to either the value function J or its derivative. A stationary optimal policy then corresponds to a triple (f, Π, V) such that (i) given the policy Π, the function V is a fixed point of the relation (2.14); given the pseudo-value function V and the prior f, the policy Π solves the maximization problem treated in section 1; and (iii) given the policy Π, the distribution f is an invariant distribution, *i.e.,* a fixed point of relation (2.4).

This characterization of a stationary optimal policy reduces our problem to a much more mathematically tractable one than solution of (2.8) for the value function $J(f)$. We need only solve for two real-valued functions of a single real variable, the functions $V(q)$ and $\Lambda(q)$; a probability distribution $f(q)$ over values of that same single real variable; and a real number q^*. These can be solved for using standard methods of function approximation and simulation of invariant distributions, of the kind discussed for example in Miranda and Fackler (2002).[29]

3 A Model of Monopolistically Competitive Price Adjustment

Let us now numerically explore the consequences of the model of price adjustment developed in section 2, in the context of an explicit model of the losses from infrequent price adjustment of a kind that is commonly assumed, both in the literature on canonical (full-information) SDP models and in ET models of inflation dynamics. The economy consists of a continuum of monopolistically competitive producers of differentiated goods, indexed by i. In the Dixit-Stiglitz model of monopolistic competition,[30] each firm i faces a demand curve of the form

$$y_t(i) = Y_t \left(\frac{p_t(i)}{P_t}\right)^{-\epsilon}$$

for its good, where $p_t(i)$ is the price of good i,

$$Y_t \equiv \left[\int y_t(i)^{\frac{\epsilon-1}{\epsilon}} di\right]^{\frac{\epsilon}{\epsilon-1}} \tag{3.1}$$

is the Dixit-Stiglitz index of aggregate output (or real aggregate demand), $\epsilon > 1$ is the constant elasticity of substitution among differentiated goods, and

$$P_t \equiv \left[\int p_t(i)^{1-\epsilon} di\right]^{\frac{1}{1-\epsilon}} \tag{3.2}$$

is the corresponding aggregate price index. An individual firm takes as given the stochastic evolution of the aggregate market conditions $\{P_t, Y_t\}$ in considering the effects of alternative paths for its own price.

Abstracting from the costs of information and the fixed costs associated with price reviews, a firm's objective is to maximize the present value of profits

$$\mathrm{E} \sum_{t=0}^{\infty} R_{0,t} \Pi_t(i), \tag{3.3}$$

[29] The specific approach used to compute the numerical results described in the next section is discussed further in the Appendix.

[30] See, for example, Woodford (2003, chap. 3) for details of this model and of the derivation of the profit function.

where $\Pi_t(i)$ denotes the real profits (in units of the composite good defined in (3.1)) of firm i in period t, and $R_{0,t}$ is a stochastic discount factor, discounting real income in any state of the world in period t back to its equivalent value in terms of real income in period zero. The expectation operator E indicates an unconditional expectation, *i.e.*, an expectation under the firm's prior about possible evolutions of the economy from period zero onward, before receiving any information in period zero about the economy's state at that time. In the case of a representative-household model, in which we furthermore assume that aggregate output Y_t is also the equilibrium consumption of the composite good by the representative household, and assume time-separable isoelastic (or CRRA) preferences with a constant intertemporal elasticity of substitution $\sigma > 0$, then the stochastic discount factor is given by

$$R_{0,t} = \beta^t \left(\frac{Y_0}{Y_t}\right)^{\sigma^{-1}},$$

where $0 < \beta < 1$ is the representative household's utility discount factor. We can then express the objective (3.3) as (a positive multiple of) an objective with an exponential (non-state-contingent) discount factor,

$$\mathrm{E}\sum_{t=0}^{\infty} \beta^t \tilde{\Pi}_t(i), \tag{3.4}$$

as assumed above in (2.1), if we define

$$\tilde{\Pi}_t(i) \equiv Y_t^{-\sigma^{-1}} \Pi_t(i).$$

Under the Dixit-Stiglitz model of monopolistic competition, the real revenues of firm i in period t are equal to

$$Y_t P_t^{\epsilon-1} p_t(i)^{1-\epsilon}.$$

If we also assume an isoelastic disutility of work effort, sector-specific labor markets, an isoelastic (or Cobb-Douglas) production function, and efficient labor contracting (or a competitive spot market for labor in each sector), then the complete expression for marginal-utility-weighted profits is of the form

$$\tilde{\Pi}_t(i) = Y_t^{1-\sigma^{-1}} P_t^{\epsilon-1} p_t(i)^{1-\epsilon} - \frac{\lambda}{1+\omega}[Y_t P_t^{\epsilon} p_t(i)^{-\epsilon}]^{1+\omega}, \tag{3.5}$$

where $\omega \geq 0$ measures the combined curvatures of the production function and the disutility-of-labor function,[31] and $\lambda > 0$ is a positive constant. If we define $\bar{Y} > 0$ as the full-information/flexible-price equilibrium level of output — which is to say, the level of output such that if $Y_t = \bar{Y}$, profits are maximized by a price $p_t(i) = P_t$ — then we can write

$$\tilde{\Pi}_t(i) = \bar{Y}^{1-\sigma^{-1}} \tilde{\pi}(Q_t(i), \tilde{Y}_t),$$

where both the normalized price

$$Q_t(i) \equiv \frac{p_t(i)\bar{Y}}{P_t Y_t}$$

and the output gap

$$\tilde{Y}_t \equiv \frac{Y_t}{\bar{Y}}$$

are scale-independent quantities that differ from 1 only to the extent that prices fail to perfectly adjust to current market conditions, and the normalized profit function is given by

$$\tilde{\pi}(Q, \tilde{Y}) \equiv \tilde{Y}^{2-\epsilon-\sigma^{-1}} Q^{1-\epsilon} \left[1 - \frac{1}{\mu(1+\omega)} \tilde{Y}^{(\sigma^{-1}+\omega)-(1+\omega\epsilon)} Q^{-(1+\omega\epsilon)}\right], \qquad (3.6)$$

using $\mu \equiv \epsilon/(\epsilon - 1) > 1$ for the desired markup of price over marginal cost.

In the present paper, I shall restrict attention to a "partial equilibrium" analysis of price adjustment by a firm, or by a group of firms that comprise only a negligible fraction of the entire economy, with information costs and menu costs of the kind discussed in section 2, but in an economy in which measure 1 of all firms are assumed to immediately adjust their prices fully in response to any variations in aggregate nominal expenditure, so that for most firms $Q_t(i) = 1$ at all times. (Note that this last assumption corresponds to optimal behavior under full information and no cost of changing prices.) As a consequence, it follows that the aggregate price index (3.2) will equal $\mathcal{Y}_t/\bar{Y}$ in each period, where $\mathcal{Y}_t$ denotes aggregate nominal expenditure, and

[31] The notation follows Woodford (2003, chap. 3), where the model is further explained. The allowance for sector-specific labor markets increases the degree of strategic complementarity between the pricing decisions of firms in different sectors, which allows larger real effects of nominal rigidities for reasons discussed, for example, in Burstein and Hellwig (2007). If one instead assumes that all firms hire the same homogeneous labor input in a single competitive spot market, (3.5) still applies, but in this case ω reflects only the curvature of the production function (*i.e.*, the diminishing marginal productivity of labor), so that it is harder to justify assigning ω a value that is too large.

