Federal Reserve Bank of Minneapolis
Research Department Staff Report 520

November 2015

Coarse Pricing Policies*

Luminita Stevens

Federal Reserve Bank of Minneapolis
and University of Maryland

ABSTRACT

The puzzling behavior of inflation in the Great Recession and its aftermath has increased the need to better understand the constraints that firms face when setting prices. Using new data and theory, I demonstrate that each firm's *choice* of how much information to acquire to set prices determines aggregate price dynamics through the patterns of pricing at the micro level, and through the large heterogeneity in pricing policies across firms. Viewed through this lens, the behavior of prices in recent years becomes less puzzling, as firms endogenously adjust their information acquisition strategies. In support of this mechanism, I present micro evidence that firms price goods using plans that are sticky, coarse, and volatile. A theory of information-constrained price setting generates such policies endogenously, and quantitatively matches the discreteness, duration, volatility, and heterogeneity of policies in the data. Policies track the state noisily, resulting in sluggish adjustment to shocks. A higher volatility of shocks does not reduce monetary non-neutrality and generates slight inflation, while progress in the technology to acquire information results in deflation.

Keywords: rigid prices, rational inattention, inflation dynamics
JEL classification: E3, E5

*This paper supersedes the earlier working papers Stevens (2011; 2012). I am grateful to Mike Woodford and Ricardo Reis for their valuable suggestions and support. I would also like to thank Boragan Aruoba, Ryan Chahrour, Doireann Fitzgerald, Christian Hellwig, Pat Kehoe, Ben Malin, Filip Matějka, Virgiliu Midrigan, Emi Nakamura, Jaromir Nosal, Ernesto Pasten, Chris Sims, Jón Steinsson, and participants at numerous seminars and conferences for helpful comments. Camilo Morales-Jimenez and Dun Jia provided excellent research assistance. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. Contact: stevens@econ.umd.edu.

1 Introduction

The behavior of inflation in the Great Recession and its aftermath has increased the need to better understand how firms set prices and what constraints they face when responding to changes in economic activity. On the one hand, the U.S. experienced only a mild disinflation during the Great Recession.[1] On the other hand, after a modest pick-up, inflation started declining again in 2012, despite strengthening economic activity and falling unemployment. This paper uses new data and theory to argue that information frictions, specifically each firm's *choice* of how much information to acquire to set prices, depending on its own characteristics and market conditions, play a key role in determining aggregate price dynamics, through the patterns of pricing at the micro level, and through the large heterogeneity in pricing policies across products.

In support of the mechanism of endogenous information acquisition, I present evidence that firms appear to price goods according to simple price plans that are updated relatively infrequently. These simple price plans suggest that firms optimize by seeking to economize on the costs of monitoring and responding to continuously evolving market conditions. Empirically, I identify price plans by searching for breaks in individual price series. To ensure a large degree of generality, breaks are identified by any change in the distribution of prices charged over time, using an adaptation of the Kolmogorov-Smirnov test. I apply this break test to product level prices from AC Nielsen's new Retail Scanner Data. This database has weekly point-of-sale data for a very large number of products sold in grocery, drug, mass merchandiser, and other stores all across the U.S., from 2006 through 2012.[2]

First, I establish three facts about pricing policies at the product level. First, policies are *sticky*, changing every seven months, even though individual prices change every three weeks. Second, policies are *coarse*, typically consisting of only three distinct price points, despite the large weekly frequency of within-policy price changes. This finding points to the "disproportionate importance" of a few price points, consistent with similar evidence at the *series* level documented by Klenow & Malin (2010) using micro data underlying the CPI. The discreteness of prices coupled with the high frequency of adjustment suggests that while the timing of price changes within a policy is quite flexible, the level to which the price adjusts is more rigid. Third, policies are *volatile*, with prices changing by 10% in absolute value within policies and shifting by 11% in absolute value once the policy is

[1] See Hall (2011), Ball & Mazumder (2011) and Del Negro, Giannoni & Schorfheide (2015).

[2] This data set is very new. It has also been used by Beraja, Hurst & Ospina (2014) to analyze dynamics in regional price indices. The data is from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. For more information see http://research.chicagobooth.edu/nielsen/.

updated. Hence, the volatility of prices in the micro data reflects prices alternating among a small set of relatively dispersed price points that are updated relatively infrequently. A theory of information-constrained optimization can endogenously generate such simple price plans that crudely track the optimal full information price.

Making the distinction between policy changes and raw price changes when characterizing product-level price dynamics proves useful because changes along the two dimensions may be driven by different forces. For instance, within-policy volatility may be primarily driven by transitory shocks or price discrimination motives, while the shift in average prices across policies may be driven by more persistent shocks. Hence, this distinction can be used to discriminate among different theories of price setting and among different potential sources of price volatility. The dynamics of price and policy adjustment over time illustrate this point: while the rate of policy adjustments was particularly elevated during the Great Recession, neither the rate nor the size of price changes showed any change. A potential explanation for this pattern is that the Great Recession was a period of heightened uncertainty to which firms responded not by making their pricing plans more complex (which would affect the frequency and size of raw price changes), but rather, by keeping their price plans simple and reviewing them often, until the uncertainty was resolved.

Next, I document substantial heterogeneity in the type of pricing policies employed across different products. All products can be classified into one of three types: single-price, one-to-flex, and multiple-price. Products characterized by *single-price policies* (SPP), such as those generated by the canonical time-dependent or state-dependent models, represent 10% of all product series. The prices of these products adjust much less frequently and by smaller amounts conditional on adjustment, hence these products face a relatively low volatility of their desired price that does not warrant designing complex pricing policies. Products characterized by *one-to-flex policies* (OFP), in which a single sticky price is accompanied by transitory price changes to and from it, and in which none of the transitory price levels are revisited, represent 19% of all product series. Such patterns are consistent with prior theoretical work in which firms flexibly deviate from regular prices that are assumed to be rigid (Kehoe & Midrigan (2010) and Guimaraes & Sheedy (2011)). These products face a higher volatility in their desired price, compared with the SPP products. They adjust policies more frequently and by larger amounts. Nonetheless, the policies themselves are neither very volatile nor complex. The muted within-policy volatility suggests that OFP products have a relatively high volatility in their desired price, but also a potentially high cost of implementing more complex pricing policies. Hence, they adapt by updating their policies more frequently and by shifting their prices by larger amounts upon adjustment. Finally, underscoring the presence of rigidity beyond the modal price within each price plan, 71% of product series are characterized by *coarse multiple price policies* (MPP), consisting of a small number of rigid price points that are revisited over the life of the policy. The volatility of the data is concentrated in these series: policies are updated every 6.5 months and shift by 12% upon adjustment, and, despite the low cardinality of the pricing policy (only four distinct prices are typically charged between policy changes), the frequency and absolute size of within-policy price changes are both large, at 33% and 10%, respectively. Hence MPP products face highly volatile market conditions, to which they respond in two ways: first, they choose more complex, though nonetheless coarse, pricing policies, and second, they update their policies frequently, and upon adjustment, they shift by large amounts.

The attempt to categorize products by the type of pricing policy employed, rather than by the frequency of price adjustment alone, is also a novel, useful way to characterize the heterogeneity in the data. In particular, I show that inflation dynamics during the Great Recession varied substantially across the three policy types: while the inflation rates for all types of products moved in tandem leading up to the fourth quarter of 2008, once inflation started to fall, it fell twice as far for MPP products than for the SPP and OFP products. Moreover, SPP and OFP products continued to raise prices throughout the crisis, while MPP products actually cut prices. This evidence is consistent with that provided by Gilchrist, Schoenle, Sim & Zakrajsek (2014), who find that at the peak of the crisis, firms operating in competitive markets lowered their prices significantly, relative to firms operating in less competitive markets. These heterogeneous responses across policy types can also account for the increased dispersion in price changes documented by Vavra (2014), as some types of products are more responsive to changes in economic conditions than others. The information-based theory presented in this paper predicts precisely these effects: firms that operate in more volatile or more competitive markets have an incentive to acquire more information about market conditions, and hence they will choose more complex pricing policies, and they will respond to shocks more aggressively. These findings also underscore the importance of studying price data in its entirety, rather than filtering out transitory price volatility: transitory volatility is in fact crucial to pinning down the type of pricing policy employed by different firms and, as we have seen, the type of policy chosen by firms in turn affects how these firms respond to shocks, and hence it affects aggregate inflation dynamics.

The empirical analysis adds to a large literature on product-level price patterns (see Klenow & Malin (2010) and Nakamura & Steinsson (2013) for reviews). That literature has focused a lot on transitory sales from rigid regular prices. I depart from that approach by interpreting both the transitory and the regular price levels as chosen to be jointly optimal, as part of an integrated pricing policy. This integrated approach suggests a departure from existing theoretical work on micro price patterns, which either imposes distinct technologies for changing regular versus sales prices (e.g. Kehoe & Midrigan (2010) or Guimaraes & Sheedy (2011)), or abstracts from transitory price changes altogether. The coarseness and rigidity of pricing policies is instead consistent with the simple price plans hypothesized by Eichenbaum, Jaimovich & Rebelo (2011), in which firms are assumed to choose prices from a small set that is updated infrequently, subject to a cost. The theory presented in this paper generates such policies endogenously in a dynamic model of information choice. I show that once one allows the firm to occasionally revise its pricing policy, coarse, discrete pricing and large transitory volatility arise endogenously in an otherwise standard infinite-horizon price setting model. In this context, heterogeneity in pricing policies, namely the coexistence of single-price, one-to-flex, and multiple-price policies, also arises naturally if one allows firms to differ in the volatility of the shocks to their profit functions, the curvature of their profit functions, or the managerial or informational costs of monitoring market conditions and of redesigning their policies.

In the model, firms choose the type and quantity of information to acquire about the state of the world. A firm's policy specifies (i) how information is acquired and used to set prices and (ii) since the policy itself can be reviewed, how information is acquired and used to decide whether or not to undertake a policy review. The theory builds on papers in the imperfect information and rational inattention literature, primarily Reis (2006), Woodford (2009), and Matějka (2010), combining both fixed and variable costs of information. First, to generate infrequent breaks to new policies, I introduce a fixed cost that enables the firm to learn the state and to revise its policy. The firm can choose to implement a single-price or a multiple-price policy, depending on the trade-offs it faces between the expenditure required to design a more complex policy and the profits gained from tracking market conditions more closely. If the firm finds it optimal to implement a multiple-price policy, then between reviews it acquires a signal in each period to decide which price to charge from the menu of available prices specified by the policy currently in effect. Expenditure on this pricing signal is linear in the signal's desired informativeness about current economic conditions. Additionally, the firm also monitors market conditions in order to decide whether its current policy has become obsolete relative to the evolution of market conditions. In order to make this review decision, the firm receives a second signal in each period, which indicates the desirability of paying the fixed cost and updating its policy. This second signal can also be chosen to be more or less precise, depending on the value that the firm places on accurately timing the policy revisions. Hence, I model a dual decision problem that specifies rules for making a *review decision* and a *pricing decision* in each period. The measurement of the information acquired to make each decision follows the rational inattention literature (Sims, 2003; 2006), using Shannon's (1948) relative entropy function. How much information to acquire in order to make each decision is under the firm's control, as firms facing different environments (either in the cross-section or over time) may choose different information expenditure levels for one or both decisions.

The setup can be seen as modeling the relationship between headquarters (which decides and communicates the policy) and the branch level (which implements the policy day-to-day). Alternatively, the setup can also be seen as a reduced-form representation of the relationship between the manufacturer (or distributor) and the retailer: the overall policy is the result of (relatively infrequent) negotiations between the two parties, while the exact implementation of the policy (for instance, when to implement a sale) is largely left to the discretion of the retailer.[3]

The firm's optimal policy consists of three elements. The first element is an endogenous hazard function that specifies the probability of conducting a policy review conditional on the current state, for all states and periods between reviews. This hazard function determines the frequency with which the firm revises its policy and its accuracy in making this revision. It generalizes to the case of policy adjustments the hazard function for individual adjustment first postulated by Caballero & Engel (1993) for employment and derived in the context of an information choice model for price setting by Woodford (2009). Its shape is determined jointly by the shape of the firm's profit function and by the shape of the distribution of shocks that accumulate between policy reviews, and it plays a key role in determining the shape of the distribution of prices that are charged between reviews. The second element of the firm's policy is the set of prices to be charged until the next review. This set is determined by optimality conditions that equate the net benefit of charging different prices over the life of the policy. Depending on the trade-off between pricing accuracy and information costs, the firm chooses a smaller or a larger set of prices with which to track the state between reviews, with a larger set of prices requiring more expenditure on information between reviews. The third element is a conditional distribution that specifies which price to charge conditional on the current state, given the set of prices chosen. Together with the evolution of market conditions, this distribution determines the frequency with which the firm charges different prices between reviews and the extent to which the firm makes pricing mistakes in each period. Prices vary stochastically with the state, and policy reviews are stochastically state-dependent and independent of the time elapsed since the last review. The random relationship between each of the two decisions and the current state is a result of the firm's need to economize on information costs.

The setup delivers several novel results. First, the model can be parameterized to endoge-

[3] See, for example, Anderson, Jaimovich & Simester (2012) for a discussion of the pricing practices of a US national retailer.

nously yield discrete prices in an infinite horizon setting with normally distributed shocks. The resulting optimal policy is updated infrequently and specifies a small set of prices relative to the set of prices that would be charged under full information. Both the coarseness and the stickiness of the resulting policy reflect the firm's desire to economize on the costs of monitoring market conditions. The paper discusses cases in which the optimal pricing policy is discrete versus continuous and illustrates how the support of the pricing policy evolves as a function of model parameters.

Second, either a single-price or a multiple-price policy may be optimal, depending on parameter values, such as the costs of processing information, the volatility of shocks, and the curvature of the profit function. Third, among multiple-price policies, a smaller or a larger set of prices may be chosen, also depending on parameter values. Hence, the theory can generate heterogeneity in the complexity of pricing policies chosen by firms in different sectors or over time. In particular, the theory can generate the SPP, OFP and MPP types of policies documented in the empirical part of the paper.

Finally, I show quantitatively that the model can be parameterized to match the discreteness, duration and volatility of policies in the data. Generating pricing patterns consistent with the data requires moderate expenditure on information.

Allowing the firm to choose how much information to acquire in order to make its policy and pricing decisions is critical to generating both discreteness in price levels and heterogeneity in pricing policies across products. But information choice also has strong implications for aggregate dynamics. I explore the model's implications for three types of scenarios: short-run, medium-run and long-run.

First, in the short run, I obtain a sluggish response to nominal shocks that is completely divorced from the frequency of price changes. Moreover, the firm's choice to change prices between policy reviews does not reduce the model's implied aggregate rigidity relative to that implied by the single-price-policy parameterization. The impulse response functions are essentially identical for both single-price and multiple-price policies. This finding reflects the inter-dependence between the firm's pricing policy and its review policy: the SPP firm chooses to spend more resources on designing an accurate review policy and no resources on the pricing policy, while the MPP firm chooses to spend more resources on the more complex pricing policy, but then needs less accuracy in the timing of policy reviews. The fact that high price volatility does not necessarily imply low monetary neutrality has been discussed in the literature, by Kehoe & Midrigan (2010) and Eichenbaum et al. (2011). However, this paper generates this result in the context of a model in which the firm chooses its policy optimally, rather than having certain aspects of the policy be exogenously assumed. In particular, it endogenously generates the price plans postulated by Eichenbaum et al. (2011).

Second, in terms of responses to structural changes, the model predicts that higher volatility does not generate higher aggregate flexibility. This result stands in contrast to the predictions of full-information state-dependent pricing models, and it reflects the endogenous response of the firm's information acquisition strategy: although the firm increases information expenditure, it nevertheless has less information relative to the uncertainty it faces in the new, higher volatility environment. Given the information costs it faces, it is not optimal for the firm to completely undo the effects of increased volatility. Hence on net, even though the firm acquires more information than before, it still generates the same degree of non-responsiveness as before. Additionally, the increase in volatility also generates a modest degree of inflation, as firms raise prices to protect themselves against the losses that come from underpricing in a more uncertain environment.

Finally, in terms of longer run structural changes, increased competition and progress in the technology to acquire information both result in modest deflation. Increased competitive pressures imply that each firm faces larger potential losses from mispricing. Hence each firm acquires more information to maintain its profits. The firm's increased ability to track market conditions in turn implies that it can charge a lower price on average. Similarly, technological progress that lowers information costs also results in more complex pricing policies that better track market conditions, thereby also implying lower prices. Hence, in addition to the other factors highlighted in the literature, such as better monetary policy or smaller shocks, low modern inflation rates may also be partially attributable to information costs trending down and to competitive pressures rising over time.