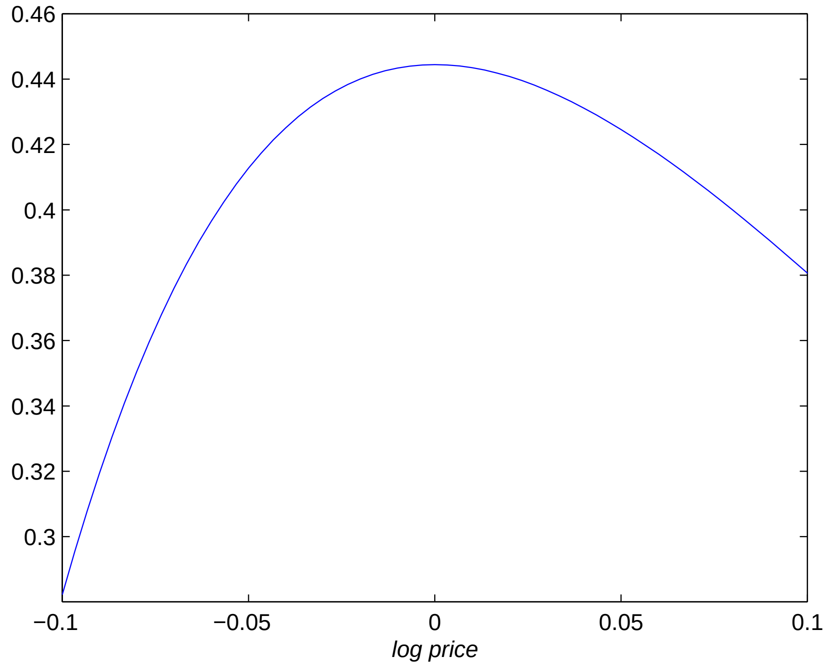

Figure 2: The normalized profit function $\pi(q)$.

hence that $\tilde{Y}_t = 1$ in each period. In fact, for purposes of our characterization of the pricing decisions of an individual firm, all that matters is the assumption that the price index P_t perfectly tracks aggregate expenditure, so that $\tilde{Y}_t = 1$ at all times; it does not matter if this is true (as in the model of Caplin and Spulber, 1987) without individual firms each adjusting their prices to track aggregate expenditure.

The advantage of this simplification is that normalized profits each period depend only on the individual firm's normalized price in that period, allowing us to work with a unidimensional state space (under the further simplifying assumption of a random walk in aggregate expenditure), as in the "generalized Ss" framework of Caballero and Engel (1993a, 2007). This is a tremendous computational simplification. Considering this special case amounts to an investigation of the consequences of costly information and costs of reviewing prices for price adjustment while abstracting from the effects of slow adjustment of prices *elsewhere* in the economy on the adjustment of an individual firm's or sector's prices; essentially, we consider the non-neutrality of purely nominal disturbances while abstracting from strategic complementarities

Table 1: Resource expenditure on information, for alternative values of θ. Each share is measured in percentage points.

θ	s_κ	s_θ	r_θ
0	.006	0	100
.0004	.006	.0023	14.3
.004	.008	.0030	1.3
.04	.012	.0010	.04
.4	.014	.0000	.0000
4	.014	.0000	.0000
40	.014	.0000	.0000
∞	.014	0	0

in the pricing decisions of firms in different sectors. Since it is plausible to assume that such complementarities are important (Woodford, 2003, chap. 3), we should expect this analysis to underestimate the quantitative magnitude of the real effects of nominal disturbances. Nonetheless, analysis of this simple case is useful, because we can examine the solution in greater detail. And consideration of this case can already allow us to answer one question about the general equilibrium version of the model, namely, whether a neutrality result of the kind obtained by Caplin and Spulber (1987) holds. For if such a result *did* obtain in the general-equilibrium model, then the assumption made here would be correct (the aggregate price index would perfectly track aggregate expenditure), and the behavior of each firm in the general-equilibrium model would be identical to our partial-equilibrium analysis. Thus the fact that we find substantial non-neutrality even in a partial-equilibrium analysis already implies that in a general-equilibrium analysis, nominal disturbances will not be neutral. (I defer a full analysis of the general-equilibrium case to future work.)

In the case that $\tilde{Y}_t = 1$ at all times, the normalized profit function can be written as

$$\pi(q) \equiv \tilde{\pi}(e^q, 1) = e^{-(\epsilon-1)q} - \frac{\epsilon - 1}{\epsilon(1+\omega)} e^{-(1+\omega)\epsilon q} \tag{3.7}$$

as in (2.1) above, where $q \equiv \log Q$.[32] The function is determined by two parameters,

[32] This definition coincides with the one in sections 1-2 if we adopt units for measuring output in

ω and ϵ (or ω and μ). For any values of these parameters, the profit function is increasing for $q < 0$, maximized at $q = 0$, and decreasing for $q > 0$. Moreover, it is asymmetric, in that $\pi(q) > \pi(-q)$ for any $q > 0$. (This asymmetry gives rise to an asymmetry in the optimal hazard function, discussed below.) Figure 2 plots this function for the illustrative parameter values $\epsilon = 6$, $\omega = 0.5$.[33]

Now let us suppose furthermore (again in order to simplify the analysis) that aggregate nominal expenditure $\{\mathcal{Y}_t\}$ evolves according a random walk (2.2), where the innovation ν_t is drawn independently each period from a distribution $N(\bar{\pi}, \sigma_z^2)$. The shocks indicated by the innovations $\{\nu_t\}$ are understood to be purely monetary in character; they result from random variations in monetary policy, not associated with any changes in preferences or technology. Under this specification (and the stipulation that P_t perfectly tracks $\mathcal{Y}_t$), the current value of $\mathcal{Y}_t$ completely summarizes everything about the aggregate state of the economy at date t that is relevant to the pricing problem of a firm (*i.e.*, all information that is available in principle about the current values or future evolution of both real aggregate demand and the aggregate price index). Hence a firm (which is assumed to know its own current price) has no need of any information about current or past market conditions other than the current value of $\mathcal{Y}_t$, or equivalently, the current value of its normalized price $q_t(i)$, as in the kind of dynamic problem discussed in section 2. In fact, the firm's decision problem is of exactly the kind discussed there, where the period profit function $\pi(q)$ is given by (3.7), and the shock distribution $g(\nu)$ is $N(\bar{\pi}, \sigma_z^2)$.

3.1 The Stationary Optimal Policy

Given a specification of the profit function $\pi(q)$, the discount factor $0 < \beta < 1$, the shock distribution $g(\nu)$, and the cost parameters $\kappa, \theta > 0$, one can solve numerically for the value function $V(q)$, the target normalized price q^*, the optimal hazard function $\Lambda(q)$, and the invariant distribution $f(q)$ that constitute a *stationary optimal*

which $\bar{Y} = 1$. The advantage of the normalization proposed here is that q can now be interpreted as the gap between the log price and the one that would be chosen under full information and no costs of price changes; thus the absolute magnitude of q is meaningful, and not just the gap $q - q^*$.

[33]This value of ϵ implies a degree of market power such that the steady-state markup of prices over marginal cost is 20 percent. The value of ω corresponds to the degree of curvature of the disutility of output supply that would be implied by a Cobb-Douglas production function with a labor coefficient of 2/3 and a linear disutility of work effort.

Table 2: The optimal value of q^* for alternative values of θ.

θ	q^*
0	.00003
.0004	.00003
.004	.00005
.04	.00008
.4	.00009
4	.00009
40	.00009
∞	.00009

policy for a firm in this environment, using an algorithm of the kind discussed in Appendix B. Here I present illustrative numerical results for the profit function shown in Figure 2, together with parameter values $\beta = 0.9975$ (corresponding to a 3 percent annual rate of time preference, on the understanding that model "periods" represent months), $\kappa = .002$ (the cost of a price review is 0.2 percent of monthly steady-state revenue), and a range of possible values for the information cost θ. I assume zero drift in aggregate nominal expenditure, or alternatively in the general level of prices, so that $\bar{\pi} = 0$, and an innovation standard deviation $\sigma_z = .001$.[34]

Table 1 lists the alternative values of θ that are considered,[35] and in each case indicates the implied cost to the firm of inter-review information collection (*i.e.,* the cost of the information on the basis of which decisions are made about the scheduling of price reviews), as well as the cost to the firm of price reviews themselves, both as average shares of the firm's revenue. (These two shares are denoted s_θ and s_κ respectively.) The table also indicates how the assumed information used by the firm in deciding when to review its prices compares to the amount of information that would be required in order to schedule price reviews optimally; the information used is fraction r_θ of the information that would be required for a fully optimal decision,

[34] This corresponds to a standard deviation for quarterly innovations in the (annualized) inflation rate of approximately 70 basis points.

[35] The bottom line of the table describes limiting properties of the stationary optimal plan, as the value of θ is made unboundedly large, *i.e.,* in the "Calvo limit".

given the firm's value function for its continuation problem in each period (which depends on the fact that, at least in the future, it does *not* expect to schedule price reviews on the basis of full information). A value of $\theta = 0.04$, for example, might seem high, in that it means that the cost per nat of information is 4 percent of the firm's monthly steady-state revenues. (Alternatively, the cost per bit of information is 2.8 percent of monthly revenue.) However, under the stationary optimal policy, the firm only uses information each month in deciding whether to review its pricing policy with a cost equivalent to .001 percent of steady-state revenue. And since this is .04 percent of the information that would be required to make a fully optimal decision, this specification of the information cost implies that it would only cost about 2.5 percent of monthly revenue for the firm to make a fully optimal decision.[36] Considered in this way, an information cost of $\theta = 0.04$ does not seem especially high.[37] It seems high when expressed as a cost per bit (or cost per nat), because I allow the signal s to be designed so as to focus on precisely the information needed for the manager's decision; once I have done so, one can only explain imprecision in the decisions that are taken under the hypothesis that the information content of s must be quite small, or alternatively, that the marginal cost of increasing the information content of the signal s is quite high.[38]

The optimal policy of an individual firm is specified by the reset value for the normalized price, q^*, and the hazard function $\Lambda(q)$. (An advantage of the univariate case considered here is that the hazard is a function of a single real variable, and can easily be plotted.) Table 2 shows the optimal value of q^* for a range of values for the information cost θ, and Figure 3 plots the corresponding optimal hazard functions.

[36] Here I refer to the cost of making a fully optimal decision in one month only, taking for granted that one's problem in subsequent months will be the information-constrained problem characterized here, and not to the cost of making a fully optimal decision each month, forever. In Table 1, the information cost of a fully optimal decision is computed using the value function $V(q)$ associated with the stationary optimal policy corresponding to the given value of θ.

[37] The value $\theta = 4$ would instead imply that the information required for a fully optimal price-review decision each month would cost about twice steady-state revenue, a prohibitive expense; and for $\theta = 40$, the cost would be about 20 times steady-state revenue.

[38] It is important to understand that the parameter θ does not represent a cost-per-letter of having a staff member read the *Wall Street Journal*; it is instead intended to represent a cost of getting the attention of the manager who must make the decision, once the staff have digested whatever large amount of information may have been involved in the preparation of the signal s that must be passed on to the manager.

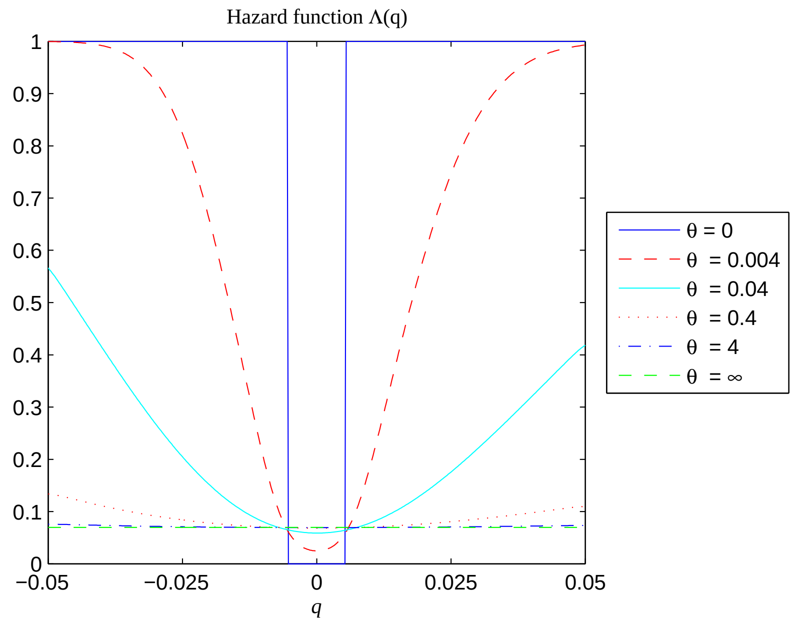

Figure 3: The optimal hazard function $\Lambda(q)$, for alternative values of θ.

One observes that q^* is positive, though quite small.[39] The optimal reset value is slightly positive because of the asymmetry of the profit function seen in Figure 2. Because the losses associated with a price that is too low are greater than those associated with a price that is too high by the same number of percentage points, it is prudent to set one's price slightly higher than one would if one expected to be able to adjust the price again in the event of any change in market conditions, in order to reduce the probability of having a price that is too low. The size of the bias that is optimal is slightly higher the more costly is interim information; but in no case is it very large.

In the case that $\theta = 0$, the optimal hazard function has the "square well" shape associated with standard SDP models: there is probability 0 of adjusting inside the Ss thresholds, and probability 1 of adjusting outside them. For positive values of θ, one instead has a continuous function taking values between 0 and 1, with the lowest values in the case of price gaps near zero, and the highest values for large price gaps of either sign. When θ is small (though positive), as in the case $\theta = 0.004$ shown in

[39] A value of .00008, for example, means that when a firm reviews its price, it sets the price 0.008 percent higher than it would in a full-information/flexible-price economy.

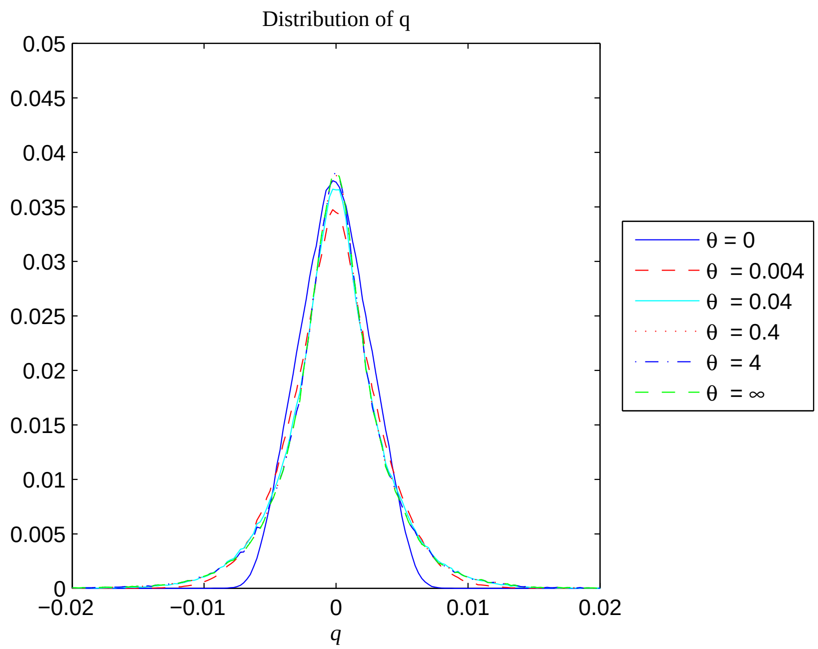

Figure 4: The invariant distribution $f(q)$, for alternative values of θ.