The theoretical contribution of this paper builds on the very large literature of price setting under imperfect information, in particular the work of Woodford (2009), Matějka (2010), and Reis (2006).[4] In Matějka (2010) the decision of which price to charge in each period is based on noisy signals, chosen subject to an information processing capacity limit, and resulting in errors in the *size* of price adjustment. That paper shows that assuming a uniform distribution for shocks yields discrete prices and it links the model to micro facts on the distribution of markups and the transitory nature of many price changes. In Woodford (2009), once the firm decides to change its price, the price charged is the optimal one, hence, unlike in Matějka (2010), there is no error along the size of adjustment margin. On the other hand, that paper studies a dynamic model and generates errors in the *timing* of price adjustments. That paper links the model to micro facts about the distribution of filtered *regular* price changes and also discusses aggregate non-neutrality. It connects the Calvo (1983) and menu

[4]Other models of price setting with endogenous information acquisition include Maćkowiak & Wiederholt (2009, 2010), Matějka & McKay (2011), Paciello (2012), Paciello & Wiederholt (2014) and Pasten & Schoenle (2014).

cost models as two extremes of the information-constrained model and shows the response to a monetary shock on impact as a function of the severity of the friction. In the present paper, both the *timing* and the *size* of price adjustments are subject to mistakes, because the firm acquires noisy signals in order to make both decisions. The theory also extends the discreteness results of Matějka (2010) to a dynamic, infinite-horizon model with persistent, normally distributed shocks, showing that rational inattention generates discreteness in a much wider range of settings. Finally, it generates quantitative results regarding micro price patterns and also explores the implications of this micro friction beyond the sluggishness of the aggregate price level, also looking at the effects on aggregate prices of heightened volatility, progress in information acquisition technologies, and increasing competition.

2 Empirical Evidence

The Data I use the new Retail Scanner Data provided by AC Nielsen, which contains the weekly sales of products in stores from 90 retail chains across the U.S. between January 2006 and December 2012. The data's product categories represent approximately 27% of the total goods consumption measured by the BLS's Consumer Expenditure Survey. Product categories include health and beauty care, dry grocery, non-food grocery (e.g., household cleaners), dairy, frozen foods, alcohol, and general merchandise (e.g., glassware, kitchen gadgets). I exclude the Deli, Packaged Meat, and Fresh Produce departments. I further limit the sample to data from the store with the largest number of observations from each chain. Some series have missing observations. I keep only series with contiguous observations that are at least 52 weeks long. The resulting sample contains more than 140 million observations for approximately $150,000$ universal product codes, from 88 stores. The average series length is 147 weeks and the maximum is 365 weeks.[5]

A drawback of using this retail scanner data is the relatively narrow product coverage: food, drug, and some general merchandise. On the other hand, it has the advantages of (*i*) high frequency (versus the BLS's monthly or bimonthly sampling), (*ii*) long price series (versus the BLS's much smaller number of observations per series and frequent product substitutions), and (*iii*) very large number of products within the categories (versus the BLS's much narrower sampling within product groups). Moreover, the scanner data covers products whose prices are highly volatile and exhibit precisely the sharp, transitory price swings that have come to the forefront of the price dynamics literature. The expenditure-weighted median weekly frequency of price changes is 23.0% and the expenditure-weighted

[5]The dataset is similar to, though much larger (in terms of both product and geographic coverage) than other scanner data sets such as Dominick's Finer Foods, which, when cleaned in the same way, yields less than 5,000 UPCs.

median size of price changes is 13.6% in absolute value. Hence, any rigidity uncovered in this subset of consumer goods provides a lower bound on the rigidity in the overall CPI.[6]

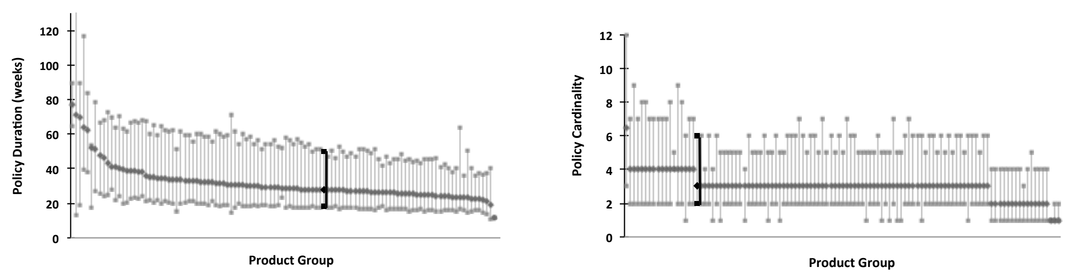
The Break Test The empirical method is based on the Kolmogorov-Smirnov test, which tests whether two samples are drawn from the same distribution. Building on tests that estimate the location of a single break in a series (Deshayes & Picard (1986) and Carlstein (1988)), I adapt the test to identify an unknown number of breaks at unknown locations in a series. The method uses an iterative procedure similar to that employed by Bai & Perron (1998) who sequentially estimate multiple breaks in a linear regression model. Specifically, I first test the null hypothesis if no break in a given price series; upon rejection, I estimate the location of the break; I then iterate on the resulting sub-samples until I identify all breaks in a series. The strength of the method depends on its ability to correctly identify the timing of breaks. In simulations, I find that the break test correctly identifies breaks 91% of the time across a mixture of different data generating processes and it finds the *exact* location of the break 94% of the time (in the remaining cases, it is off by 2 periods). In simulations restricted to generate policy realizations that last at least 5 weeks, the test finds virtually all breaks.

This method allows for the interpretation that all prices are potentially chosen to be jointly optimal, as part of an integrated pricing policy that the firm implements and occasionally updates. In principle, the test can identify any salient changes in both the support and the shape of the distribution of prices over time. Hence, it is less restrictive than filters that focus on the modal or high price within a pre-specified window. The empirical method, its robustness across different simulated data generating processes, and a comparison with filters that seek to identify changes in regular or reference prices are detailed in the Appendix.

Policy Adjustment The first empirical result is that the identified pricing policies are relatively *sticky*: policy changes typically occur every 6.5 months, even though raw prices change every three-to-four weeks. For a comparison, papers that seek to filter out transitory price volatility report the duration of prices ranging from 7.8 months to 12.7 months for

[6]All reported statistics are weighted by the expenditure share of each product group. I exclude price changes that are smaller than 1% in absolute value (10.6% of all price changes). In the full sample, the weighted frequency and size of price changes are 4 percentage points higher and 1 percentage point lower, respectively. However, as argued by Eichenbaum, Jaimovich, Rebelo & Smith (2014), very small price changes may reflect measurement error and bias price statistics. In the *Retail Scanner Data*, a price observation is the volume-weighted average price of the product for a particular week. Prices reflect bundling (e.g. 2-for-1 deals) and discounts associated with the retailer's coupons or loyalty cards. Variation in bundling or in the fraction of customers getting such discounts from one week to the next may induce spurious small price changes. The use of volume-weighted average prices also implies that my analysis provides only a lower bound of the degree of discreteness in prices.

regular prices in grocery store data (Kehoe & Midrigan (2010), Eichenbaum et al. (2011)) and from 6.7 months to 14 months for regular prices in the CPI (Klenow & Kryvtsov (2008), Nakamura & Steinsson (2008), Kehoe & Midrigan (2010)). This variation highlights the fact that price statistics are sensitive to the definition of permanent versus transitory price changes and to the filters implemented to identify them. An advantage of the break test over such price filters is precisely the fact that it sidesteps the need to take a stand on how to define and identify regular versus transitory price changes, which is the source of a big portion of the dispersion in estimates in the existing literature (beyond that arising from data coverage differences).[7] In Figure 1, panel a shows the median implied duration for each product group, ordered from highest to lowest, as well as the interquartile range. There is considerable heterogeneity across products, but most policies last between 5 and 15 months.



(a) Duration (b) Coarseness

Figure 1: Policy heterogeneity across product groups.

Data: AC Nielsen Retail Scanner Data. Median and interquartile range for duration (panel a) and cardinality (panel b) of policy realizations. The expenditure-weighted statistics for the full sample are in black.

Second, policies are *coarse*: the median number of distinct prices per policy is 3, and the vast majority of policy realizations have less than six distinct prices, as shown in panel b of Figure 1. Moreover, there is no strong correlation between the duration and the cardinality of policies, suggesting that firms value simplicity when choosing their pricing policies.

On the other hand, inside these policy realizations, prices are *volatile*, despite the low cardinality of the policy, as shown in Figure 2, which plots the frequency and size of within-policy price changes for the different product groups and for the full sample. The weighted median weekly frequency of within-policy price changes is 25%, and the weighted median size (in absolute value) of within-policy price changes is 9.7%. Hence, although the data rejects the

[7]The Appendix compares the performance of the break test to that of the filters in simulated as well as actual data. I find that among the different filters, the best performing one in simulations is that proposed by Kehoe & Midrigan (2010). Specifically, there exists a parameterization of that filter that can match the accuracy of the break test. For all the cited studies, I report the monthly implied duration = -1/ln(1-median monthly frequency).

hypothesis of single price policies, such as the canonical time-dependent or state-dependent models, it nonetheless suggests that firms choose to implement simple price plans, with a small number of price points among which to alternate, despite potentially volatile market conditions that might warrant frequent and large price changes. The data also suggests particular sources of firm heterogeneity: since there is a strong positive correlation between the frequency and the absolute size of within-policy price changes, we can rule out differences in menu costs of price adjustment alone (which would generate a negative correlation), and consider differences in the volatility of the market conditions that firms face. This evidence also highlights two dimensions of flexibility in within-policy price adjustment: flexibility in the timing of adjustment and flexibility in the level to which the price adjusts. While the level seems quite rigid, the timing appears much more flexible. The theory proposed in the next section generates the coexistence of these two features of the data.

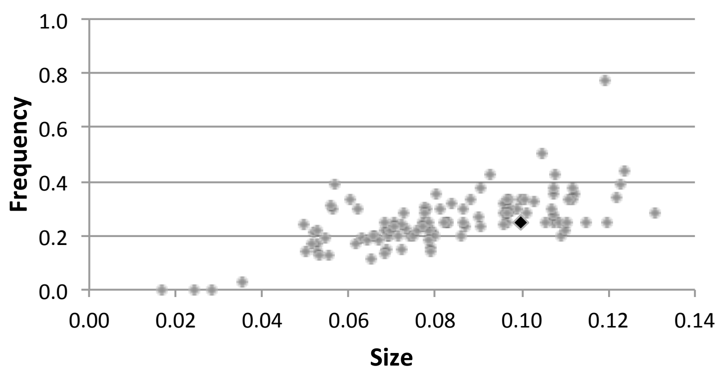


Figure 2: Frequency and size of within-policy price changes across product groups.

Data: AC Nielsen Retail Scanner Data. The median frequency is plotted against the median absolute size of within policy price changes. The outlier in the top right corner is Greeting Cards/Party Needs. The expenditure-weighted statistics for the full sample are in black.

Finally, policy changes are associated with large *shifts* in prices: the expenditure-weighted median shift in the weighted average price across consecutive policy realizations is 11.1% versus 9.7% for within-policy price changes. Policy shifts are computed by taking the average weighted price within a policy and computing the absolute value of the change in this average price. These magnitudes are consistent with prior studies which have found that prices typically change by 10%-11% (e.g. Klenow & Kryvtsov (2008)). The novelty here is distinguishing within-policy price changes from shifts in the average price level across policies, since they may be driven by different forces. For instance, within-policy volatility may be primarily driven by transitory shocks or price discrimination motives, while the shift in average prices across policies may be driven by more persistent shocks. Hence, this distinction can be used to discriminate among different theories of price setting and among different potential sources of price volatility.

Policy Heterogeneity Heterogeneity in pricing practices across goods is a very well-known and very strong feature of the data, which I confirm in the Nielsen data set. I depart from the existing literature that focuses on heterogeneity across goods in the frequency and size of price changes alone, and instead, I categorize products by the type of pricing policy they employ. Based on the finding that policies typically consist of a small set of prices, I characterize policy types in terms of the rigidity in the set of prices observed over the life of realized policies within a series. All products can be grouped into three categories: products characterized by *single-price policies* (SPP); products characterized by *one-to-flex policies* (OFP), in which a single sticky price is accompanied by transitory price changes to and from it, and in which none of the transitory price levels are revisited over the life of the policy;[8] and products with *coarse multiple-price policies* (MPP), in which at least two prices are revisited over the life of a policy. Figure 3 shows the share of products that fall under each policy type, illustrating the predominance of MPP products across product groups.

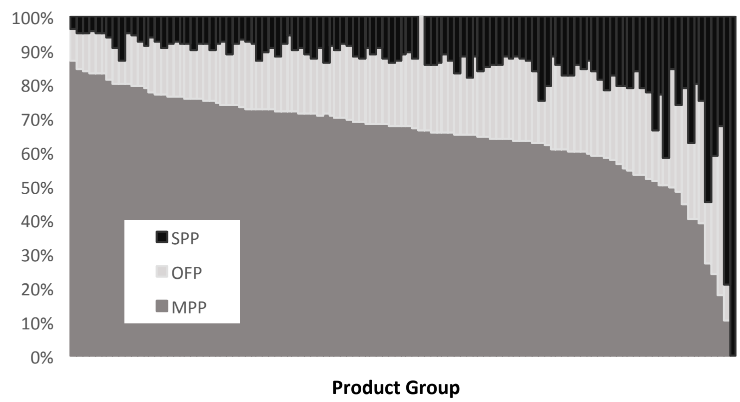


Figure 3: Policy types across product groups.

Data: AC Nielsen Retail Scanner Data. Breakdown of series by policy type (coarse multiple-price, one-to-flex and single-price) in each product group.

In allocating products to different categories, I assume that the determinants of a firm's choice of whether to pursue a single-price, one-to-flex, or coarse multiple-price plan do not change over time. Hence, all product series that have at least one coarse MPP realization are labeled as pursuing a coarse MPP strategy. All product series that have no such realizations, but have at least one OFP policy realization are assigned to the one-to-flex category. Finally, all products that consist entirely of single-price policies are counted in the SPP category. This categorization exhausts all product series: no series are characterized by purely flexible

[8]A price level is revisited if the price returns to that level before a break occurs in the series.

policies.[9] Table I presents statistics overall and by policy type.

Table I: Statistics by Pricing Policy

	Single-price	One-to-flex	Multi-price	All
Fraction of series (%)	9.6	19.4	71.1	100
Policy duration (median, weeks)	57.3	28.3	26.7	27.8
Policy cardinality (median)	1	2	4	3
Policy shift (median, %)	8.2	10.5	11.7	11.1
Frequency of price changes within (%)	0.0	8.3	33.3	25.0
Size of price changes within (%)	4.8	6.8	10.0	9.7

Data: AC Nielsen Retail Scanner Data. All statistics are expenditure-weighted. *Policy shift* is the change in absolute value in the weighted average price across policy realizations of a series. *Size of price changes within* is non-zero for single-price policies because the category includes series in which policies exhibit a single deviation from the modal price.

Single-Price Policies The workhorse time-dependent or state-dependent models of rigid price setting generate *single-price* policies. In the data, SPP series represent 9.6% of all product series.[10] These products adjust much less frequently and by less when they do adjust: the median policy duration is 57 weeks versus 28 weeks for all products, and the median policy shift is 8% versus 11% for all products. Hence, these products appear to face a relatively low volatility of their desired price that does not warrant the design and implementation of complex pricing policies.

One-to-Flex Policies Motivated by prior empirical studies that highlight the importance of transitory price changes, recent theoretical work has developed models in which firms can flexibly deviate from a rigid regular price, thereby generating a *one-to-flex* pattern. For example, Kehoe & Midrigan (2010) allow menu cost firms to "rent" a one-period price change for free, while Guimaraes & Sheedy (2011) allow firms to update the sales price flexibly while keeping a Calvo restriction on the regular price. In the data, OFP series

[9]A policy is defined as flexible if no price levels are repeated over the life of a policy realization.

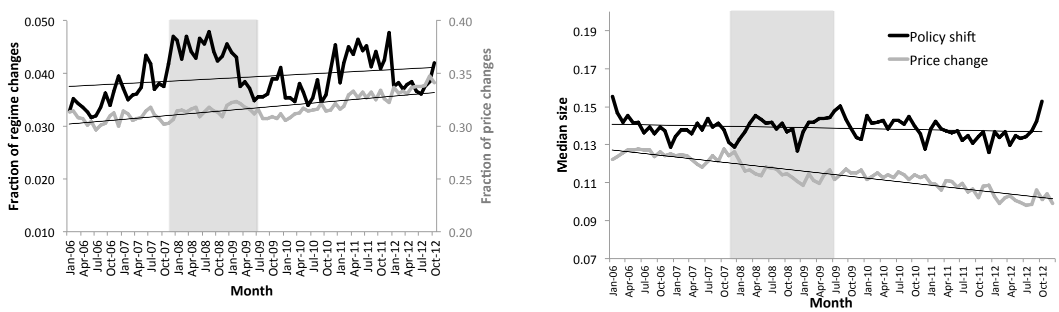
[10]This category allows for policies to exhibit a single deviation from the modal price. A single deviation over the life of a policy realization suggests that transitory price changes are not a meaningful aspect of the firm's pricing policy. Series characterized by purely single-price policies, with no deviations at all, represent only 2% of the data.

represent 19.4% of all product series. The statistics for these products suggest that they face a higher volatility in their desired price, compared with the SPP products. In particular, the median policy duration is much shorter, at 28 weeks, and the median shift in average prices across policy realizations is more than two percentage points higher, at 10.5%. However, the policies themselves are not very volatile or complex. First, 56% of the realized policies within OFP series are in fact single-price policies. Second, the median frequency with which prices adjust inside policies is only 8.3% (versus 25% for all products) and the median size of within-policy price changes is only 6.8%, three percentage points lower than the median for all products. The muted within-policy volatility suggests that the OFP products face a relatively high volatility in their desired price, but also a relatively high cost of implementing complex pricing policies.