the figure, the hazard function is still barely above 0 for small price gaps, and rises rapidly to values near 1 for price gaps that are only a small distance outside the "zone of inaction" under full information. But for larger values of θ, the optimal hazard function is significantly positive even for price gaps near zero, and increases only slightly for price gaps far outside the full-information "zone of inaction". In the case that $\theta = 4$, the optimal hazard function is essentially constant over the entire interval for which the function is plotted in Figure 3, though also in this case, the hazard rate eventually rises to values near 1 for a large enough negative price gap. One also observes that for any given normalized price q, there is a positive limiting value for $\Lambda(q)$ that is approached in the case of any large enough value of θ. For example, one can see in the figure that for values in the interval $-0.01 \leq q \leq 0.01$, the optimal hazard rate is essentially the same positive value for all values of θ equal to 0.04 or higher. This limiting value of the optimal hazard rate is the same positive value for all values of q, though the convergence can only be observed in the figure for values of q in an interval around zero; thus one obtains the Calvo model (in which $\Lambda(q) = \bar{\Lambda}$ for all q) as a limiting case of the present model, in which θ is made unboundedly large.

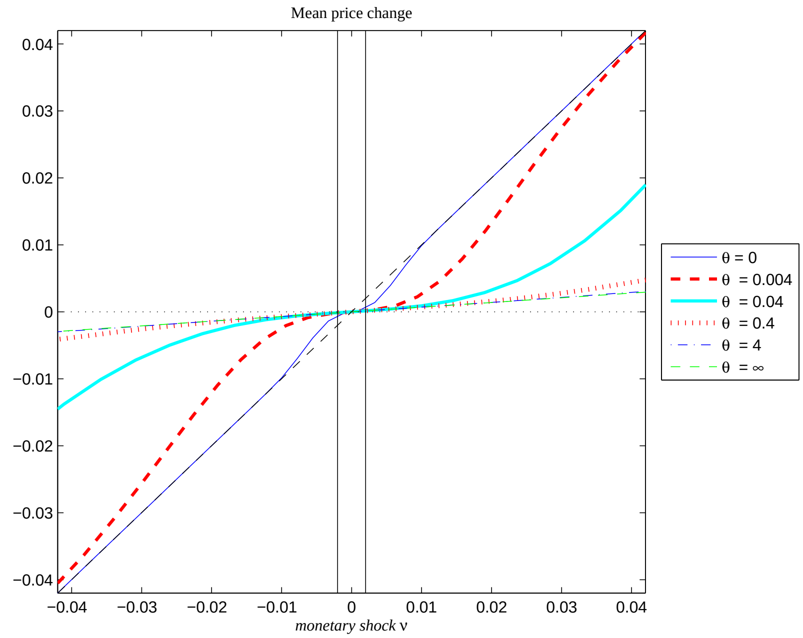

Figure 5: The function $h(\nu)$, for alternative values of θ. The dashed line on the diagonal shows the benchmark of perfect neutrality.

The invariant distribution $f(q)$ implied by the optimal policy $(\Lambda(q), q^*)$ is shown in Figure 4 for each of these same values of θ. As the value of θ is increased, the range of variation in the normalized log price in equilibrium falls slightly. This is because the hazard rate becomes larger near q^*, so that a firm's normalized log price is less certain to wander away very far from q^* before it is reconsidered (and returned to the value q^*); hence the long-run distribution of normalized log prices is more concentrated in a neighborhood of q^*. The invariant distribution converges to a well-defined limiting distribution (the one associated with the limiting Calvo policy) as θ is made large; in fact, it is evident from the figure that the invariant distribution has nearly converged once θ is equal to 0.04 or larger. This is not surprising, given that for values of θ of this magnitude, the optimal hazard function has nearly converged, for the range of values of q that occur with appreciable probability in the limiting invariant distribution.

Figures 3 and 4 together imply that the Calvo model should provide a reasonable approximation to the dynamics of price adjustment as long as θ is on the order of

0.04 or larger. While the optimal hazard function has not yet converged, when θ is no larger than 0.04, for *large* values of q, it is reasonably constant (and close to its limiting value) on the interval $-0.01 \leq q \leq 0.01$; and in the invariant distribution implied by this policy, q will remain within that interval *most of the time.* Hence the Calvo model should be a good approximation, not only in the case of very large values of θ, but even in the case of more moderate values — θ need only be large enough to make the optimal hazard function relatively constant over the range of values of q that occur with high frequency under the equilibrium dynamics. This only requires that θ be large relative to κ.[40]

Figure 3 also indicates that in the case of large enough shocks for the constant-hazard approximation to be a poor one, the hazard function is not perfectly symmetric. It is particularly clear in the case that $\theta = 0.04$ that the hazard rate rises more steeply in the case of *negative* price gaps than in the case of *positive* price gaps of the same size. (This in turn is due to the asymmetry of the profit function shown in Figure 2.) In fact, this asymmetry is of the same sign as has been found to best fit U.S. data on both aggregate inflation dynamics (Caballero and Engel, 1993a) and on the distribution of individual price changes (Caballero and Engel, 2006). The present model provides an economic explanation for asymmetry of that kind.[41]

3.2 Monetary Non-Neutrality

A key question is to what extent an increase in aggregate nominal expenditure results in an immediate increase in the general level of prices, or alternatively, in an increase in aggregate real activity. Here we have assumed that most prices adjust immediately in full proportion to the increase in nominal expenditure, so that there cannot be any affect on aggregate real activity; but we can still ask what the effect is on average prices among those firms (assumed to represent a negligible share of the economy

[40]Note that since $L(0) = -\kappa$, it follows from (1.9) that if $e^{-\kappa/\theta}$ is a small fraction, then $\bar{\Lambda}$ must exceed $\Lambda(0)$ by a correspondingly small fraction. But this means that $\Lambda(x)$ cannot rise very much above its minimum value $\Lambda(0)$ over the range of x values that occur with any substantial probability.

[41]I do not attempt here to ask to what extent the model can reproduce the *degree* of asymmetry that would best fit the data when a theoretically unconstrained hazard function is estimated. The "partial equilibrium" case considered here is hardly realistic enough for such an exercise to be of interest. But the connection between the asymmetry of the profit function and the *sign* of the asymmetry in the optimal hazard function seems likely to carry over to more complex environments.

as a whole) that are subject to the information costs and costs of reviewing their prices. If we have aggregate neutrality even when individual prices do not all adjust in response to a shock, as in the model of Caplin and Spulber (1987), then we should find that the *average* price of sticky-price firms adjusts exactly in proportion to the increase in aggregate demand, even though the individual prices of such firms do not.

A quantity of interest is therefore

$$h(\nu) \equiv \mathrm{E}[\Delta p_t(i)|\nu_t = \nu],$$

the average price increase (among the sticky-price firms) resulting from an innovation of size ν in aggregate nominal expenditure. Here the expectation is conditional upon the value of the most recent shock, but integrating over all possible histories of disturbances prior to the current period. Note that the average price change resulting from a given shock ν_t depends on the distribution of price gaps that happens to exist at the time that the shock occurs, as has frequently been stressed by Caballero and Engel; the average price change is thus a nonlinear function of both the current shock and the previous history of shocks. But by integrating over the possible previous histories we obtain an *average* answer to the question of how much prices change in response to a given size of shock; this provides a useful measure of monetary non-neutrality that can be easily plotted, as it is a function of a single real variable.

It follows from our characterization of a stationary optimal policy that

$$h(\nu) = -\int [\tilde{q} - q^* - \nu]\Lambda(\tilde{q} - \nu)\tilde{f}(\tilde{q})d\tilde{q},$$

where $\tilde{f}(\tilde{q})$ is the invariant distribution of values for $\tilde{q}_t(i)$, firm i's normalized log price *before* the period t innovation in aggregate nominal expenditure. (After a shock ν_t, the normalized log price is $q_t(i) = \tilde{q}_t(i) - \nu_t$, and if the price is then reviewed, the resulting price change will be of size $-[\tilde{q}_t(i) - \nu_t - q^*]$.) The invariant distribution $\tilde{f}(\tilde{q})$ will furthermore consist of a continuous density $(1 - \Lambda(\tilde{q}))f(\tilde{q})$ plus an atom of size $\bar{\Lambda}$ at $\tilde{q} = q^*$.[42] Hence we can alternatively write

$$h(\nu) = \bar{\Lambda}\Lambda(q^* - \nu) - \int [q - q^* - \nu]\Lambda(q - \nu)(1 - \Lambda(q))f(q)dq,$$

[42]Since the distribution contains an atom, and is not a continuous density, writing it as a function $\tilde{f}(\tilde{q})$ involves an abuse of notation. But this way of writing an integral with respect to the probability measure $\tilde{f}$ is used by analogy with the way I have previously written integrals with respect to the measure f, and should create little confusion.

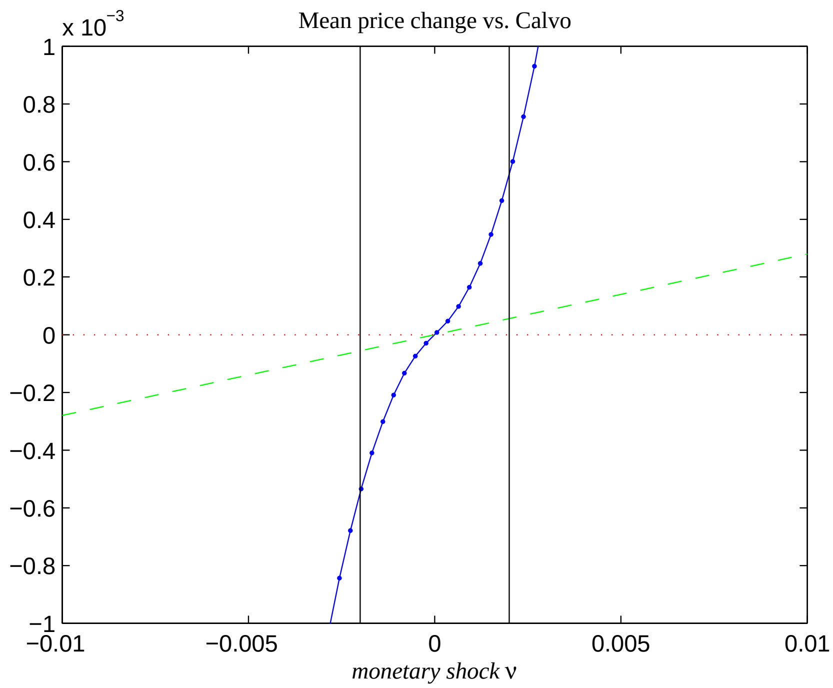

Figure 6: A closer view of the function $h(\nu)$, for the case $\theta = 0$. The dashed line shows the prediction of the Calvo model for purposes of comparison.

in terms of the quantities q^* and the functions $\Lambda(q)$ and $f(q)$ that are computed in the solution for the stationary optimal policy. (See Appendix B for further discussion of the computation of $h(\nu)$.)

There are two simple benchmarks with which it is useful to compare the function $h(\nu)$ obtained for the model with information-constrained price review decisions. One is the benchmark of *perfect neutrality*. In this case (as, for example, when firms have full information and no cost of reviewing or changing prices), $h(\nu) = \nu$, a straight line with a slope of 1. Another useful benchmark is the prediction of the *Calvo model* of price adjustment, when calibrated so as to imply an average frequency of price change equal to the one that is actually observed, $\bar{\Lambda}$. In this case, $h(\nu) = \bar{\Lambda}\nu$, a straight line with a slope $\bar{\Lambda} < 1$.[43] We wish to consider to what extent either of these simple

[43]The Calvo model predicts that for each of the fraction $\bar{\Lambda}$ of firms that review their price in the current period, the current log price change is equal to ν_t plus a sum of past disturbances, the average value of which, when integrates over the possible past disturbances, is ν_t. For each of the remaining fraction $1 - \bar{\Lambda}$ of firms, the log price change is zero. Averaging over all firms, one obtains

theories is similar to the actual shape of the function $h(\nu)$.

Figure 5 plots the function $h(\nu)$, for each of the several possible values of θ considered in Table 1. The figure also plots the benchmark of full neutrality (shown as a dashed line on the diagonal).[44] One observes that in all cases, there is less than full immediate adjustment of prices to a purely monetary shock, in the case of small shocks ($0 < h(\nu) < \nu$ for small $\nu > 0$, and similarly $\nu < h(\nu) < 0$ for small $\nu < 0$). However, there is greater proportional adjustment to larger shocks, and in fact (though this cannot be seen in all cases from the part of the plot shown in the figure) in each case the graph of $h(\nu)$ eventually approaches the diagonal (the benchmark of full neutrality) for large enough shocks of either sign. The size of shocks required for this to occur is greater the larger is θ. In the case that $\theta = 0$, one sees from the figure that there is essentially full adjustment to shocks larger than .01 in absolute value.[45] When instead $\theta = 0.004$, the convergence to full adustment is still evident, but has not quite occurred at the boundaries of the figure. For higher values of θ, nearly full adjustment occurs only for shocks much larger than any shown in the figure, though one can see from the figure (at least in the case that $\theta = 0.04$) that $h(\nu)$ increases more than proportionally with increases in ν.

Even in the case of small shocks, while there is not full adjustment to monetary shocks in the month of the shock, the average price increase is *many times larger* than would be predicted by the Calvo model, in the case of sufficiently small values of θ. Figure 6 shows a magnified view of the graph of $h(\nu)$ for small values of ν, in the case $\theta = 0$, with the prediction of the Calvo model also shown by a dashed line. (The vertical axis has been stretched so as to make the slope of the line representing the Calvo prediction more visible.) The slope of the curve $h(\nu)$ near the origin is several times greater than $\bar{\Lambda}$, the slope predicted by the Calvo model.

However, for larger values of θ, the Calvo model provides quite a good approximation, in the case of small enough shocks. Figure 7 shows a similarly magnified view of the graph of $h(\nu)$ in the case $\theta = 0.04$. One observes that the prediction of the Calvo model is quite accurate, except in the case of large shocks, when it under-predicts the

an average log price change of $\bar{\Lambda}\nu_t$.

[44]The Calvo benchmark cannot be plotted as any single line in this figure, as it depends on the value of $\bar{\Lambda}$, and the value of $\bar{\Lambda}$ is different for the different values of θ, as shown in Table 3.

[45]These are still quite large shocks: note that they are more than 10 standard deviations away from the mean. The range of shocks that are two standard deviations or less from the mean is indicated by the two vertical lines in the figure.