Coarse Multi-Price Policies Underscoring the presence of rigidity beyond the modal price within each price plan, 71.1% of series contain *coarse* multiple-price policies (MPP). The volatility of the data is concentrated in these series. The median policy duration for these products is 27 weeks, one week shorter than that of the one-to-flex products, and the median shift in prices across policy realizations is higher, at 11.7%. In contrast to the policies of OFP products, the policies of MPP products are highly volatile: the median frequency of within-policy price changes is 33.3%, four times that of the OFP series, and the absolute size of within-policy price changes is 10.0%, more than three percentage points higher than that of the OFP series. Despite this volatility, these policies exhibit considerable discreteness in price levels: only four distinct prices are typically charged over the life of a policy realization. These statistics suggest that these products face highly volatile market conditions, and they adjust in two ways: first, they choose more complex – though nevertheless coarse – pricing policies, which consist of a small menu of prices; and second, they update their policies relatively more frequently, and upon adjustment, they shift by larger amounts.

The prevalence in the data of coarse multiple-price policies presents a challenge for existing models of price setting, and is instead consistent with the hypothesis of Eichenbaum et al. (2011), who propose that firms choose from a small set of prices that is updated relatively infrequently. The theory developed in Section 3 uses costly information to generate such plans endogenously, and yields the kind of heterogeneity in pricing policies that is observed in the data, as a function of heterogeneity in the volatility of shocks relative to the the costs of acquiring information.

Policy Adjustment During the Great Recession An advantage of the new AC Nielsen data is that it covers the Great Recession and its aftermath, enabling me to document how the patterns in pricing policies changed over that period. I find that making the distinction between policy changes and raw price changes is essential for uncovering these patterns. Figure 4 shows the dynamics of price and policy adjustment over time: the left panel shows the time series for the fraction of policy changes (left axis) and of price changes (right axis), and the right panel shows the time series for the median size of policy shifts and of price changes. The size of a policy shift is obtained by computing the average weighted price within a policy, and taking the absolute value of the change in this average price. All series are seasonally adjusted weekly statistics, averaged to monthly values.



(a) (b)

Figure 4: Adjustment during the Great Recession.

Data: AC Nielsen Retail Scanner Data. Panel a shows the time series for the fraction of policy changes (left axis) and of price changes (right axis). These are seasonally adjusted weekly fractions, averaged to monthly values. Panel b shows the time series for the median size of policy shifts and of price changes. The size of a policy shift is obtained by computing the change in the average weighted price within a policy. The shading marks the Great Recession. Linear trend lines are in black.

A striking feature of the data is that the rate of policy adjustments was particularly elevated during the Great Recession (rising from an average of 3.6% prior to the start of the Great Recession to a high of 4.8% at the peak of the crisis in September 2008). One potential explanation for this pattern is that the Great Recession was a period of heightened volatility, which led firms to increase the frequency with which they reviewed their policies. This interpretation is bolstered by the increase in the rate of policy adjustments in 2011, which was another period of increased uncertainty due to the Euro zone crisis and rising and highly volatile oil prices: the rate of policy changes rose once again, and stayed elevated throughout 2011, before declining sharply in early 2012.[11]

In contrast, the rate of price changes showed no cyclical pattern. Likewise, the size of

[11] This evidence is consistent with that of Anderson, Malin, Nakamura, Steinsson & Simester (2015), who find that an increase in oil prices in the 2007-2009 period had a significant effect on the frequency of *regular* prices posted by a particular retailer. A similar force may be driving the increase rate of policy changes in the Nielsen data in 2011.

adjustment, for both policies and prices, also showed no cyclical change during the Great Recession or during 2011. These findings suggest that firms did not appear to make their pricing policies more complex, which would generate higher rates of price adjustment and would also affect the size of policy and price adjustments. Rather, they made simple plans and they kept reviewing these plans often, until uncertainty was reduced. In terms of more medium-run trends, the data show a modest linear decline in the size of adjustment over the entire sample period, in tandem with the slight increase in the rate of adjustment observed over the same period. However, the sample is too short to determine if these trends reflect a structural, more permanent change in pricing behavior that is likely to continue, or if they only reflect a transitory phenomenon. A more permanent, structural change would suggest an increase in the complexity of pricing policies implemented by firms. The information-based theory presented in the next section predicts such an increase in complexity (and the associated increases in the rates of adjustment and decreases in the size of adjustment) as a result of technological progress that reduces the costs of acquiring information in order to make pricing decisions.

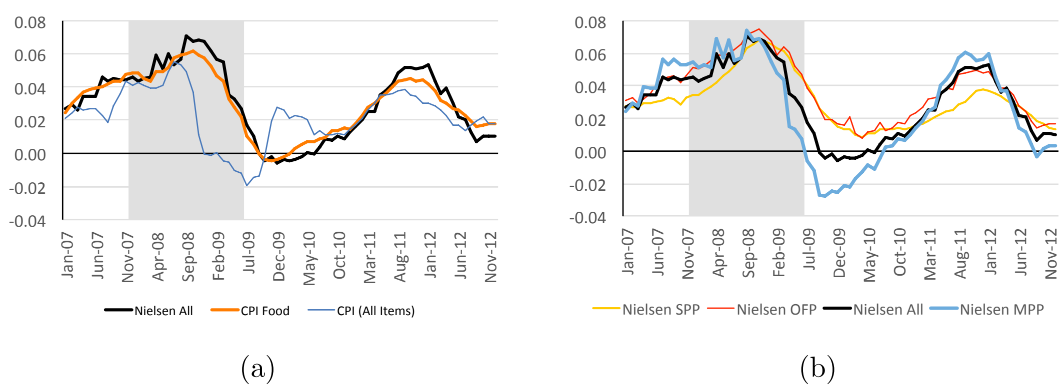


Figure 5: Inflation in AC Nielsen and the CPI.

Data: BLS and AC Nielsen Retail Scanner Data. Annual inflation rates in the AC Nielsen data set versus the CPI (panel a) and in the AC Nielsen data set by policy type: single-price, one-to-flex and coarse multiple-price (panel b).

Inflation Dynamics During the Great Recession A second interesting pattern pertains to the dynamics of inflation during and in the aftermath of the Great Recession. Figure 5 shows annual inflation series for different data samples: panel a compares the AC Nielsen inflation with that of the CPI, and panel b shows the AC Nielsen inflation for the three different types of products: single-price, one-to-flex and multi-price. Overall, the series diverge the most at the peak of the crisis, starting in the fourth quarter of 2008. First, the AC Nielsen inflation rate tracks the CPI inflation rate for Food and Beverages very well over the entire sample period.[12] Second, the data show a clear differential response of the different product types to the aggregate shock, as quantified in Table II: while the inflation rates for all types of products moved largely in tandem leading up to the fourth quarter of 2008, once inflation started to fall, it fell much faster for the MPP products than for the SPP and OFP products: from September 2008 to October 2009, inflation for MPP products had fallen by 10 percentage points, while inflation for SPP and OFP products had fallen by less than half that amount. Likewise, once inflation started falling again at the end of 2011, it fell by more than twice as much for MPP products than for SPP and OFP products. Hence, the MPP products, which likely face more volatile market conditions in general, responded more aggressively to the aggregate shocks. This finding suggests that the degree of state-dependence in policies differs significantly across products.

Table II: Changes in Inflation by Pricing Policy

	Single-price	One-to-flex	Multi-price	All
January 2007 - September 2008	3.6	4.0	5.0	4.4
September 2008 - October 2009	-4.1	-5.2	-10.2	-7.5
October 2009 - September 2011	0.8	2.8	8.9	5.6
September 2011 - September 2012	-1.2	-3.4	-6.5	-4.5

Data: AC Nielsen Retail Scanner Data. Entries show the peak changes in inflation (in percentage points) between the beginning and the end of four sub-periods, capturing the major swings over the sample period.

Moreover, SPP and OFP products continued to raise prices throughout, as inflation never fell below 1%, while MPP products actually cut prices, as their inflation rate reached a low of -2%. This evidence is consistent with that provided by Gilchrist et al. (2014), who find that at the peak of the crisis, firms operating in competitive markets lowered their prices significantly, relative to firms operating in less competitive markets. These heterogeneous responses across policy types can also account for the increased dispersion in price changes documented by Vavra (2014), as some types of products are more responsive than others.

[12] AC Nielsen diverges from the overall CPI starting in the fall of 2008: the overall CPI exhibits a sharp drop, concentrated in the fourth quarter of 2008, a sharp rebound in the third quarter of 2009, and then more modest fluctuations until the end of 2012. On the other hand, the AC Nielsen inflation rate falls a quarter later, and it shows no sharp uptick, only a gradual rebound starting at the end of 2009. Starting in 2011, the Nielsen inflation rate once again begins to track the overall CPI inflation quite well, and it exhibits the decline that has become the latest puzzle in the literature. Beraja et al. (2014) also show that the AC Nielsen price index tracks the Food CPI.

The information-based theory presented in the next section predicts precisely these effects: firms that operate in more volatile or more competitive markets have an incentive to acquire more information about market conditions, and hence they will choose more complex pricing policies and respond to shocks more aggressively.

These findings also underscore the importance of studying price data in its entirety, rather than eliminating transitory price volatility: transitory volatility is in fact crucial to pinning down the type of pricing policy employed by different firms and, as we have seen, the type of policy chosen by firms in turn affects how these firms respond to shocks, and hence it affects aggregate inflation dynamics.[13]

3 Theory

The empirical evidence supports a theory of price setting that generates coarse, infrequently updated price plans. In this section, I develop a theory of endogenous information acquisition that can generate such price plans, and that further predicts heterogeneity in the complexity of price plans chosen by different firms.

3.1 Setup

I study the price setting problem of an information-constrained firm who sets prices in a stochastic environment. Obtaining any information about the state of the world is costly. The firm's management chooses a policy that specifies (*i*) how information is acquired and used to set prices and (*ii*) since the policy itself can be reviewed, how information is acquired and used to decide whether or not to undertake a policy review. If a review is warranted, the management team pays a fixed cost to learn the state and to redesign the policy. Between policy reviews, the firm monitors market conditions and uses this information to implement its chosen policy. The firm's pricing policy specifies a menu of prices, a rule for determining which price to charge in each period over the life of the policy, as a function of the information obtained in each period, and a rule for determining the information to be acquired for this purpose. The firm's review policy specifies a rule for determining in each period whether or not the policy has become obsolete, such that a review is warranted, as a function of the information obtained, and a rule for determining what information to acquire in order to

[13] In allocating products to different policy types, I assume that the choice of whether to pursue an SPP, OFP or MPP policy does not change over time. However, in a longer data sample, it would be interesting to see if firms in fact do change their policy type in response to large shocks. For example, it may be the case that products whose prices were set according to SPP or OFP rules prior to the Great Recession responded to the crisis by switching to an MPP policy. One implication of such endogenous shifts in policy choices would be that the differences that I document between the different policy types underestimate the true differences.

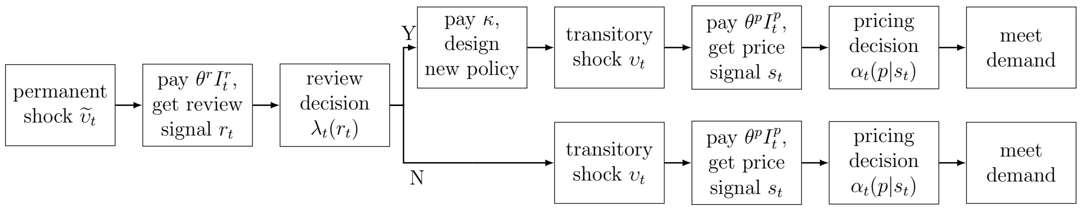


Figure 6: Model: Sequence of events in each period.

make this review decision.

The firm's per-period profit $\pi(p - x)$ is a function of the gap between its actual log price in the period p and its target log price x. The profit function is a smooth real-valued function with a unique global maximum at $p = x$. The target price is a linear combination of exogenous shocks, both transitory and permanent: $x_t = \widetilde{x}_t + \upsilon_t$, with $\widetilde{x}_t = \widetilde{x}_{t-1} + \widetilde{\upsilon}_t$, where $\widetilde{\upsilon}_t$ and υ_t, are drawn independently from some known distributions. In the frictionless benchmark, the firm observes the realized shocks perfectly and sets $p_t = x_t$ in each period.

The information-constrained firm maximizes its discounted profit stream net of monitoring and policy review costs,

$$\max_{\left\{I_t^r, I_t^p, \delta_t^r, p_t\right\}} E_0 \sum_{t=0}^{\infty} \beta^t \left[\pi(p_t - x_t) - \theta^r I_t^r - \theta^p I_t^p - \kappa \delta_t^r\right], \tag{1}$$

where $\beta \in (0, 1)$ is the discount factor, $I_t^r \geq 0$ is the quantity of information acquired in period t in order to make the review decision, $\theta^r > 0$ is the cost per unit of information for this decision, $I_t^p \geq 0$ is the quantity of information acquired in period t in order to make the pricing decision, $\theta^p > 0$ is the cost per unit of information for this decision, δ_t^r is equal to 1 if management reviews the policy in period t and 0 otherwise, and $\kappa > 0$ is the fixed cost associated with a policy review. The policy chosen at the time of each review specifies the rules that govern I_t^r and δ_t^r, to be applied starting in the next period, and I_t^p and p_t, which come in effect immediately, within the period. Information for each of the two decisions is acquired in the form of two endogenous signals: a review signal and a price signal. Figure 6 presents the sequence of events in each period.

The measurement of the information flows I_t^r and I_t^p follows the rational inattention literature (Sims (2003)), using Shannon's (1948) mutual information. Information flow is the reduction in entropy that results from observing an endogenously designed signal on the state of the economy. Choosing to acquire a larger quantity of information implies obtaining a more costly, but more precise signal. Hence, the firm faces a trade-off between closely tracking market conditions and economizing on information expenditure. The setup also allows the firm to choose to acquire no information for one or both decisions. In this case, decisions are based on the firm's prior, which is updated whenever there is a policy review.[14]

The two monitoring costs θ^r and θ^p are not necessarily equal, since the two decisions may be the responsibility of different managers, each with his or her own cost of processing information. For each manager, the unit cost determines the information processing capacity that the manager allocates to his or her problem. I assume that the quantity of information required for each problem is small relative to each manager's total capacity, such that each unit cost may be taken as fixed. Moreover, following Woodford (2009), the same unit cost applies to all types of information that may be relevant for each manager's problem. There is no free memory and there is no free transmission of information between the two managers.[15]

For simplicity, payment of the fixed cost κ enables the management team to receive *complete* information about the state at the time of the review, as in Reis (2006) and Woodford (2009). The assumption that this cost is fixed may be rationalized via economies of scale in the review technology. Hence the model nests both flow and lumpy acquisition of information.

3.2 The Firm's Problem

The Review Policy Let $\widetilde{\omega}_t$ denote the complete state at the time of the receipt of the review signal in period t. It includes the current realization of the permanent shock, $\widetilde{v}_t$, and the full history of shocks, signals, and decisions through period $t-1$. Suppose that the firm decides to review its policy. The new review policy is implemented starting in period $t+1$.

Definition 1. A *review policy*, implemented following a policy review in an arbitrary state $\widetilde{\omega}_t$ in period t, is defined by

1. $\mathcal{R}_t$, the set of possible review signals;

2. $\{\rho_{t+\tau}(r|\widetilde{\omega}_{t+\tau})\}_\tau$, the sequence of conditional probabilities for all $r \in \mathcal{R}_t$, all $\widetilde{\omega}_{t+\tau}$, and all $\tau > 0$ until the next review;

3. $\overline{\rho}_t(r)$, the unconditional frequency with which the decision-maker anticipates receiving each signal r, for all $r \in \mathcal{R}_t$, until the next review;

[14] Unlike Maćkowiak & Wiederholt (2009, 2010), and Matějka (2010), I employ a cost per unit of information acquired, rather than a fixed capacity to process information, which allows the firm to vary the quantity of information acquired, depending on its objective, market conditions, and the costs of obtaining information.

[15] Other dynamic inattention papers assume that the entire history of past signals is available for free in each period (e.g., Maćkowiak & Wiederholt (2009)). Conversely, as in Woodford (2009), I interpret the information friction as a processing friction that applies regardless of where the information is stored when not in use (externally, or in one's memory). Knowing the full history for free is not necessary in the current setup, given the firm's ability to occasionally review its policy.

4. $\lambda_t : \mathcal{R}_t \rightarrow [0,1]$, the decision rule for conducting reviews, with $\lambda_t(r)$ specifying the probability of conducting a review when the signal r is received, for all $r \in \mathcal{R}_t$.

The quantity of information expected, at the time of the review, to be acquired in the implementation of this review policy in each period until the next review is

$$J^r_{t+\tau} = E_t \left\{ I\left(\rho_{t+\tau}\left(r|\widetilde{\omega}_{t+\tau}\right), \overline{\rho}_t\left(r\right)\right)\right\}, \tag{2}$$

$$I\left(\rho, \overline{\rho}\right) \equiv \sum_{r \in \mathcal{R}_t} \rho\left(r|\widetilde{\omega}\right)\left[\log \rho\left(r|\widetilde{\omega}\right) - \log \overline{\rho}\left(r\right)\right], \tag{3}$$

where E_t denotes expectations conditional on the state $\widetilde{\omega}_t$, on a policy review having taken place in that state, and on the policy implemented at that time. This quantity is given by the average distance between the unconditional frequency of review signals over the life of the policy, $\overline{\rho}_t$, and each conditional distribution, $\rho_{t+\tau}$.

The Pricing Policy In each period, the price signal is received after the review decision has been made, and after the realization of the transitory shock, υ_t. For any $\tau \geq 0$, let $\omega_{t+\tau}$ denote the complete state at the time of the receipt of the price signal in period $t+\tau$. As above, suppose that the firm conducts a policy review in an arbitrary state $\widetilde{\omega}_t$. The new pricing policy applies starting in period t.

Definition 2. A *pricing policy*, implemented following a policy review in an arbitrary state $\widetilde{\omega}_t$ in period t, is defined by

1. $\mathcal{S}_t$, the set of possible price signals;

2. $\{\phi_{t+\tau}\left(s|\omega_{t+\tau}\right)\}_\tau$, the sequence of conditional probabilities of receiving the price signal s, for all $s \in \mathcal{S}_t$, all $\tau > 0$, and all $\omega_{t+\tau}$ until the next review;

3. $\overline{\phi}_t\left(s\right)$, the unconditional frequency with which the decision-maker anticipates receiving each price signal s, for all $s \in \mathcal{S}_t$, until the next review;

4. $\alpha_t : \mathcal{S}_t \times \mathbb{R} \rightarrow [0,1]$, the decision rule for price-setting, with $\alpha_t\left(p|s\right)$ specifying the probability of charging price $p \in \mathbb{R}$ when the price signal s is received, for all $s \in \mathcal{S}_t$.

The quantity of information expected to be acquired in the implementation of this pricing policy in each period until the next review is

$$J^p_{t+\tau} = E_t \left\{ I\left(\phi_{t+\tau}\left(s|\omega_{t+\tau}\right), \overline{\phi}_t\left(s\right)\right)\right\}, \tag{4}$$

$$I\left(\phi,\overline{\phi}\right)=\sum_{s\in\mathcal{S}_t}\phi\left(s|\omega\right)\left[\log\phi\left(s|\omega\right)-\log\overline{\phi}\left(s\right)\right], \tag{5}$$

where E_t denotes expectations conditional on the state $\widetilde{\omega}_t$, on a policy review having taken place in that state, and on the policy implemented at that time.

The first three elements in each of the two definitions can be thought of as the interface between the manager and her environment, while the fourth element maps the information received through this interface into the manager's actions.

These definitions are very general. The sets of possible signals $\mathcal{R}_t$ and $\mathcal{S}_t$ can include any variables that may be useful for the decisions at hand. It is important to note that nothing in the specification rules out continuous distributions. The sets $\mathcal{R}_t$ and $\mathcal{S}_t$ have been written as countable sets only for expository purposes, but it is only once we specify the objective function and the shock processes that the optimal signals will endogenously turn out to be continuous or discrete. Likewise, the sequences of conditional probabilities, $\{\rho_{t+\tau}\}_\tau$ and $\{\phi_{t+\tau}(s|\omega_{t+\tau})\}_\tau$ can be related in an arbitrary way to the state, and these relationships can vary with each future period until the next review. The only assumption is that all information, including knowledge of the passage of time or past events, is subject to the same unit cost of information. As a result, the two signal structures must each be defined relative to a single frequency ($\overline{\rho}_t$ and $\overline{\phi}_t(s)$), and each decision-maker must apply a single decision rule (λ_t and α_t), both chosen at the time of the review.[16]

The Cheapest Signal Structure The amount of information that is *used* by the decision-maker quantifies the reduction in uncertainty that is reflected in the agent's final decision (for example, review or do not review). Let $\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})$ denote the probability with which the decision-maker anticipates undertaking a policy review in state $\widetilde{\omega}_{t+\tau}$ in period $t+\tau$, and let $\overline{\Lambda}_t$ denote the unconditional probability of a review across all states, under the current policy,

$$\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)\equiv\sum_{r\in\mathcal{R}}\lambda_t\left(r\right)\rho_{t+\tau}\left(r|\widetilde{\omega}_{t+\tau}\right), \tag{6}$$

$$\overline{\Lambda}_t\equiv\sum_{r\in\mathcal{R}}\lambda_t\left(r\right)\overline{\rho}_t\left(r\right). \tag{7}$$

[16] Suppose that between reviews, the decision-maker had free access to either the entire history of past signals or the number of periods that have elapsed since the last review. In that case, the firm's policy would specify separate frequencies and decision rules for each history of prior signals, or for each period between reviews. Such a specification would complicate the model but, more importantly, it would take the model farther away from the empirical evidence, which underscores simplicity in the pricing policies chosen by firms, which most often consist of no more than three or four distinct price points.

Similarly, let $f_{t+\tau}(p|\omega_{t+\tau})$ denote the probability that the firm charges price p in state $\omega_{t+\tau}$ in period $t+\tau$, and let $\overline{f_t}(p)$ denote the unconditional probability that price p is charged over the life of the policy,

$$f_{t+\tau}(p|\omega_{t+\tau}) \equiv \sum_{s\in\mathcal{S}} \alpha_t(p|s)\,\phi_{t+\tau}(s|\omega_{t+\tau}), \tag{8}$$

$$\overline{f_t}(p) \equiv \sum_{s\in\mathcal{S}} \alpha_t(p|s)\,\overline{\phi}_t(s). \tag{9}$$

Lemma 1. *The most efficient policy, implemented following a policy review in an arbitrary state $\widetilde{\omega}_t$ in period t, defines $\{0,1\}$ as the set of possible review signals r, and specifies*

1. $\{\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})\}_\tau$*, the sequence of conditional probabilities of receiving $r=1$ (conduct a review) in state $\widetilde{\omega}_{t+\tau}$, period $t+\tau$;*
2. $\overline{\Lambda}_t$*, the anticipated unconditional frequency of reviews;*
3. $\mathcal{P}_t$*, the set of prices charged until the next review;*
4. $\{f_{t+\tau}(p|\omega_{t+\tau})\}_\tau$*, the sequence of conditional probabilities of charging price p for all $p\in\mathcal{P}_t$, all $\tau>0$ and all $\omega_{t+\tau}$ until the next review;*
5. $\overline{f}_t(p)$*, the anticipated unconditional frequency of prices, for all $p\in\mathcal{P}_t$.*

At the time of the review, the quantities of information expected to be acquired in the implementation of this policy in each period until the next review are

$$I^r_{t+\tau} = E_t\left\{I\left(\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau}), \overline{\Lambda}_t\right)\right\}, \ \forall\tau>0, \tag{10}$$

$$I^p_{t+\tau} = E_t\left\{I\left(f_{t+\tau}(p|\omega_{t+\tau}), \overline{f_t}(p)\right)\right\}, \ \forall\tau\geq 0. \tag{11}$$

Proof. Both the review decisions and prices are distributed independently of the state, conditional on the review and price signals. By the data-processing inequality (Cover & Thomas (2006)), the relative entropy between decisions and states is weakly less than the relative entropy between signals and states. If decisions are random functions of the signals, then the inequality is strict. □

This result is not only intuitive, but it also formally defines the cheapest policy that the firm can employ in order to make its review and pricing decisions. It extends the results of Woodford (2008) to the case of pricing policies that consist of more than a single price. The quantity $I^r_{t+\tau}$ defined in equation (10) is the smallest quantity of information that the review manager can acquire and still make exactly the same review decisions as when acquiring $J^r_{t+\tau}$, defined in equation (2). Likewise, the quantity $I^p_{t+\tau}$ defined in equation (11) is the smallest quantity of information that the pricing manager can acquire and still make exactly the same decisions as when acquiring $J^p_{t+\tau}$, defined in equation (4). For instance, it would not be optimal for the policy to differentiate between states in which the decision-maker takes the same action, since by merging such signals, information costs would be reduced with no loss in the accuracy of the decision. Moreover, it would also not be efficient to randomize the decision upon receipt of the signal, since it would be cheaper to reduce the mutual information between the signal and the state instead.

Reformulating the signalling mechanism in this way also leads to a simplification in solving the firm's problem: the firm's choices are reduced to five objects: $\overline{\Lambda}_t$, $\{\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})\}_\tau$, $\mathcal{P}_t$, $\overline{f}_t(p)$, and $\{f_{t+\tau}(p|\omega_{t+\tau})\}_\tau$. The first two objects define the firm's review policy, determining the frequency with which it undertakes reviews and the extent to which the timing of these reviews is tied to the state. The last three objects define the firm's pricing policy, determining the set of prices to charge between reviews and the degree to which the choice of which price to charge in what state is tied to the state.

If we eliminate the choice of a pricing policy, and instead restrict the firm to choose a single price to be charged between reviews, then the setup collapses to that of Woodford (2009), who studies the problem of a firm choosing when to update its price based on receipt of an endogenously chosen noisy signal. The information problem at the time of each review becomes choosing the sequence of conditional probabilities of a price change and the unconditional frequency of price changes. On the other hand, if we eliminate the review decision and assume that the firm obtains a signal based on which it sets its price in each period, the problem becomes a repeated static pricing problem similar to that solved by Maćkowiak & Wiederholt (2009) and Matějka (2010). The per-period information problem then becomes choosing the support for the price distribution and the conditional probability of charging each price in each state of the world. Hence, I model a dual decision problem that specifies rules for making both a review decision and a pricing decision in each period, and that determines the interdependence between the two decisions.

The Stationary Formulation The firm's problem can be written in terms of the innovations to the state since the last review. At the time of a policy review in period t, the firm learns the complete state, $\widetilde{\omega}_t$. First, let the *news states* $\widetilde{\varpi}_\tau$ and ϖ_τ denote the innovations in the complete states $\widetilde{\omega}_{t+\tau}$ and $\omega_{t+\tau}$ since the review in state $\widetilde{\omega}_t$. In particular, $\widetilde{\varpi}_\tau$ (which is relevant for the review decision) includes the history of permanent shocks between period $t+1$ and period $t+\tau$, the history of transitory shocks between period t and period $t+\tau-1$, and the history of prices between period t and period $t+\tau-1$. The news state ϖ_τ (relevant for the pricing decision) includes $\widetilde{\varpi}_\tau$ and the transitory shock in period $t+\tau$. Second, let $\widetilde{y}_\tau \equiv \widetilde{x}_{t+\tau} - \widetilde{x}_t$ denote the normalized pre-review target price, defined as the innovation in the pre-review target price since the last review, and let $y_\tau \equiv \widetilde{y}_\tau + \upsilon_t$ denote the normalized post-review target price, where υ_t is the transitory shock realization. Finally, let $q \equiv p - \widetilde{x}_t$ denote the normalized price. The normalized variables $\widetilde{y}_\tau$, y_τ, $\widetilde{\varpi}_\tau$, and ϖ_τ, are distributed independently of the state $\widetilde{\omega}_t$. Hence, the firm's problem can be expressed without any reference to either the date t or the state $\widetilde{\omega}_t$ in which the review takes place.

Problem. A firm undertaking a policy review in any state and period chooses $\overline{\Lambda}$, $\{\Lambda_\tau(\widetilde{\varpi}_\tau)\}_{\tau>0}$, Q, $\overline{f}(q)$, and $\{f_\tau(q|\varpi_\tau)\}_{\tau\geq 0}$ to solve

$$\overline{V} = \max E\left[\Pi_0(\varpi_0) + \sum_{\tau=1}^{\infty} \beta^\tau \Gamma_\tau(\widetilde{\varpi}_{\tau-1}) W_\tau(\varpi_\tau)\right], \tag{12}$$

where $\Pi_\tau(\varpi_\tau)$ is the per-period profit expected under the pricing policy in effect, prior to receiving the price signal for that period, and net of the cost of that signal,

$$\Pi_\tau(\varpi_\tau) \equiv \sum_{q\in Q} f_\tau(q|\varpi_\tau)\,\pi(q-y_\tau) - \theta^p I\left(f_\tau(q|\varpi_\tau), \overline{f}(q)\right), \tag{13}$$

and $\Gamma_\tau(\widetilde{\varpi}_{\tau-1})$ denotes the probability, expected at the time of the review, that the review policy in effect continues to apply τ periods later, with $\Gamma_1(\cdot) \equiv 1$ and

$$\Gamma_\tau(\widetilde{\varpi}_{\tau-1}) \equiv \prod_{k=1}^{\tau-1}\left[1-\Lambda_k(\widetilde{\varpi}_k)\right], \forall \tau > 1. \tag{14}$$

The continuation value $W_\tau(\varpi_\tau)$ is given by

$$W_\tau(\varpi_\tau) \equiv (1-\Lambda_\tau(\widetilde{\varpi}_\tau))\,\Pi_\tau(\varpi_\tau) + \Lambda_\tau(\widetilde{\varpi}_\tau)\left(\overline{V}-\kappa\right) - \theta^r I\left(\Lambda_\tau(\widetilde{\varpi}_\tau), \overline{\Lambda}\right). \tag{15}$$

Conditional on the current policy surviving all the review decisions leading to a particular state $\widetilde{\varpi}_\tau$, the firm pays the cost of the review signal. It then continues to apply the current policy with probability $1-\Lambda_\tau(\widetilde{\varpi}_\tau)$, in which case it attains expected profits $\Pi_\tau(\varpi_\tau)$, and it undertakes a policy review with probability $\Lambda_\tau(\widetilde{\varpi}_\tau)$, in which case it pays the review cost κ and expects the maximum attainable value, $\overline{V}$.

3.3 The Optimal Policy

I obtain the solution to the firm's problem in steps, deriving each element of the optimal policy taking the other elements as given. The derivation is in the Appendix.

The first result is that the optimal policy conditions directly on the normalized targets $\widetilde{y}$ and y, rather than on the complete news states, $\widetilde{\varpi}$ and ϖ. The firm chooses to allocate no attention to learning about past actions, past signals, or the passage of time. This outcome reflects the fact that all these types of information have equal cost per unit of information. Since the firm would like to have knowledge of past events or the passage of time only insofar as this knowledge is informative about the current normalized target, the firm chooses to learn directly about this target.

The second result is that the optimal policy specifies time-invariant functions for both the review policy and the pricing policy, even though I allow the firm to choose conditional distributions that are indexed by time. This outcome is a direct consequence of the first point discussed above. Since the firm chooses to learn directly about the current target, its signal problem for each decision is the same in every period, subject to the requirement that across periods, it must be consistent with the anticipated frequency with which each choice is expected to be made over the life of the policy.

The Optimal Review Policy. *Let the pricing policy be fixed. The optimal hazard function for policy reviews is given by*

$$\frac{\Lambda\left(\widetilde{y}\right)}{1-\Lambda\left(\widetilde{y}\right)}=\frac{\overline{\Lambda}}{1-\overline{\Lambda}}\exp\left\{\frac{1}{\theta^{r}}\left[\overline{V}-\kappa-V\left(\widetilde{y}\right)\right]\right\}, \tag{16}$$

where $V\left(\widetilde{y}\right)$ is the firm's continuation value under the current policy and $\overline{V}=V\left(0\right)$ is the firm's continuation value upon conducting a policy review. The optimal anticipated frequency of policy reviews is given by

$$\overline{\Lambda}=\frac{E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau-1}\right)\Lambda\left(\widetilde{y}_{\tau}\right)\right\}}{E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau-1}\right)\right\}}, \tag{17}$$

where $\Gamma\left(\widetilde{y}^{\tau-1}\right)$ is the probability that the policy in effect continues to apply τ periods later, as a function of the history of the pre-review normalized target prices, $\widetilde{y}^{\tau-1}$, with $\Gamma\left(0\right)\equiv 1$, and $\Gamma\left(\widetilde{y}^{\tau-1}\right)\equiv\prod_{k=1}^{\tau-1}\left[1-\Lambda\left(\widetilde{y}_{k}\right)\right]$ for $\tau>1$.

First, in determining whether or not to undertake a review, the firm considers the gain from undertaking a review, $\overline{V}-V\left(\widetilde{y}\right)$, relative to the cost of the review, κ, but it does so imperfectly. In order to economize on information costs, the optimal review signal neither rules out a review nor indicates a review with certainty. For low values of the unit cost θ^r, the firm can afford to acquire more information in order to make its review decision, and hence this decision becomes increasingly precise. In the limit, as $\theta^r \to 0$, the review policy approaches a fully state-dependent review policy, as in Burstein (2006).[17] At the other extreme, as $\theta^r \to \infty$, $\Lambda(\widetilde{y}) \to \overline{\Lambda}$ for all $\widetilde{y}$, generating Calvo-like policy reviews.

Although omitted in order to simplify notation, the review hazard function depends not only on the current normalized target price $\widetilde{y}$, but also on the firm's pricing policy, which determines the per-period profit expected under the current policy. If we restrict the firm to choose a single price between reviews, then the review hazard function becomes a function of the gap between the firm's current log price and its normalized target price. The hazard function then becomes of the same form as that derived by Woodford (2009) for *price* reviews in a model in which the firm chooses, based on imperfect signals, when to update its price.