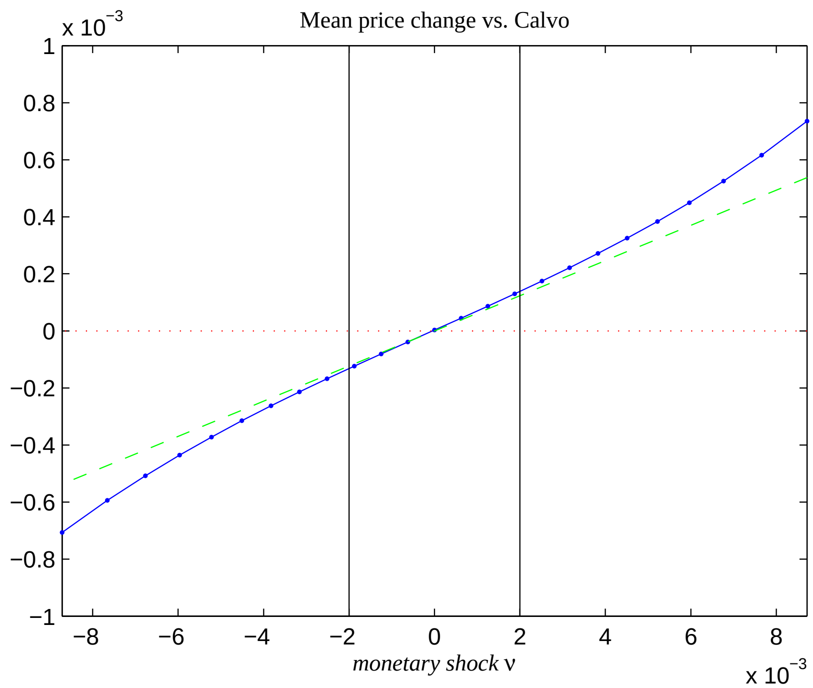

Figure 7: A closer view of the function $h(\nu)$, for the case $\theta = 0.04$. The dashed line again shows the prediction of the Calvo model. The two vertical lines indicate shocks of ± 2 standard deviations in magnitude.

average price change. For even larger values of θ (not shown here), the approximation is even better, and the range over which the approximation is accurate extends to even larger shock sizes.

Even for information costs of this magnitude, of course, the Calvo model becomes quite a poor approximation in the case of very large shocks. Figure 8 shows the graph of $h(\nu)$ in the case that $\theta = 0.04$ again, but now for a larger range of values for ν. Both of the two simple baselines, the full-neutrality prediction and the Calvo prediction, are shown by dashed lines. One observes that the Calvo model is a good approximation to the actual shape of $h(\nu)$ in the case of small enough shocks, while the full-neutrality benchmark is a good approximation in the case of large enough shocks of either sign.

While the Calvo model remains a poor approximation in the case of large shocks, when θ takes an intermediate value, it may nonetheless be a good approximation most of the time. The vertical lines in Figures 7 and 8 indicate shock sizes that are

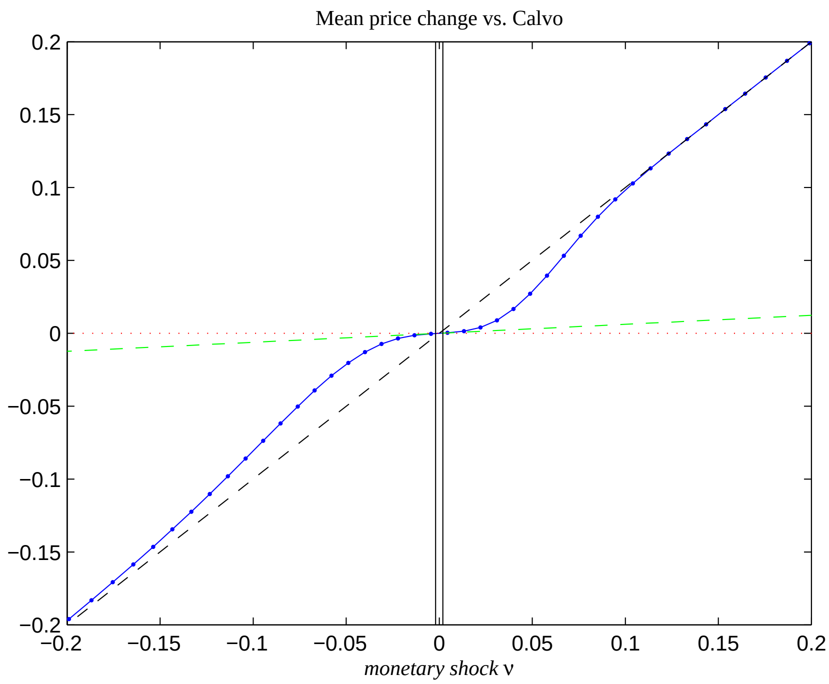

Figure 8: The function $h(\nu)$ for the case $\theta = 0.04$ again, but for a larger range of shock sizes. The dashed lines indicate the full-neutrality benchmark (the steeper line) and the Calvo benchmark (the flatter line).

plus or minus two standard deviations in magnitude; thus under the assumed shock process, shocks should fall within this range about 95 percent of the time. Within this range (as can be seen most clearly in Figure 7), the Calvo model is quite a good approximation. The same is true for even larger values of θ; shocks of the size required for the Calvo approximation to become inaccurate become progressively less likely, the larger is θ.

One way of measuring the extent to which the inaccuracy of the Calvo approximation matters in general is by considering the slope of a linear regression of the log price change on the size of the current aggregate shock. Suppose that we approximate the function $h(\nu)$ by a linear equation,

$$\Delta p_t(i) = \alpha + \beta \nu_t + \epsilon_t(i),$$

where the residual is assumed to have mean zero and to be orthogonal to the aggregate shock, and estimate the coefficients α and β by ordinary least squares. Under the

Table 3: The coefficient β from a regression of log price changes on the current monetary shock, for alternative values of θ. The value of $\bar{\Lambda}$ implied by the stationary optimal policy in each case is shown for purposes of comparison. (Both quantities reported in percentage points.)

θ	$\bar{\Lambda}$	β
0	2.8	31.0
.004	4.0	8.2
.04	6.2	6.8
.4	6.87	6.94
4	6.93	6.94
40	6.97	6.97
∞	6.98	6.98

full neutrality benchmark, β would equal 1; the Calvo model predicts that β should equal $\bar{\Lambda}$.

The values of β obtained from simulations of the stationary optimal policies corresponding to the different values of θ are given in Table 3, which also reports the values of $\bar{\Lambda}$ implied by each of these policies. One observes that the Calvo model under-predicts the flexibility of prices very substantially in the full-information case ($\theta = 0$), which is to say, in a standard SDP model of the kind studied by Golosov and Lucas (2007). For the parameter values assumed here, I find that the correct linear response coefficient is more than 10 times as large as the one predicted by the Calvo model. In the case of only small positive information costs, $\theta = 0.004$, the Calvo model also under-predicts, but only by a factor of 2. If $\theta = 0.04$, the correct coefficient is only about 10 percent larger than the prediction of the Calvo model. The Calvo model is even more accurate if information costs are larger; for example, if $\theta = 0.4$, it under-predicts the immediate price response by only 1 percent. In the limiting case of unboundedly large θ, the Calvo model is perfectly accurate.

For shocks that are large, but not large enough for nearly full adjustment to occur immediately, the average response of the price level to a monetary shock falls somewhere between the predictions of the Calvo model and the full-neutrality benchmark. It is interesting to note that while each of these benchmarks is completely antisymmetric (the effect of a negative shock is precisely the effect, with the sign reversed, of a positive shock of the same size), the effects of shocks of an intermediate magnitude are *asymmetric*. A positive shock results in more nearly complete price adjustment, on average, than does a negative shock of the same size. (Compare, for example, the effects of shocks of size ± 0.1 in Figure 8: there is nearly complete adjustment of the average price in the case of the positive shock, but much less than complete adjustment to the negative shock.) This is a direct result of the asymmetry of the optimal hazard function, already observed in Figure 3. (Because firms with prices that are too low are more likely to immediately adjust their prices than firms with prices that are too high, more adjustment occurs immediately in response to a positive shock than to a negative shock.) The result implies, in turn, that the effects of a contraction of nominal aggregate demand on real activity will be greater than the effects of an expansion of nominal aggregate demand by the same number of percentage points; for more of the positive demand disturbance will be dissipated in an immediate increase in prices than occurs in the case of a negative disturbance. This conclusion, of course, echoes a feature often found in old-fashioned Keynesian models, which assumed that prices (or wages) were "downwardly rigid" but not upwardly rigid to the same extent. The present model justifies similar behavior as a consequence of optimization; but the reason here is not any resistance to price declines — instead, firms are *more worried* about allowing their prices to remain too low than they are about allowing them to remain too high.

3.3 The Distribution of Price Changes

The model makes predictions, of course, about the complete distribution of individual price changes, and not only the mean response of prices to a shock. These are of interest, among other reasons, because they can be compared with evidence from micro data sets to compare the degree of empirical realism of alternative models. It does not make sense to attempt any detailed comparison of the predictions of this "partial equilibrium" model to the properties of actual price distributions, because the maintained hypothesis — that the general price level perfectly tracks variations in aggregate nominal spending — is plainly false in the settings from which the actual price data have been collected. Nonetheless, a brief consideration of the way in which alternative values of θ affect the distribution of price changes in this model may be

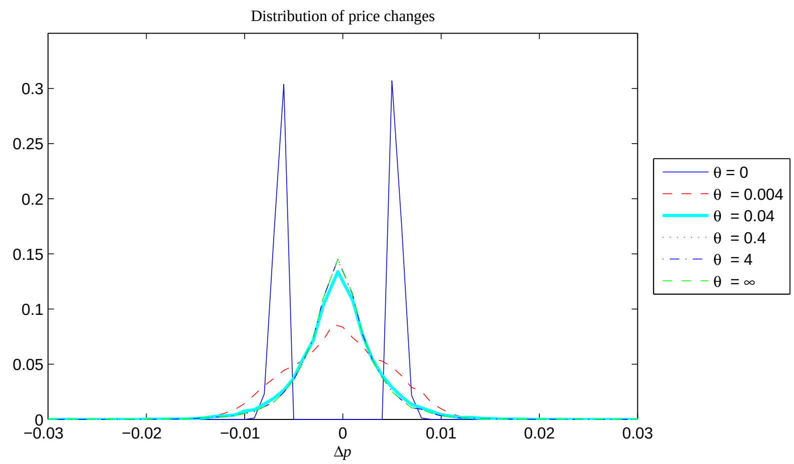

Figure 9: The (unconditional) distribution of individual price changes, for alternative values of θ.

useful in judging the likely importance of the proposed information friction.

For each of the stationary optimal policies discussed above corresponding to alternative values of θ, we can plot a long-run frequency distribution of individual price changes, obtained by stochastic simulation.[46] These distributions are shown in Figure 9 for the same six values of θ as are compared in Figure 5. For the case $\theta = 0$, we obtain the kind of distribution familiar from previous studies of standard (full-information) SDP models (see, *e.g.*, Midrigan, 2006): the distribution is largely concentrated around two spikes (one positive and one negative), corresponding to the size of price changes that occur when one just reaches the upper and lower Ss triggers respectively.[47]

This prediction is dramatically changed by the introduction of limited attention. Even when the information costs remain relatively small (the case $\theta = 0.004$), the implied distribution of price changes is now unimodal, with a peak near zero. However, in this case, there remains a relatively large amount of the probability mass far

[46] Of course, this is to be interpreted as only the distribution of price changes by the small number of firms subject to the information costs; under the "partial equilibrium" assumption, in the economy as a whole the distribution of price changes is identical to the distribution $g(\nu)$, regardless of the value of θ.

[47] The distribution of price changes does not consist of two atoms at those exact numerical values because of the assumption of discrete time, which does not allow all firms to change price when their normalized price is just crossing one of the trigger points.

from zero relative to the variance; that is, the distribution is platykurtic. For values of θ equal to 0.04 or higher, instead, the distribution remains fairly similar; that is, even for the value $\theta = 0.04$, the distribution is not too different from the limiting distribution as θ becomes unboundedly large. For each of these values (as for the Calvo model), the distribution is *leptokurtic.* That is, there are a larger number of cases in which the price changes are either small (relative to the standard deviation) or large, than would be the case under a normal distribution. As Midrigan (2006) notes, a number of data sets on individual price changes have this property, which poses a problem for standard SDP models.

This observation allows us a conjecture about which of the alternative values of θ considered in the previous tables and figures are more likely to be empirically realistic. In order for the model to generate a leptokurtic distribution of price changes, the information cost θ must not be *too small;* in the calibrated example considered here, one needs to have a value on the order of $\theta = 0.04$ or higher. But these are exactly the cases in which we have concluded that, most of the time, the Calvo model will provide a fairly accurate approximation. This is not accidental, as the Calvo model generates a leptokurtic distribution of price changes (in the case of normally distributed innovations in the profit-maximizing price), while any case in which the hazard function remains substantially smaller for price gaps near zero than it is on average will tend instead to generate a platykurtic distribution, if not a bimodal one.

A serious effort to determine the most realistic parameterization of the model by seeking to match properties of empirical distributions of price changes will depend upon extending the analysis to more complex cases than the one considered here. Here it has only been possible to present a "partial equilibrium" analysis: the dynamics of the average price of firms subject to menu costs and information costs, as well as the distribution of individual price changes by such firms, have been considered in a setting in which it is assumed that the *aggregate* price level adjusts immediately in proportion to any change in aggregate nominal expenditure. This simple case is convenient to analyze because individual firms' decisions depend on no aggregate state variables (under the assumption that aggregate nominal expenditure is a random walk), so that the adjustment hazard is a function of a single real variable, the individual firm's "price gap," as in the generalized Ss framework of Caballero and Engel (1993a, 2007). Analysis of this case has sufficed to show that in the present model, random variations in aggregate nominal expenditure will affect aggregate real activity, rather than being neutral as in Caplin and Spulber (1987). It is also relatively easy to see, in this case, in what ways the Calvo model is and is not an accurate approximation to the dynamics of price adjustment in a model of information-constrained state-dependent pricing.

But realistic predictions about the overall distribution of price changes would require us to solve for the "general equilibrium" dynamics of prices when *all* firms are subject to menu costs and information costs, and the response of the aggregate price level to aggregate shocks is determined by aggregating the decisions of the population of such firms. While the extension of the model equations in section 2 to this case is mathematically straightforward, the "general equilibrium" case is computationally much more challenging. In an exact solution, the state space of the model will be infinite-dimensional, even when the dynamics of aggregate nominal expenditure are as simple as those assumed here (as is true of full-information SDP models except in extremely special cases). Hence a solution for the approximate equilibrium dynamics is likely to be possible only under an assumption of "bounded rationality" in the spirit of Krusell and Smith (1998), as for example in the work of Midrigan (2006). This is an important topic for future work.

4 Conclusion

I have presented a model in the timing of price changes results from optimizing behavior on the part of firms subject to a fixed cost of conducting a review of existing pricing policy. Standard models of state-dependent pricing, however, are generalized by assuming that a firm's policy with regard to the timing of price reviews is designed to economize on the cost of being continuously informed about market conditions during the intervals between full-scale reviews. The introduction of interim information costs softens the distinction, emphasized in prior contributions, between the dynamics of price adjustment in models with exogenous timing of price adjustments and models with state-dependent pricing, by attenuating both the "selection effect" emphasized by Golosov and Lucas (2007) and the relative importance of the "extensive margin of price adjustment" emphasized by Caballero and Engel (2007). In the limiting case of sufficiently large interim information costs, the predicted dynamics of price adjustment are identical to those of the Calvo (1983) model of staggered price-setting.