Second, for a given hazard function, the frequency of reviews is chosen to minimize the expected cost of the review signal over the expected life of the policy. The cost of the review signal in future periods is more heavily discounted, and this discounting is reflected in the expression for $\overline{\Lambda}$ in equation (17).

Furthermore, the hazard function for policy reviews together with the evolution of exogenous shocks determine the distribution of states that the firm expects to encounter over the life of the policy. Let $\widetilde{g}_\tau$ denote the distribution of pre-review target prices in period $\tau \geq 1$, with $\widetilde{g}_1(\widetilde{y}) = h_{\widetilde{\nu}}(\widetilde{y})$ and

$$\widetilde{g}_\tau(\widetilde{y}_\tau) = \int [1 - \Lambda(\widetilde{y}_{\tau-1})]\, \widetilde{g}_{\tau-1}(\widetilde{y}_{\tau-1})\, h_{\widetilde{\nu}}(\widetilde{y}_\tau - \widetilde{y}_{\tau-1})\, d\widetilde{y}_{\tau-1}, \tag{18}$$

for $\tau > 1$, where $h_{\widetilde{\nu}}$ is the distribution of the permanent innovation. If we define $\widetilde{G}$ as the *discounted* distribution of states over the life of the policy,

$$\widetilde{G}(\widetilde{y}) = \frac{\sum_{\tau=1}^{\infty} \beta^\tau \widetilde{g}_\tau(\widetilde{y})}{\int \sum_{\tau=1}^{\infty} \beta^\tau \widetilde{g}_\tau(z)\, dz}, \tag{19}$$

then we can express the anticipated frequency of reviews more compactly, as

$$\overline{\Lambda} = \int \Lambda(\widetilde{y})\, \widetilde{G}(\widetilde{y})\, d\widetilde{y}. \tag{20}$$

[17] Burstein (2006) considers a full-information model in which the firm faces a fixed cost of changing its pricing policy; the policy then specifies the entire sequence of time-varying future prices, which are however chosen based on the information available at the time of the review, and cannot be made contingent on future states. Hence, unlike the present paper, he considers a model in which the firm obtains no information between reviews.

The Optimal Pricing Policy. *Let the review policy be fixed. For a given support Q, the optimal conditional distribution of prices is given by*

$$f(q|y) = \overline{f}(q) \frac{\exp\left\{\frac{\pi(q-y)}{\theta^p}\right\}}{\sum_{\widehat{q}\in Q} \overline{f}(\widehat{q}) \exp\left\{\frac{\pi(\widehat{q}-y)}{\theta^p}\right\}}, \tag{21}$$

and the unconditional distribution of prices is given by

$$\overline{f}(q) = \frac{E\left\{\sum_{\tau=0}^{\infty} \beta^{\tau} \Gamma(\widetilde{y}^{\tau}) f(q|y)\right\}}{E\left\{\sum_{\tau=0}^{\infty} \beta^{\tau} \Gamma(\widetilde{y}^{\tau})\right\}}. \tag{22}$$

Moreover, these distributions specify the unique optimal pricing policy among all pricing policies with support Q.

For a given set of prices in the support of the pricing policy, the probability of setting a particular price in a particular state is high, relative to the overall probability of charging that price across all states, when the value of doing so is high relative to the average value that the firm can expect in this particular state across all the prices in the support. However, the relationship between the state and the price is noisy: the pricing policy places positive mass on all prices in the support, for each target price y. This noise reflects the desire to economize on the information cost associated with receiving the price signal in each period.

The anticipated frequency of prices is chosen to minimize the total cost of the price signal over the expected life of the policy. The optimal frequency is equal to the (discounted) weighted average of the conditional price distribution over all post-review states that the firm expects to encounter until the next review, given the firm's review policy, which determines the probability of surviving to a particular state. In particular, let g_τ denote the distribution of post-review target prices in period τ, with $g_0(y) = h_\nu(y)$ and

$$g_\tau(y) = \int [1 - \Lambda(y-\nu)]\, \widetilde{g}_\tau(y-\nu)\, h_\nu(\nu)\, d\nu, \tag{23}$$

$\forall \tau > 0$, for all y, where h_ν is the distribution of the transitory innovation, ν. If we define

$$G(y) = \frac{\sum_{\tau=0}^{\infty} \beta^{\tau} g_\tau(y)}{\int \sum_{\tau=0}^{\infty} \beta^{\tau} g_\tau(z)\, dz}, \tag{24}$$

then the optimal frequency with which the decision-maker anticipates charging each price over the life of the policy is the marginal distribution corresponding to f,

$$\overline{f}(q) = \int f(q|y)\, G(y)\, dy. \tag{25}$$

Static Transformation Rather than designing a separate signalling mechanism to accommodate the distribution of relevant states in each period, $\widetilde{g}_{\tau}$ and g_{τ}, the firm designs a *single* signalling mechanism that can accommodate all possible distributions until the next review, reflecting the fact that it has no knowledge of which distribution is "active" at any point in time, with distributions further into the future discounted relatively more.

The part of the objective that depends on the firm's pricing policy can now be written directly in terms of the discounted distribution of normalized target prices as

$$\int G(y)\,\Pi(y)\,dy, \tag{26}$$

where $\Pi(y)$ is the expected profit under the current pricing policy, net of the cost of the pricing policy, when the target price is y,

$$\Pi(y) = \sum_{q\in Q} f(q|y)\,\pi(q-y) - \theta^{p} I\left(f(q|y), \overline{f}(q)\right). \tag{27}$$

Through this formulation, the dynamic pricing problem has been transformed into a *static* rational inattention problem for a distribution of states given by G and an objective function given by π. The pricing objective specified in equation (26) is strictly concave in both f and $\overline{f}$. Therefore, equations (21) and (25), which characterize f and $\overline{f}$ for a given support, describe the optimal policy on a fixed support, Q, and have the same form as the equations that characterize the solution to the static rate distortion problem for a memoryless source (Shannon (1959)).

The Optimal Pricing Support. *Let the distribution of states, G, be fixed, and let the probability distributions f and $\overline{f}$ satisfy (21) and (25) for all $q \in Q$. Let*

$$Z\left(q;\overline{f}\right) \equiv \int G(y)\,\frac{\exp\left[\frac{\pi(q-y)}{\theta^{p}}\right]}{\sum_{\widehat{q}\in Q}\overline{f}(\widehat{q})\exp\left[\frac{\pi(\widehat{q}-y)}{\theta^{p}}\right]}dy. \tag{28}$$

Then, the set Q is the optimal support of the pricing policy if and only if

$$Z\left(q;\overline{f}\right) \quad \begin{cases} =1 & \text{if} \quad q\in Q, \\ \leq 1 & \text{if} \quad q\notin Q. \end{cases} \tag{29}$$

The associated probability distribution satisfies the fixed point $\overline{f}(q) = \overline{f}(q)\,Z\left(q;\overline{f}\right), \forall q \in Q$.

The value $Z\left(q;\overline{f}\right)$ represents the value of charging the price q relative to the value of charging other prices $\widehat{q} \in Q$, on average, across all possible states y. The optimal signalling mechanism equates this value across all prices in the support. Moreover, it requires that charging any other price would yield a weakly lower average value. If one can find a set of prices Q that satisfy the conditions in (29), then this set characterizes the uniquely optimal solution at the information cost θ^p.

Following Rose (1994), one can establish a pair of useful necessary conditions. Let $f(q|y)$ and $\overline{f}(q)$ satisfy the optimality conditions in equations (21) and (25). The points of support must satisfy

$$\int G(y|q) \frac{\partial \pi(q-y)}{\partial q} dy = 0, \tag{30}$$

$$\int G(y|q) \left[\frac{\partial^2 \pi(q-y)}{\partial q^2} + \frac{1}{\theta^p} \left(\frac{\partial \pi(q-y)}{\partial q} \right)^2 \right] dy \leq 0, \tag{31}$$

for all $q \in \mathcal{Q}$.

The interpretation of the first condition is that the price signal q received in any period must maximize the expected single-period profit under the conditional distribution for y implied by the signal that is received. I use this condition to determine the values of q for a given pair of distributions, $f(q|y)$ and $\overline{f}(q)$, and a given cardinality of the set $\mathcal{Q}$.

The second order condition implies that if a set $\mathcal{Q}$ of a given cardinality is such that prices in this set satisfy the optimality conditions for $f(q|y)$, $\overline{f}(q)$, and q, but do not satisfy equation (31), then the size of the set $\mathcal{Q}$ must be increased. Hence, equation (31) provides a way to verify if, for a given information cost, the solution must necessarily involve more than one price, as discussed further below.

Threshold Information Cost I establish a bound on the unit cost of the price signal such that, for any cost below this bound, the optimal policy necessarily involves more than one price. A single-price policy, if optimal, is defined by the price

$$\overline{q} = \arg\max_q \int G(y) \pi(q-y) dy. \tag{32}$$

The threshold cost of the price signal that is sufficiently low such that the single-price policy is not optimal is given by

$$\overline{\theta}^p \equiv \frac{\int G(y) \left(\frac{\partial}{\partial q} \pi(q-y) \right)^2 dy}{\int G(y) \left(\frac{\partial^2}{\partial q^2} \pi(q-y) \right) dy}, \tag{33}$$

where the derivatives are evaluated at $\overline{q}$.[18]

Solution Method I use equations (21), (25) and (29) numerically to find the optimal support. The numerical algorithm builds on algorithms from the information theory literature, namely Arimoto (1972), Blahut (1972), Csiszár (1974), and Rose (1994). The algorithm is detailed in the Appendix.

4 Micro Results

This section explores the implications for price adjustment of the information structure developed thus far, in a standard model of price-setting under monopolistic competition. Generating pricing patterns consistent with the data requires moderate expenditure on information.

4.1 Model of Price Setting

I consider the problem of monopolistically competitive firms that set prices subject to uncertainty in demand and productivity. I assume that all aggregate variables evolve according to the full-information, flexible price equilibrium, and focus on the price adjustment of a set of information-constrained firms of measure zero. The Appendix maps a standard monopolistically competitive economy into this setup. The profit function is

$$\pi(q-y) = e^{(1-\varepsilon)(q-y)} - \frac{\varepsilon - 1}{\varepsilon\gamma\,(1+\nu)} e^{-\varepsilon\gamma(1+\nu)(q-y)}, \tag{34}$$

where $\varepsilon > 1$ is the elasticity of substitution among Dixit-Stiglitz varieties, $\gamma \geq 1$ captures decreasing returns to scale in production and $\nu \geq 0$ is the inverse of the Frisch elasticity of labor supply. The profit function is concave, with a unique maximum at $q = y$.

The target price is a linear combination of all the shocks in the economy: permanent monetary shocks, permanent idiosyncratic quality shocks, which affect both the demand for an individual product and the cost of producing it, and i.i.d. idiosyncratic quality shocks. The log of money supply follows a random walk process, $m_t = m_{t-1} + \mu_t$, where $\mu_t \sim \mathcal{N}\left(\overline{\mu}, \sigma_\mu^2\right)$ is independent over time and from any other disturbances. The idiosyncratic permanent quality shock also follows a random walk, $z_t\left(i\right) = z_{t-1}\left(i\right) + \xi_t\left(i\right)$, where $\xi_t\left(i\right) \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$, independent over time and from the other shocks. The idiosyncratic i.i.d. shock is $\zeta_t\left(i\right) \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$.

[18] Note that the threshold is not always finite. In particular, in a model with a quadratic objective and a Gaussian distribution G, the solution "breaks" to a continuous support on the entire real line for any $\theta^p < \infty$.

The law of motion for the normalized pre-review state $\tau > 0$ periods after a review is

$$\widetilde{y}_{\tau}(i) = \widetilde{y}_{\tau-1}(i) + \mu_{\tau} + \xi_{\tau}(i). \tag{35}$$

This law of motion is embedded in $\widetilde{G}(\widetilde{y})$, the discounted distribution of pre-review target prices that the firm expects to encounter over the life of the policy, determined in Section 3. The law of motion for the normalized target price that enters the firm's period profit function is $y_0(i) = \zeta_0(i)$ and

$$y_{\tau}(i) = \widetilde{y}_{\tau}(i) + \zeta_{\tau}(i), \tag{36}$$

for $\tau > 0$. This law of motion is embedded in $G(y)$, the discounted distribution of target prices after the review decision, and after the realization of the transitory shock in each period, determined in Section 3.

4.2 Empirical Targets

I parameterize the model at the weekly frequency, targeting the duration, discreteness, and volatility of pricing policies for coarse multiple-price policy (MPP) products. Variation in parameters then yields heterogeneity in pricing policies, including SPP and OFP-like policies. Figure 7 shows a sample price series for a multiple-price policy firm, along with the target price that would be charged in the full information, flexible price benchmark. The shading marks the timing of policy reviews as identified by the break test. Consistent with the data, the theory generates large, transitory volatility among a small number of infrequently updated price levels. Overall, the firm's actual price tracks the target price well, especially in the medium-run, although in the short run the firm frequently makes mistakes, given the noise in both its review signal and its price signal.

Parameterization The parameters that determine the shape of the firm's profit function, shown in the top panel of Table III, are set to commonly used values in the literature. The elasticity of substitution is $\varepsilon = 5$. Variation in ε changes the asymmetry of the profit function, and hence the firm's incentives to acquire information. A higher elasticity implies larger losses from setting a price that is too low relative to the optimal full information price. The inverse of the exponent on the firm's production function is $\gamma = 1.5$ and the inverse of the Frisch elasticity of labor supply is $\nu = 0$. Variations in these two parameters change both the curvature and asymmetry of the profit function: higher values imply larger losses from charging a price that is different from the optimal full information price, especially in the case of prices that are too low relative to the optimum. Finally, the weekly discount factor is $\beta = 0.9994$, which implies an annual discount rate of 3%.

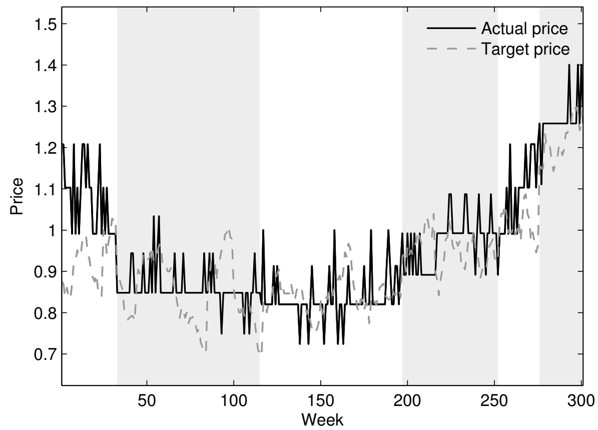


Figure 7: Simulated price series.

Simulation of actual and target price. Shading marks policy reviews identified by the break test.

The middle panel of Table III shows the parametrization of the shocks, and the bottom panel shows the parameterization of the information costs. For the money supply process, $\overline{\mu} = 0.0004$ and $\sigma_\mu = 0.0015$. These values imply an annualized inflation rate of 2%, with an annualized standard deviation of 1%. The standard deviation of the permanent idiosyncratic shock, $\sigma_\xi = 0.027$, is chosen jointly with the information costs $\kappa = 1.5$, $\theta^r = 5$ and $\theta^p = 0.09$, to target the frequency of policy reviews, the average shift in prices across reviews, the overall frequency of price changes, and the cardinality of the pricing policy. Introducing transitory shocks has limited quantitative effects on the firm's review policy. On the other hand, transitory shocks increase both the frequency and size of price changes, and, if large enough, they can also increase the cardinality of the firm's pricing policy. For simplicity, I exclude transitory shocks from the baseline results.

Table IV shows the model's ability to match statistics from the micro data. The second column presents statistics for MPP products from the data, and the third column presents statistics from the baseline parameterization of the model. Subsequent columns present results for alternative parameterizations, in which I vary the three information costs.

Multiple-Price Policies The baseline parameterization yields multiple-price policies that can match the duration, discreteness and volatility of pricing policies for MPP products in the Retail Scanner data. Specifically, in terms of the four targets, the model generates (*i*) a 3.2% frequency of policy reviews versus 3.1% in the data, (*ii*) a 31.3% frequency of price changes versus 33.3% in the data, (*iii*) a median number of distinct prices per policy realization of 4, as in the data, and (*iv*) a shift in average prices across policy realizations (computed as

Table III: Baseline Parameterization

Parameter	Symbol	Value	Explanation/Target
Elasticity of substitution	ε	5	Implied markup of 25%
Inverse production fn. exponent	γ	1.5	Decreasing returns to scale
Inverse Frisch elasticity	ν	0	Indivisible labor
Discount factor	β	0.9994	Annual discount rate of 3%
Mean of money supply shock	$\overline{\mu}$	0.0004	Annual inflation rate of 2.1%
Std. dev. of money supply shock	σ_μ	0.0015	Annual standard deviation of 1.1%
Std. dev. of permanent idio. shock	σ_ξ	0.027	Frequency of price changes
Fixed cost of a policy review	κ	1.5	Frequency of policy reviews
Cost of review signal	θ^r	5	Shift in mean prices across policies
Cost of price signal	θ^p	0.09	Cardinality of pricing policy

the median shift in absolute value in the weighted average price across consecutive policy realizations) of 10.2%, versus 11.7% in the data.

In terms of additional statistics, the model generates (*i*) an overall median absolute size of price changes of 13.1% versus 13.2% in the data, (*ii*) a frequency of the modal price of 71.6% versus 53.5%, and (*iii*) a frequency with which the maximum price per policy is the modal price of 12.9% versus 62.2%. In general, statistics that relate to the volatility of prices or policy realizations can be improved upon by varying parameters within the existing framework.

However, one set of statistics that are difficult to reconcile with the data have to do with the shape of the distribution of prices between policy reviews: in the data, the most frequently quoted price is often among the highest prices realized between breaks, and deviations are often price cuts from the mode, rather than price spikes. Conversely, the theory generates a much lower value for this statistic because the most frequently quoted price is often among the medium or low prices of the policy. As a result, the model over-estimates the frequency of price spikes relative to price cuts. This outcome reflects the shape of distribution of shocks that the firm is facing and the fact that the information-constrained firm fears under-pricing. The relevant distribution for the shape of pricing policy is $G(y)$, the discounted distribution of target prices. As the Gaussian shocks accumulate over time, $G(y)$ features the largest mass in the middle of the distribution. Generating a pricing policy with the mode at the high price would require giving the firm an additional specific incentive to undertake price cuts relative

Table IV: Quantitative Results

	Data	Base	High θ^p	High θ^r	High κ
Targets					
Frequency of policy reviews (%)	3.1	3.2	4.1	3.8	3.0
Frequency of price changes (%)	33.3	31.3	4.1	34.5	39.1
Cardinality of the pricing policy	4	4	1	6	5
Shift in prices across policies (%)	11.7	10.2	14.7	8.9	10.8
Other statistics					
Overall size of price changes (%)	13.2	13.1	14.7	14.6	12.7
Freq. of modal price (%)	53.5	71.6	100	72.3	61.1
Freq. max is mode (%)	62.2	12.9	100	13.1	10.2
Information expenditure (% of Full Info profits)					
On reviews	-	10.2	12.2	12.1	12.7
On review signal	-	1.7	3.8	0.6	2.0
On price signal	-	5.8	-	8.5	7.6
Total info expenditure	-	17.7	16.0	21.3	22.3
Profits, excluding info costs (% FI)	-	88.8	84.7	88.6	88.5

to price increases (for instance to temporarily gain market share), and by limiting the losses that the firm faces when potentially underpricing (for instance, by allowing stockouts).[19]

How well does the firm do with this complex pricing policy? First, the firm's expected profit, excluding information costs, is 88.8% of the benchmark full-information profit. Hence, overall, the firm's policy tracks market conditions fairly closely. Second, the firm spends approximately 18% of the full-information profits on the design and implementation of its policy.[20] The breakdown is as follows: 10.2% is spent on reviewing the policy; since the cost of the review signal is quite high at $\theta^r = 5$, the firm spends only 1.7% of the full information profits on monitoring market conditions to determine if a review is warranted; finally, the cost of the price signal is moderate, so the firm spends 5.8% on monitoring market conditions to determine which price to charge in each period. Hence, net of information costs, the

[19] Matějka (2010) generates a distribution of prices with slightly larger mass at the high price by assuming a uniform distribution for $G(y)$, in which case the asymmetry in the price distribution is determined solely by the asymmetry of the profit function.

[20] For comparability across parameterizations, I report information costs as a percent of the benchmark full information profit, rather than as a percent of the (varying) information-constrained expected profit.

information-constrained firm achieves 71% of the full information, flexible price profits.

4.3 Heterogeneity and Interdependence

Varying parameters that plausibly differ across firms and over time, such as the costs of acquiring information or the volatility of firm's target price, yields heterogeneity in the resulting pricing policies. To illustrate the interaction between the firm's pricing policy and its review policy, the last three columns of Table IV present results for variation in information costs relative to the baseline parameterization.

Cost of Pricing Policy First, I consider an increase in θ^p, the cost of monitoring market conditions to set the price in each period between reviews, to $\theta^p = 0.32$ from $\theta^p = 0.09$. This cost directly affects the pricing policy chosen by the firm to be implemented between reviews. If this cost is high enough, the firm will choose a coarser pricing policy, and at $\theta^p = 0.32$ the firm chooses a pricing policy consisting of a single price. The direct effect is that the frequency of price changes decreases and the size of price changes increases.

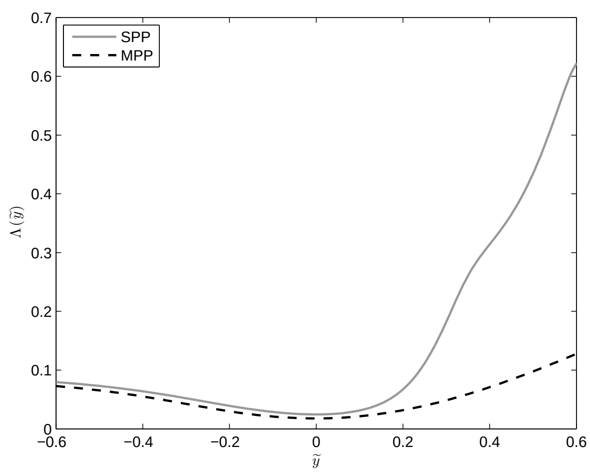


Figure 8: Hazard functions for policy reviews.

Hazard function for policy reviews for $\theta^p = 0.32$, which yields single-price policies (SPP), and for $\theta^p = 0.09$, which yields multiple-price policies (MPP).

Crucially, since the review policy and the pricing policy are chosen to be *jointly* optimal, the change in the pricing policy affects the review policy as well. Specifically, the firm now undertakes policy reviews more frequently, and acquires more precise information to determine whether or not to review its policy, more than doubling its expenditure on the review signal. Hence, the firm partially makes up for its more costly price signal by spending more resources on its review policy.

Figure 8 plots the hazard functions $\Lambda(\widetilde{y})$ implied by the high θ^p, single-price policy and by the low θ^p, multiple-price policy, as a function of the normalized pre-review state, $\widetilde{y}$. First, note that both review hazard functions are asymmetric, reflecting the asymmetry of the firm's profit function: since the firm's losses are larger when the permanent component of the firm's target price $\widetilde{y}$ is high relative to the firm's prices, the review hazard function is designed to trigger a review with higher probability in such states, thus "killing off" more quickly states that generate larger losses. This asymmetry implies that contractionary aggregate demand shocks will have larger effects than expansionary aggregate demand shocks. Second, in the MPP case, the possibility of adjusting prices between reviews, albeit imperfectly, enables the firm to undertake less frequent policy reviews and to spend less on acquiring information regarding the timing of reviews. As a result, the MPP hazard function is lower, flatter, and much less asymmetric than the SPP hazard function.

Cost of Review Policy Next, I consider an increase in the cost of monitoring market conditions to determine whether a review is warranted, to $\theta^r = 20$ from $\theta^r = 5$. The higher this cost is, the less information the firm's review signal contains about the evolution of market conditions. The firm's hazard function for policy reviews, $\Lambda(\widetilde{y})$, is flatter, and the frequency of reviews, $\overline{\Lambda}$, is higher. In turn, the flatter hazard function significantly affects the optimal pricing policy, increasing the threshold $\overline{\theta}^p$ below which multiple-price policies are optimal. In other words, the higher is the cost of undertaking policy reviews, the more complex a pricing policy the firm will choose between reviews. As shown in the fifth column of Table IV, the cardinality of the pricing policy increases to 6 from 3 prices. Since the review signal is so costly, the firm essentially spends no resources on designing an informative review signal, and instead partially compensates by implementing a more complex pricing policy and by undertaking policy reviews more frequently.

Cost of A Review Finally, I consider an increase in the fixed cost of policy reviews, to $\kappa = 2$ from $\kappa = 1.5$. The direct effect of this increase is that the hazard function for policy reviews shifts down, and the frequency of reviews declines. To compensate for the more costly reviews, the firm chooses to acquire a more precise review signal, such that the hazard function for reviews becomes slightly steeper. These changes in the review policy have implications for the firm's pricing policy, since they imply that the distribution of states relevant for the pricing decision has fatter tails, as shown in Figure 9: since the frequency of reviews has decreased, the distribution of potential states G is now more dispersed. As a result, the threshold $\overline{\theta}^p$ below which multiple-price policies are optimal increases and the firm designs more complex pricing policies that are characterized by both higher cardinality and higher precision of the signal that dictates which price to charge in each period.

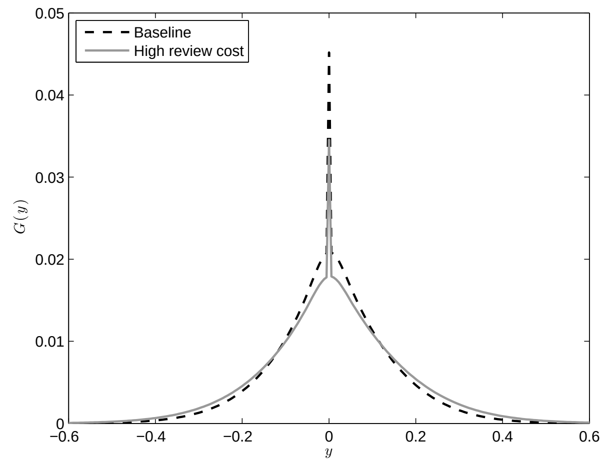


Figure 9: The firm's prior under different review policies.

Distribution of post-review states relevant to the pricing decision for baseline parameterization and for a high cost of reviews, κ. The atom at $y = 0$ reflects the state in which a review has just occurred (in the absence of transitory shocks).

4.4 Discreteness

I obtain discrete prices in an infinite-horizon pricing model with Gaussian shocks. The shape of the firm's objective π and the shape of the distribution of the state G determine the firm's pricing policy between reviews, and whether or not this pricing policy has a continuous or a discrete support. The objective function is asymmetric and the fact that the firm can occasionally revise its policy yields a distribution of the states that are relevant to the pricing decision whose support – while unbounded – is skewed and has negative excess kurtosis. I show numerically that these effects are strong enough to generate a discrete support for a finite cost of the price signal.

Recalling the optimality conditions that determine the optimal price support, note that the solution is continuous if $Z\left(q; \overline{f}\right) = 1$ for all $q \in \mathbb{R}$. In this case, equations (21) and (25) are necessary and sufficient to fully characterize the unique optimal pricing policy for a given review policy. On the other hand, the solution is necessarily discrete if one can find a set of prices that satisfy equations (21) and (25), but which yield either $\overline{f}(q) = 0$ or $Z\left(q; \overline{f}\right) < 1$ for any point in this set. The function Z represents the value of charging each normalized price q, and the optimal signalling mechanism equates this value across all prices in the support and furthermore requires that all other prices yield a weakly lower value.

Figure 10 illustrates how the firm's pricing policy evolves as a function of the cost of the price signal θ^p, keeping the review policy fixed. The panels plot the evolution of $Z\left(q;\overline{f}\right)-1$ as a function of q, for decreasing levels of the information cost. Single-price and multiple-price policies are optimal for different ranges of θ^p, and the cardinality of the solution increases as the cost of information is decreased.

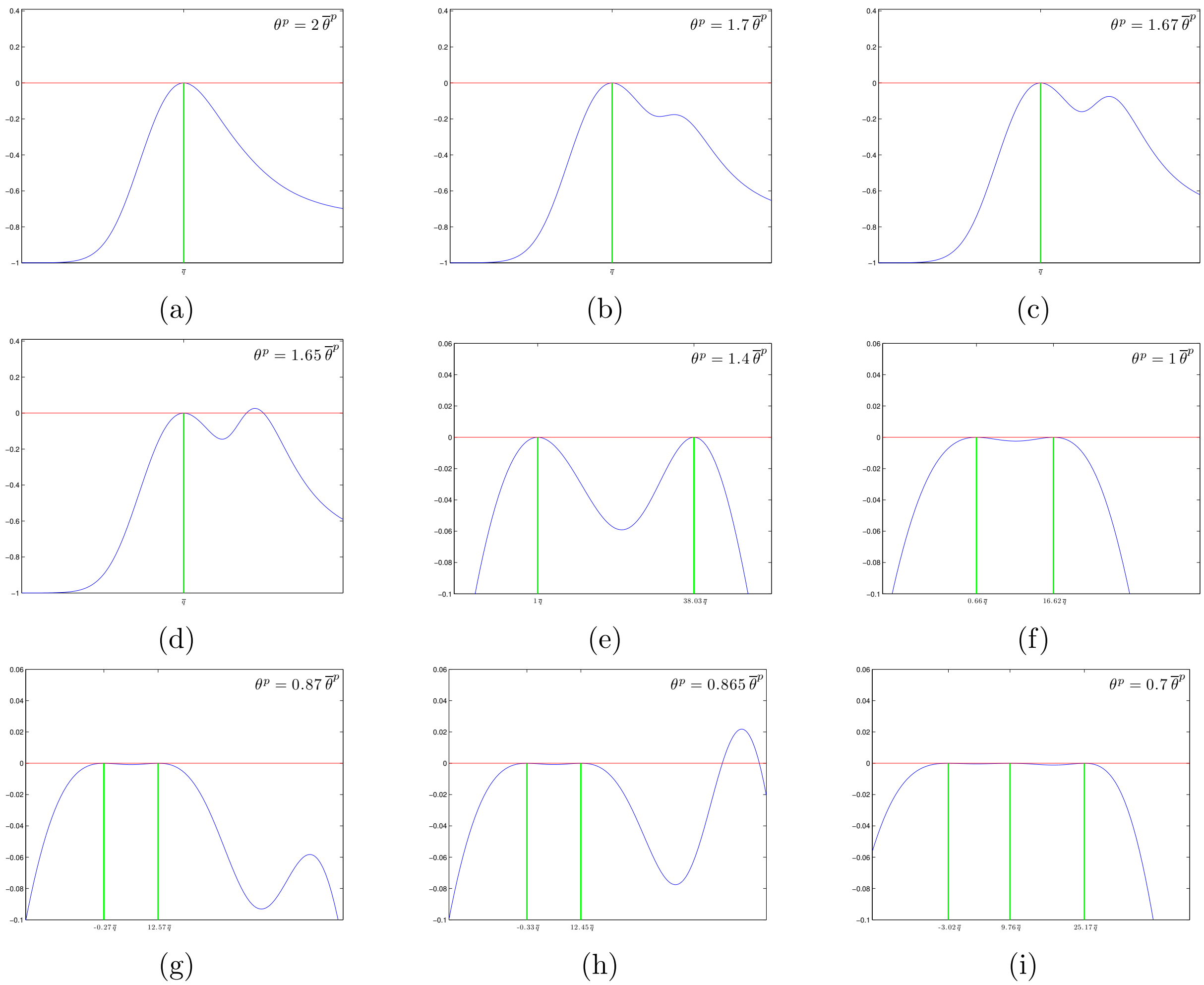


Figure 10: Growth of new mass points in the price distribution.

Simulation. The panels plot the function $Z\left(q;\overline{f}\right)-1$ as the cost of information θ^p is reduced. The points of support, for which $Z\left(q;\overline{f}\right)=1$ and $\overline{f}(q)>0$, are shown as multiples of $\overline{q}$, the price that would be charged under the single-price policy.

Consider first the optimal pricing policy for a very high information cost. In this case, the solution converges to a singleton, $\mathcal{Q}=\{\overline{q}\}$. The function Z is below 1 everywhere except at $\overline{q}$. As the information cost falls, the function Z increases for all points around $\overline{q}$. However, the growth occurs at a much faster rate in the range that will contain the new mass point. Eventually, $Z>1$, triggering the addition of a new mass point to the optimal support. Moreover, there is no other fast-growing area over the entire range of q, such that the transition from the single-price to the multiple-price policy occurs with the growth of a *single* new mass point. This is due to the asymmetry of the problem: new mass points are added one by one to the support, spreading out over a wider and wider range of possible prices. In a setup that retains the skinny tails of the distribution of states relative to the objective function (such that discreteness remains optimal) but instead employs a symmetric objective and a symmetric distribution of states, the singleton price would "break" into two and be replaced by a price below $\overline{q}$ and a price above $\overline{q}$ simultaneously. As the cost of information is further reduced, a low price and a high price would continue to be added symmetrically. In the quadratic-normal setup, for any finite information cost, $Z\left(q;\overline{f}\right) = 1$ for all $q \in \mathbb{R}$, as the optimal price support "breaks" to the entire real line immediately.

Since the firm's pricing problem has been transformed into a static problem, I can relate the optimal solution to existing results in both the rational inattention and the information theory literatures. At one extreme, a perfectly symmetric setup with a normally distributed state and a quadratic objective yields a signal whose support is the entire real line. See, for example, Sims (2003) or Maćkowiak & Wiederholt (2009) in economics and Cover & Thomas (2006) as a reference text in information theory.[21] At the other extreme, a setup in which the state is drawn from a distribution with bounded support yields a signal with a discrete support, regardless of the shape of the objective function, as shown by Matějka (2010) and Matějka & Sims (2010) in economics and by Fix (1978) in the information theory literature.

Departures from these extremes no longer guarantee a clear outcome. In the general case, the signal endogenously allocates more precision to the regions of the state space that have the potential to generate larger losses from inaccurate signals. Asymmetry in the objective function implies that more attention needs to be allocated to the steeper part of the objective, since that part generates larger losses from deviating from the full-information optimum. On the other hand, depending on the distribution of shocks, attention is allocated first to the area with more mass, and negative excess kurtosis requires less attention in the tails. Hence all of these features have the potential to generate discrete solutions.