At a minimum, this result means that there is no reason to regard the predictions of (full-information) "menu cost" models as more likely to be accurate than the predictions of the Calvo model, simply on the ground that the former models have firmer foundations in optimizing behavior. Both models appear as nested (extreme) cases of the more general model presented here, so that the question of which special case is more reliable as an approximation is a quantitative matter, rather than one that can be settled simply on the basis of the appeal of optimizing models.

The illustrative calculations presented in section 3 furthermore suggest that a model with interim information costs of moderate size may imply aggregate behavior fairly similar to that predicted by the Calvo model, and quite different from that predicted by a full-information menu-cost model. Further work is needed to investigate to what extent this conclusion obtains in the case of empirically realistic parameterizations. But these calculations show that it is possible for predictions of the Calvo model to be fairly accurate for many purposes — predicting the aggregate responses to disturbances of the magnitude that occur at most times — in a model that does not possess certain features of the Calvo model that are often argued to be implausible. In particular, the model with a finite positive interim cost of information does *not* imply that prices are equally unlikely to be adjusted even when a given firm's price happens over time to have become far out of line with profit maximization, or even when very large disturbances affect the economy. However, because firms are in these situations only very infrequently, the predictions of the Calvo model may nonetheless be relatively accurate much of the time.

It is important to note, however, that the implications of the present model are likely to differ from those of the Calvo model in important respects, even if a relatively large value of θ is judged to be empirically realistic. First, even if the price adjustments predicted by this model are similar to those of the Calvo model under all but extreme circumstances, the model's predictions under extreme circumstances may be of disproportionate importance for calculations of the welfare consequences of alternative stabilization policies, as argued by Kiley (2002) and Paustian (2005). And second, even in the limit of an unboundedly large value of θ (so that no interim information is available at all), the present model's predictions differ from those of the Calvo model in at least one important respect: the equilibrium frequency of price review $\bar{\Lambda}$ is endogenously determined, rather than being given exogenously. In particular, the value of $\bar{\Lambda}$ is unlikely to be policy-invariant; for example, one would expect it to be higher in the case of a higher average inflation rate, as in the generalized Calvo model of Levin and Yun (2007). For this reason as well, the present model may well have different implications than the Calvo model for the welfare ranking of alternative policy rules, as in the analysis of Levin and Yun. This is another important topic for further study.

References

[1] Bonomo, Marco, and Carlos Carvalho, “Endogenous Time-Dependent Rules and Inflation Inertia,” *Journal of Money, Credit and Banking* 36: 1015-1041 (2004).

[2] Burstein, Ariel T., and Christian Hellwig, “Prices and Market Share in a Menu Cost Model,” unpublished, U.C.L.A., September 2007.

[3] Caballero, Ricardo J., “Time Dependent Rules, Aggregate Stickiness, and Information Externalities,” Discussion Paper no. 428, Department of Economics, Columbia University, May 1989.

[4] Caballero, Ricardo J., and Eduardo Engel, “Microeconomic Rigidities and Aggregate Price Dynamics,” *European Economic Review* 37: 697-717 (1993a).

[5] Caballero, Ricardo J., and Eduardo Engel, “Microeconomic Adjustment Hazards and Aggregate Dynamics,” *Quarterly Journal of Economics* 108: 359-383 (1993b).

[6] Caballero, Ricardo J., and Eduardo Engel, “Explaining Investment Dynamics in U.S. Manufacturing: A Generalized *(S,s)* Approach,” *Econometrica* 67: 783-826 (1999).

[7] Caballero, Ricardo J., and Eduardo Engel, “Price Stickiness in Ss Models: Basic Properties,” unpublished, M.I.T., October 2006.

[8] Caballero, Ricardo J., and Eduardo Engel, “Price Stickiness in Ss Models: New Interpretations of Old Results,” *Journal of Monetary Economics* 54 (supp. 1): 100-121 (2007).

[9] Calvo, Guillermo A., “Staggered Prices in a Utility Maximizing Framework,” *Journal of Monetary Economics* 12: 383-398 (1983).

[10] Caplin, Andrew S., and Daniel F. Spulber, “Menu Costs and the Neutrality of Money,” *Quarterly Journal of Economics* 102: 703-726 (1987).

[11] Cover, Thomas M., and Joy A. Thomas, *Elements of Information Theory*, New York: Wiley-Interscience, 2d ed., 2006.

[12] Dotsey, Michael and Robert G. King, "Implications of State-Dependent Pricing for Dynamic Macroeconomic Models," *Journal of Monetary Economics* 52: 213-242 (2005).

[13] Dotsey, Michael, Robert G. King, and Alexander L. Wolman, "State Dependent Pricing and the General Equilibrium Dynamics of Money and Output," *Quarterly Journal of Economics* 114: 655-690 (1999).

[14] Gertler, Mark, and John Leahy, "A Phillips Curve with an Ss Foundation," unpublished, New York University, May 2007.

[15] Golosov, Michael, and Robert E. Lucas, Jr., "Menu Costs and Phillips Curves," *Journal of Political Economy* 115: 171-199 (2007).

[16] Kiley, Michael T., "Partial Adjustment and Staggered Price Setting," *Journal of Money, Credit and Banking* 34: 283-298 (2002).

[17] Krusell, Per, and Anthony A. Smith, Jr., "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy* 106: 867-896 (1998).

[18] Levin, Andrew, and Tack Yun, "Reconsidering the Natural Rate Hypothesis in a New Keynesian Framework," *Journal of Monetary Economics* 54: 1344-1365 (2007).

[19] Midrigan, Virgiliu, "Menu Costs, Multi-Product Firms, and Aggregate Fluctuations," unpublished, Federal Reserve Bank of Minneapolis, August 2006.

[20] Miranda, Mario J., and Paul L. Fackler, *Applied Computational Economics and Finance,* Cambridge: M.I.T. Press, 2002.

[21] Nakamura, Emi, and Jon Steinsson, "Monetary Non-Neutrality in a Multi-Sector Menu Cost Model," unpublished, Columbia University, October 2007.

[22] Paustian, Matthias, "The Role of Contracting Schemes for the Welfare Costs of Nominal Rigidities over the Business Cycle," unpublished, University of Bonn, April 2005.

[23] Phelps, Edmund S., *Seven Schools of Macroeconomic Thought,* Oxford: Clarendon Press, 1990.

[24] Reis, Ricardo, "Inattentive Producers," *Review of Economic Studies* 73: 793-821 (2006).

[25] Rubinstein, Ariel, "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory* 39: 83-96 (1986).

[26] Shannon, Claude E., "A Mathematical Theory of Communication," *Bell System Technical Journal* 27: 379-423 and 623-656 (1948).

[27] Sims, Christopher A., "Stickiness," *Carnegie-Rochester Conference Series on Public Policy* 49: 317-356 (1998).

[28] Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics* 50: 665-690 (2003).

[29] Sims, Christopher A., "Rational Inattention: A Research Agenda," unpublished, Princeton University, March 2006.

[30] Woodford, Michael, *Interest and Prices: Foundations of a Theory of Monetary Policy,* Princeton: Princeton University Press, 2003.

[31] Zbaracki, Mark, Mark Ritson, Daniel Levy, Shantanu Dutta, and Mark Bergen, "Managerial and Customer Costs of Price Adjustment: Direct Evidence from Industrial Markets," *Review of Economics and Statistics* 86: 514-533 (2004).