Fix (1978) discusses the solution to rate distortion problems and argues that, for a given information cost, the optimal support is *either* the entire real line or a discrete set of points, so that the solution cannot consist of disjoint intervals. Intuitively, there cannot be "holes" in the support of the signal unless the support is discrete. Otherwise, the decision-maker's objective could be increased by employing an alternative signalling mechanism in which precision from the continuous part of the support is moved to the sparse part of the support. Matějka & Sims (2010) undertake a broader analysis that seeks to derive analytical conditions for the optimality of a discrete solution. The analysis in this paper can be seen as

[21] In the quadratic-normal case, not only is the optimal support the entire real line, but the optimal signal is also normally distributed.

complementary to theirs, in that I demonstrate how equations (21), (25) and (29) can be used numerically to find the optimal support. The numerical algorithm builds on algorithms from the information theory literature, namely Arimoto (1972), Blahut (1972), Csiszár (1974), and Rose (1994).

5 Macro Implications

The theory generates a rich set of testable implications, beyond those for statistics in the micro data. I illustrate three specific predictions that relate to ongoing debates in the literature: (*i*) the sluggish response to shocks, and the importance for aggregate rigidity of choosing both the review policy and the pricing policy to be jointly optimal; (*ii*) the policy adjustment in response to an unanticipated increase in volatility, and the comparison of the model's predictions for aggregate non-neutrality for low versus high volatility environments; and (*iii*) the response of prices to progress in the technology to acquire and process information and to an increase in the degree of competition among firms.

5.1 Sluggish Adjustment

To assess the degree of monetary non-neutrality implied by the model, I compute the price response to a monetary shock under the assumption that all aggregate variables evolve according to the flexible price equilibrium. I report the impulse response function for the price index of a set of information-constrained firms. Hence, this response provides an upper bound for the degree of monetary neutrality, since by assuming that the aggregate price index evolves flexibly I essentially abstract from strategic complementarities in price setting.

In response to a one-standard deviation nominal shock, the price index for the baseline MPP parameterization adjusts gradually, reaching full neutrality after about one and a half years, even though prices change with a frequency of 31.3%. The low degree of aggregate flexibility reflects imperfect information along three dimensions: imprecision in the timing of policy reviews, low cardinality in the set of prices that can be charged between reviews, and imprecision in the selection of which price to charge between reviews. Of the three, the timing of policy reviews is relatively more important for aggregate price dynamics.

Importantly, the firm's ability to change prices between policy reviews does not reduce the model's implied aggregate rigidity relative to that implied by the single-price-policy parameterization. The responses, plotted in Figure 11, are very similar for both the SPP and the MPP price indices, even though under the SPP parameterization, the frequency of price changes is only 4.1%. Hence, in this model, the frequency of price changes is completely divorced from the degree of aggregate rigidity. This outcome reflects the interdependence

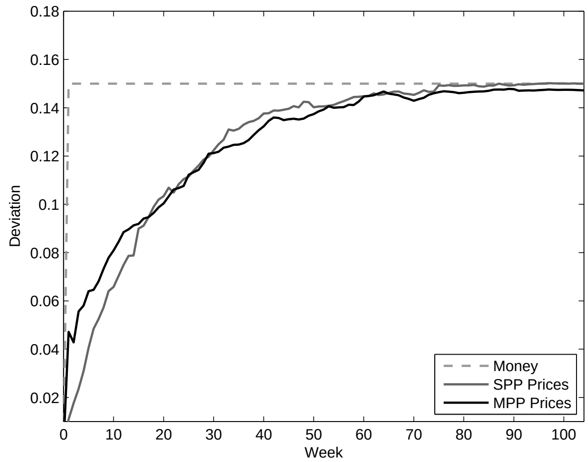


Figure 11: IRF to nominal shock, MPP versus SPP.

between the firm's pricing policy and its review policy. Since the optimal MPP review hazard function is much flatter than the optimal SPP hazard function, the MPP firm trades off accuracy in the timing of policy reviews for additional accuracy in its pricing decision between reviews. But the timing of policy reviews is an important determinant of the degree of aggregate sluggishness, since policy reviews are associated with a complete resetting of the firm's policy, based on more accurate information about the current state and in response to the sufficient accumulation of persistent shocks.

The fact that high price volatility does not necessarily imply low monetary neutrality has been discussed in the literature, with prominent examples being Kehoe & Midrigan (2010) and Eichenbaum et al. (2011). However, this paper generates this result in the context of a model in which the firm chooses its policy optimally, thereby endogenously generating the price plans postulated by Eichenbaum et al. (2011).

5.2 A Rise in Volatility

Variations in the volatility of the underlying shocks have come to the forefront of the macro literature, especially in light of the large volatility in outcomes observed in the past several years in both the U.S. and Europe.

An increase in the volatility of shocks affects both inflation and the degree of monetary non-neutrality in this model. First, it generates slight inflation, as firms raise or keep prices high to protect themselves from the potential losses associated with facing a more uncertain environment. Higher volatility increases the losses from having imprecise information about market conditions. As a result, it affects both the firm's review policy and its pricing policy. Expenditure on all ways of acquiring information increases, to compensate for the negative effect on profits of the increased volatility. Nevertheless, the increased expenditure on information is not large enough to completely offset the negative effects of facing a more volatile environment, and as an additional protective measure, the price level also rises.

Second, despite the higher volatility, the model generates the same degree of monetary non-neutrality as the baseline low volatility case. The existing literature has found that the effectiveness of monetary policy declines when volatility rises. For example, Vavra (2014) shows this result in the context of a menu cost model with stochastic volatility. In contrast to the existing literature, I find that the speed of adjustment to aggregate nominal shocks is unchanged when compared across periods of high versus low volatility, for a given parameterization of the information costs. Figure 12 shows the impulse response function of the information-constrained price index to a one standard deviation nominal shock: the response in the high volatility environment ($\sigma_\xi = 0.033$) is essentially identical to the response in the low volatility environment ($\sigma_\xi = 0.027$).

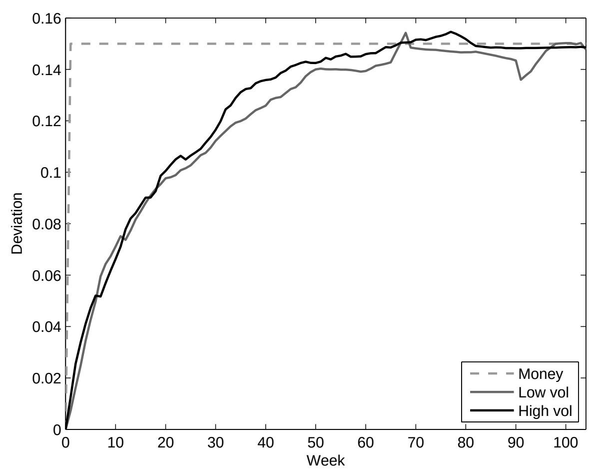


Figure 12: IRF to nominal shock, high versus low volatility periods.

This outcome reflects the endogenous response of the firm's information acquisition policy: although the firm increases information expenditure, it nevertheless has less information *relative* to the uncertainty it faces in the higher volatility environment. Given its information costs, it is not optimal for the firm to completely undo the effects of the rise in volatility. Hence it is not always the case that periods with higher volatility necessarily result in lower monetary policy effectiveness.

5.3 Competition and Progress in Information Technology

The model predicts that progress in the technology to acquire information results in modest deflation. If firms cannot perfectly track market conditions, they will set relatively high prices on average, to avoid the large losses that come from charging a price that is too low relative to the optimum. Hence, if there is progress in the technology that allows them to monitor market conditions, they can better track the optimal target price, and hence they can afford to lower their prices on average. Figure 13 illustrates this effect by showing the response of the price index of information-constrained firms to a one-time permanent, unanticipated decline in the cost of the firm's pricing policy, from $\theta^p = 0.21$ to $\theta^p = 0.08$. This decline generates an increase in the frequency of price changes and an increase in the cardinality of the firm's pricing policy from two prices to four prices with positive mass.

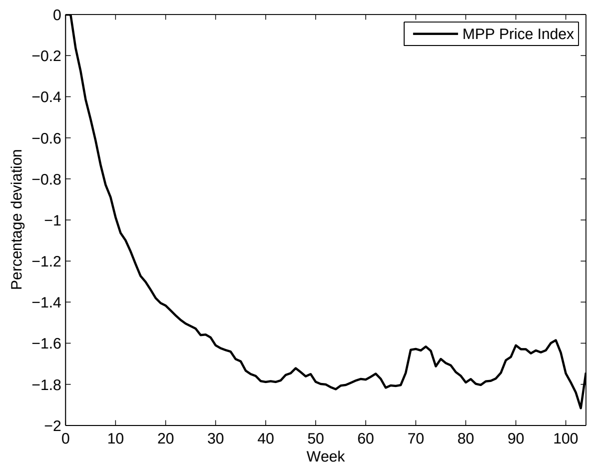


Figure 13: Response to a reduction in information costs.

Response of MPP Price Index to unanticipated decline in cost of monitoring market conditions to choose prices, from $\theta^p = 0.21$ to $\theta^p = 0.08$.

Increased competition, in the form of a higher elasticity of substitution, also generates price deflation. Stronger competitive pressure leads to larger losses when deviating from the optimal full information target price. Hence, in response, the firm increases its expenditure on all types of information, for both the review policy and the pricing policy. Larger information acquisition in turn implies a more precise pricing policy and a more precise review policy, with larger frequency and size of adjustment in both policies and prices. Hence, the endogenous acquisition of more information reinforces the deflationary pressures associated with having to charge a lower markup.

In summary, the model suggests that, in addition to the other factors highlighted in the literature, such as better monetary policy or smaller shocks, low modern inflation rates may also be partially attributable to information costs trending down and to competitive pressures rising over time.

6 Conclusion

This paper argues that firms' *choice* of how much information to acquire to set prices determines aggregate price dynamics through the patterns of pricing at the micro level, and through the large heterogeneity in pricing policies across firms. Viewed through this lens, aggregate price dynamics n the Great Recession and its aftermath, a period of high uncertainty and large shocks, becomes less puzzling, as firms endogenously adjust their information acquisition strategies.

The paper presents evidence that firms set pricing policies rather than individual prices, and develops a theory of price-setting in which firms design simple pricing policies that they update infrequently. The only friction is that all information that is relevant to the firm's pricing decision is costly. Both the decision of which price to charge from the current policy and the decision of whether or not to conduct a review and design a new policy are based on costly, noisy signals about market conditions. The precision of these signals is chosen endogenously, at the time of the policy review, subject to a unit cost for the information conveyed by each signal.

The theory generates pricing policies that are identified by discrete jumps when the policy is reviewed, and are furthermore characterized by within-policy discreteness, reflecting the firm's desire to economize on information costs between reviews. In this model, neither the frequency of policy changes, nor the frequency of price changes are sufficient statistics for the speed with which prices incorporate changes in market conditions. Nevertheless, the model generates considerable sluggishness in response to nominal shocks.

The model allows firms to vary the quantity of information acquired over time, in response to variations in market conditions, in particular in the volatility of shocks. I leave for future work the question of whether cyclicality in the acquisition of information can further account for the dynamics of inflation over the business cycle in general, or in response to very large shocks, such as the Great Recession, in particular.

The model also abstracts from other potentially important drivers of product-level price volatility, including price discrimination. Embedding price discrimination alone in an otherwise full information, flexible price stochastic model may not generate the discrete price adjustment seen in the data. However, introducing a price discrimination motive inside the information-constrained framework remains a potentially promising avenue of research. Specifically, it may help better explain the larger degree of stickiness observed at the high price within each policy.

References

Anderson, Eric, Nir Jaimovich & Duncan Simester (2012), "Price Stickiness: Empirical Evidence of the Menu Cost Model," Working Paper, Northwestern University.

Anderson, Eric, Benjamin A. Malin, Emi Nakamura, Jon Steinsson & Duncan Simester (2015), "Informational Rigidities and the Stickiness of Temporary Sales," Staff Report 513, Federal Reserve Bank of Minneapolis.

Arimoto, Suguru (1972), "An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels," *IEEE Transactions on Information Theory* 18(1): 14–20.

Bai, Jushan & Pierre Perron (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica* 66(1): 47–78.

Ball, Laurence & Sandeep Mazumder (2011), "Inflation Dynamics and the Great Recession," *Brookings Papers on Economic Activity* 2011(1): 337–381.

Beraja, Martin, Erik Hurst & Juan Ospina (2014), "The Regional Evolution of Prices and Wages During the Great Recession," Working Paper, University of Chicago.

Blahut, Richard E. (1972), "Computation of Channel Capacity and Rate Distortion Functions," *IEEE Transactions on Information Theory* 18(4): 460–473.

Burstein, Ariel T. (2006), "Inflation and Output Dynamics with State-Dependent Pricing Decisions," *Journal of Monetary Economics* 53(7): 1235–1257.

Caballero, Ricardo J. & Eduardo M.R.A. Engel (1993), "Microeconomic Adjustment Hazards and Aggregate Dynamics," *The Quarterly Journal of Economics* 108(2): 359–383.

Calvo, Guillermo A. (1983), "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics* 12(3): 383–398.

Carlstein, Edward (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics* 16(1): 188–197.

Cover, Thomas M. & Joy A. Thomas (2006), *Elements of Information Theory*, Wiley-Interscience, New York, 2 ed.

Csiszár, Imre (1974), "On the Computation of Rate Distortion Functions," *IEEE Transactions on Information Theory* 20(1): 122–124.

Del Negro, Marco, Marc P. Giannoni & Frank Schorfheide (2015), "Inflation in the Great Recession and New Keynesian Models," *American Economic Journal: Macroeconomics* 7(1): 168–196.

Deshayes, Jean & Dominique Picard (1986), "Off-Line Statistical Analysis of Change-Point Models using Nonparametric and Likelihood Methods," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, pp. 103–168, Springer Berlin Heidelberg.

Eichenbaum, Martin, Nir Jaimovich & Sergio Rebelo (2011), "Reference Prices, Costs, and Nominal Rigidities," *The American Economic Review* 101(1): 234–262.

Eichenbaum, Martin, Nir Jaimovich, Sergio Rebelo & Josephine Smith (2014), "How Frequent Are Small Price Changes?" *American Economic Journal: Macroeconomics* 6(2): 137–155.

Fix, Stephen L. (1978), "Rate Distortion Functions for Squared Error Distortion Measures," in *Proceedings of the Sixteenth Annual Allerton Conference on Communication, Control, and Computing*, pp. 704–711.

Gilchrist, Simon, Raphael Schoenle, Jae Sim & Egon Zakrajsek (2014), "Inflation Dynamics During the Financial Crisis," Working Paper, Boston University.

Guimaraes, Bernardo & Kevin D. Sheedy (2011), "Sales and Monetary Policy," *The American Economic Review* 101(2): 844–876.

Hall, Robert E. (2011), "The Long Slump," *The American Economic Review* 101(2): 431–69.

Kehoe, Patrick J. & Virgiliu Midrigan (2010), "Prices Are Sticky After All," NBER Working Paper No. 16364.

Klenow, Peter J. & Oleksiy Kryvtsov (2008), "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?" *The Quarterly Journal of Economics* 123(3): 863–904.

Klenow, Peter J. & Benjamin A. Malin (2010), "Microeconomic Evidence on Price-Setting," in *Handbook of Monetary Economics*, B.M. Friedman & M. Woodford, eds., vol. 3A, pp. 231–284, Elsevier, Amsterdam, Holland.

Maćkowiak, Bartosz A. & Mirko Wiederholt (2009), "Optimal Sticky Prices under Rational Inattention," *The American Economic Review* 99(3): 769–803.

Maćkowiak, Bartosz A. & Mirko Wiederholt (2010), "Business Cycle Dynamics Under Rational Inattention," CEPR Discussion Paper No. 7691.

Matějka, Filip (2010), "Rationally Inattentive Seller: Sales and Discrete Pricing," CERGE-EI Working Paper No. 408.

Matějka, Filip & Alisdair McKay (2011), "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," Working Paper, Boston University.

Matějka, Filip & Christopher A. Sims (2010), "Discrete Actions in Information-Constrained Tracking Problems," Working Paper, Princeton University.

Nakamura, Emi & Jón Steinsson (2008), "Five Facts about Prices: A Reevaluation of Menu Cost Models," *The Quarterly Journal of Economics* 123(4): 1415–1464.

Nakamura, Emi & Jón Steinsson (2013), "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications," *Annual Review of Economics* 5(1): 133–163.

Paciello, Luigi (2012), "Monetary Policy and Price Responsiveness to Aggregate Shocks under Rational Inattention," *Journal of Money, Credit and Banking* 44(7): 1375–1399.

Paciello, Luigi & Mirko Wiederholt (2014), "Exogenous Information, Endogenous Information and Optimal Monetary Policy," *The Review of Economic Studies* 81(1): 356–388.

Pasten, Ernesto & Raphael Schoenle (2014), "Rational Inattention, Multi-Product Firms and the Neutrality of Money," Working Paper, Central Bank of Chile.

Reis, Ricardo (2006), "Inattentive Producers," *The Review of Economic Studies* 73(3): 793–821.

Rose, Kenneth (1994), "A Mapping Approach to Rate-Distortion Computation and Analysis," *IEEE Transactions on Information Theory* 40(6): 1939–1952.

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

Shannon, Claude E. (1959), "Coding Theorems for a Discrete Source with a Fidelity Criterion," *Institute of Radio Engineers National Convention Record* 4: 142–163.

Sims, Christopher A. (2003), "Implications of Rational Inattention," *Journal of Monetary Economics* 50(3): 665–690.

Sims, Christopher A. (2006), "Rational Inattention: Beyond the Linear-Quadratic Case," *The American Economic Review* 96(2): 158–163.

Stevens, Luminita (2011), "Pricing Regimes in Disaggregated Data," Working Paper, University of Maryland.

Stevens, Luminita (2012), "Price Adjustment in a Model with Multiple Price Policies," Working Paper, University of Maryland.

Vavra, Joseph S. (2014), "Inflation Dynamics and Time-varying Volatility: New Evidence and an Ss Interpretation," *The Quarterly Journal of Economics* 129(1): 215–258.

Woodford, Michael (2008), "Inattention as a Source of Randomized Discrete Adjustment," Working Paper, Columbia University.

Woodford, Michael (2009), "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics* 56(S): 100–124.

Coarse Pricing Policies Appendix

Luminita Stevens

Federal Reserve Bank of Minneapolis &

University of Maryland

A Appendix: Empirical Method

This Appendix details the empirical method, its robustness across data generating processes, and the comparison with filters that seek to identify changes in regular or reference prices, rather than changes in pricing policies.

A.1 The Break Test

Test Statistic

Let $\{p_1, p_2, ..., p_n\}$ be a sequence of n price observations and define T_n as the set of all possible break points, $T_n \equiv \{t | 1 \leq t < n\}$. For every hypothetical break point $t \in T_n$, the Kolmogorov-Smirnov distance between the samples $\{p_1, p_2, ..., p_t\}$ and $\{p_{t+1}, p_{t+2}, ..., p_n\}$ is

$$D_n(t) \equiv \sup_p |F_{1,t}(p) - G_{t+1,n}(p)|,$$

where $F_{1,t}$ and $G_{t+1,n}$ are the empirical cumulative distribution functions of the two sub-samples, $F_{1,t}(p) \equiv \frac{1}{t}\sum_{s=1}^{t} \mathbf{1}_{\{p_s \leq p\}}$ and $G_{t+1,n}(p) \equiv \frac{1}{n-t}\sum_{s=t+1}^{n} \mathbf{1}_{\{p_s \leq p\}}$.

Following Deshayes and Picard (1986), the test statistic to test the null hypothesis of no break on a sample of size n is

$$S_n \equiv \sqrt{n} \max_{t \in T_n} \left[\frac{t}{n} \left(\frac{n-t}{n} \right) D_n(t) \right].$$

The normalization factor depends on the relative sizes of the two sub-samples, ensuring that the test is less likely to reject the null when one of the two sub-samples is relatively short, thus providing a less precise estimate of the population CDF for that sample.

If the null is rejected ($S_n > K$, where K is the critical value determined below), the estimate of the location of the break is given by Carlstein's (1988) statistic,

$$\tau_n \equiv \arg\max_{t\in T_n} \sqrt{\frac{t\,(n-t)}{n}} D_n(t).$$

To apply this method to series that may have multiple breaks at unknown locations, I first test for the existence of one break and estimate its location. I then apply the same process to each of the two resulting sub-series.

Critical Value

The only aspect of the algorithm that remains to be specified is the critical value used to reject the null of no break. The critical value (and the test statistics themselves) can be tailored to individual processes. However, good-level price series are notoriously heterogeneous, hence the specification of the test should be robust *across* different types of processes. Hence, I assume that the true data generating process for product-level prices is a mixture of different processes and I use simulations to determine a single critical value to be used across all of the simulated processes.

The existing literature on estimating breaks using Kolmogorov-Smirnov focuses on the identification of a single break. For the test of a single break at an unknown location, on observations that are drawn independently from a continuous distribution, Deshayes and Picard (1986) show that under the null hypothesis of no breaks at any $t \in T_n$,

$$S_n \to \widetilde{K} \equiv \sup_{u\in[0,1]} \sup_{v\in[0,1]} |B(u,v)|\,,$$

where $B(\cdot,\cdot)$ is the two-dimensional Brownian bridge on $[0,1]$.[1] This result provides asymptotic critical values for the test of a single break on i.i.d. data from continuous distributions. However, these values are not directly applicable to my setting. Starting from the critical values provided by Deshayes and Picard (1986), I determine the appropriate critical value using simulations in which I compare the results of the test with the true break locations. For simplicity, I use a single critical value across all sample sizes.

Simulations I simulate four processes that represent both recent theoretical models of price-setting and the most commonly observed price patterns in micro data: (i) sequences of infrequently updated *single sticky prices*, such as those generated by Calvo or simple menu cost models; (ii) sequences of *one-to-flex policies*, defined as single sticky prices accompanied by flexible deviations from these rigid modes, consistent with the dual menu cost model of Kehoe and Midrigan (2010) and with the evidence on reference prices of Eichenbaum et al. (2011); (iii) sequences of *downward-flex policies*, which consist of a single sticky price accompanied by flexible downward deviations, consistent with the dynamic version of the price discrimination model by Guimaraes and Sheedy (2011) and with the sales filtered evidence of Nakamura and Steinsson (2008); and (iv) sequences of *coarse multiple-price*

[1]For the test of a single change point at a *known* location, the normalized Kolmogorov-Smirnov statistic converges to a Brownian bridge on $[0,1]$.

policies, each consisting of a small number of distinct prices that are revisited over the life of the policy, consistent with the price plans postulated by Eichenbaum et al. (2011).

For process (*i*), the simulated series is given by

$$p_{t+1} = b_{t+1} \exp\{\varepsilon_{t+1}\} + (1 - b_{t+1})\, p_t,$$

where b_t is a Bernoulli trial with probability of success $\beta \in (0,1)$, marking the transition to a new price level, and $\varepsilon_t \sim \mathcal{N}(\mu, \sigma^2)$, i.i.d. This series also corresponds to the regular price series, p^R_{t+1}, for the multiple-price processes (*ii*), (*iii*) and (*iv*). In these cases, $b_t = 1$ marks the transition to a new policy.

For process (*ii*), the simulated series is given by

$$p_{t+1} = b_{t+1} \exp\{\varepsilon_{t+1}\} + (1 - b_{t+1}) \left[d_{t+1} p^R_t \exp\left(\varepsilon^T_{t+1}\right) + (1 - d_{t+1})\, p^R_t \right],$$

where d_t is a Bernoulli trial with probability of success $\delta \in (0,1)$, marking the transition to a new transitory price, which is given by a mean zero i.i.d. innovation, $\varepsilon^T_t \sim \mathcal{N}(0, \sigma^2_T)$.

For process (*iii*), in addition to imposing that essentially all transitory price changes are price *cuts*, by assuming that the mean of the transitory deviations is far below that of the permanent innovations, $\varepsilon^T \sim N(\mu_T, \sigma^2_T)$, with $\mu_T + 3\sigma_T < \mu - 3\sigma$, I also allow transitory prices to last up to three periods, with the maximum length of a transitory price parameterized by l_δ, with $0 \leq l_\delta \leq 3$.

Process (*iv*) is generated by collapsing the simulated values from process (*ii*) inside each policy to three bins, such that each policy consists of only three distinct prices.

These processes are parameterized to the volatility of the prices in micro data: I target a range for the mean absolute size of price changes of $10 - 15\%$, and a range for the frequency of price changes of $15 - 20\%$. Prices in the single sticky price process change with a frequency of 3%. I eliminate from simulations all policy realizations that last only one period. The performance of the test is robust to moderate variations in volatility.

Critical Values The critical value is determined using two statistics: *positive* and *negative*. The statistic *positive* reports the number of times that the test correctly rejects the null of no break on a sub-sample, as a fraction of the number of true breaks in the simulation. A low value implies that the test is not sensitive enough, such that many breaks are not identified. Correcting this requires reducing the critical value used. The statistic *negative* reports the number of times that the test incorrectly rejects the null of no break on a sub-sample that does not contain a break, as a fraction of the number of breaks estimated by the test. A high value implies that the test yields too many false positives, hence the critical value needs to be increased. Given the iterative nature of the method, the critical value determines only how soon the algorithm stops in its search for breaks: for two critical values $K_2 > K_1$, the corresponding sets of estimated break points satisfy $T_2 \subset T_1$. Hence reducing the critical value will add new breaks, without affecting the location of the existing breaks.

Table A.1 reports the performance of the break test for different critical values, starting from the asymptotic 1% and 5% significance levels provided by Deshayes and Picard (1986). The asymptotic critical values are too conservative for this setting. Using the critical value associated with the 5% significance level, the break test correctly finds only 87% of the simulated breaks on average, across all processes. The test fails to identify relatively short policy realizations, overestimating the average policy length by six periods.

TABLE A.1
BREAK TEST CRITICAL VALUE

Critical value, K	0.874	0.772	0.7	**0.61**	0.6	0.5	0.4
Positive (min, % true)	83.6	85.8	87.9	**90.1**	90.2	91.9	93.7
Positive (mean, % true)	83.9	86.5	88.5	**90.8**	90.9	93.2	95.0
Negative (max, % test)	0.2	0.8	1.8	**4.7**	5.1	10.2	35.2
Negative (mean, % test)	0.1	0.3	0.7	**1.3**	1.4	4.9	12.2
Exact synch (min, % true)	91.0	90.9	90.7	**90.5**	90.4	90.4	90.3
Exact synch (mean, % true)	93.4	93.4	93.3	**93.2**	93.2	93.2	93.1
Distance to truth (mean, weeks)	2	2	2	**2**	2	2	2
Length overshoot (mean, weeks)	+7	+6	+5	**+3**	+3	−0.2	−5

Break test simulation results for different critical values, across the four simulated processes. The critical values $K = 0.874$ and $K = 0.772$ are the asymptotic 1% and 5% significance levels provided by Deshayes and Picard (1986). *Positive (% true)* is the fraction of times that the test correctly rejects the null of no break, for each simulated process, reported as the minimum and the mean across all processes. *Negative (% test)* is number of times that the test incorrectly rejects the null of no break as a fraction of the total number of breaks found by the test, reported as the maximum and the mean across all simulated processes. *Exact synch (% true)* is the number of breaks found at the exact simulated location, as a fraction of the total number of breaks in the simulation, reported as both the minimum and the average across the four processes. *Distance to truth* is the average gap (number of periods) between the test estimate of the break location and the true location, excluding exact synchronizations, using a standard nearest-neighbor method. *Length overshoot* is the average number of periods by which the test overshoots the average length of policy realizations.

Reducing the critical value improves the test's performance: $K = 0.61$ is the threshold critical value for which the *positive* rate is at least 90% for all processes, while the *negative* rate is at most 5% for all processes. On average, across all processes, this critical value yields a 91% positive rate, and only a 1% negative rate. The average length of the policy realizations identified by the break test is longer than the true average length by three periods, reflecting the weak power in identifying policies that last between two and four periods. Restricting the simulations to policies lasting at least five weeks would ensure the identification of virtually all breaks and would eliminate the bias in the estimated average policy length..

Upon rejection of the null, I find that the change point estimate τ_k coincides exactly with the true change point 93% of the time, and is otherwise off by two periods, on average. Importantly, neither the exact synchronization nor the average distance between the estimated breaks and the true breaks, when the two are not exactly synchronized, are meaningfully affected by the choice of the critical value, since reducing the critical value does not affect the location of existing breaks, and only adds new breaks at new locations. As a result, the synchronization between the break test and the truth is consistently at 93% and the distance to the true break is consistently two periods on average.

A.2 Comparison with Filters

I compare the break test with three existing filtering methods: a v-shaped sales filter similar to those employed by Nakamura and Steinsson (2008), the reference price filter of Eichenbaum et al. (2011), and the running mode filter of Kehoe and Midrigan (2010), which is similar to that of Chahrour (2011). These filters have been proposed to uncover stickiness in product-level pricing data once one filters out transitory price changes. For these filters, a policy is identified by the regular or reference price in effect, and a break is associated with a change in the regular or reference price.

One potential advantage of the break test relative to existing price filters is that it can identify breaks without the need to specify a priori what aspects of the distribution change over time. This generality allows me to first identify breaks in price series, and then investigate what aspects of the distribution change across breaks. In contrast, v-shaped filters identify breaks based on changes in the maximum price, while reference price/running mode filters identify breaks based on changes in the modal price over time. Simulations suggest that the break test is preferable: while each filter does particularly well on specific data generating processes, the break test does well across different processes, especially when the processes are characterized by random variation in the duration of both regular and transitory prices. By using information about the entire distribution of prices, the break test also has more accuracy in detecting the *timing* of breaks compared with methods that focuses on a single statistic, such as the modal price or the maximum price. While the existing literature has focused more on the duration of regular prices, accurately identifying the timing of breaks is particularly important for characterizing within-policy volatility. Statistics such as the number of distinct prices charged, the prevalence of the highest price as the most frequently charged price, or the existence of time-trends between breaks are sensitive to the estimated location of breaks.

I apply each filter and the break test to micro data from Dominick's Finer Foods stores, which is a familiar and frequently used data set, for comparability with the existing literature. For each filter parameterization, I report the following statistics: *Filter duration*, which is the median policy duration implied by the filter, obtained by computing the mean frequency of breaks in each product category, taking the median across categories, and then computing the implied duration for the product with the median frequency as $d = -1/\ln(1-f)$; *Ratio of breaks*, the ratio of the number of breaks found by the filter to the number of breaks found by the break test, computed for each series and averaged across all series; *Exact synch*, the number of breaks that are synchronized between the two methods, as a fraction of the number of breaks found by the break test (also computed for each series and then averaged across all series); *Gap between methods*, the median distance between the break points estimated by the two methods, excluding exact synchronizations.

Standard statistics of interest vary significantly across the parameterizations of the different filters. Hence, although intuitive, filters present an implementation challenge in that they allow for substantial discretion in both setting up the algorithm and choosing the parameters that determine what defines a transitory price change and how it is identified.

V-shaped Sales Filter

The v-shaped sales filters eliminate price cuts that are followed, within a pre-specified window, by a price increase to the existing regular price or to a new regular price. I implement the v-shaped sales filter following Nakamura and Steinsson (2008).

The algorithm requires four parameters: J, K, L, F. The parameter J is the period of time within which a price cut must return to a regular price in order to be considered a transitory sale. When a price cut is not followed by a return to the existing regular price, several options arise regarding how to determine the new regular price. The parameters K and L capture different potential choices about when to transition to a new regular price. The parameter $F \in \{0, 1\}$ determines whether to associate the sale with the existing regular price or with the new one.

I apply the filter with different parameterizations to Dominick's data, varying the sale window $J \in \{3, ..., 12\}$, $K, L \in \{1, ..., 12\}$ and $F \in \{0, 1\}$. The parameter J is the most important determinant of the frequency of regular price changes. The parameters K, L and F do not significantly affect the median implied duration of the regular price, but they do affect the timing of breaks, thus affecting the synchronization of the filter with the break test. For example, fixing $J = 3$ while varying the remaining parameters of the v-shaped filter increases the synchronization in the timing of breaks between the v-shaped filter and the break test from 65% to 80%. Hence I report results for parameterizations of K, L, F that yield the highest degree of synchronization between the v-shaped filter and the break test, for each value of J.

Table A.2 presents the results. Statistics vary significantly with the parameterization, with the median implied duration of regular prices increasing from 12 to 29 weeks as I increase the length of the sale window, J. Increasing J beyond 12 weeks no longer significantly impacts statistics. This sensitivity to the parameterization of the filter is quite strong, but not entirely specific to Dominick's data: Nakamura and Steinsson (2008) report that for the goods underlying the US CPI, one can obtain different values for the median frequency of price changes in monthly data. For the range of parameters they test, they find median durations ranging between 6 and 8.3 months.

TABLE A.2
V-SHAPED SALES FILTER PERFORMANCE

Sales window, J (weeks)	3	7	12
Filter duration (median, weeks)	12	24	29
Ratio of breaks (mean, % break test)	360	177	155
Exact synch (mean, % break test)	80	64	58
Gap between methods (median, weeks)	3	5	7

V-shaped filter results for different parameterizations on Dominick's data. *Filter duration* is the implied duration for the median frequency of breaks across product categories. *Ratio of breaks* is number of breaks found by filter divided by number of breaks found by break test, averaged across series. *Exact synch* is number of breaks that are synchronized between the two methods divided by number of breaks found by break test, averaged across series. *Gap between methods* is median distance between the break points estimated by the two methods, excluding exact synchronizations.

The filter alone cannot provide a measure of accuracy, and hence enable us to pick the best parameterization. However, the break test is expected to have at least 90% accuracy in identifying breaks in the data, if the data is a mixture of the types of processes simulated above. Hence, I compute the synchronization of the different parameterizations of the v-shaped filter with the break test.

For most parameterizations, the v-shaped method yields shorter policy realizations compared with the break test, which yields a median implied duration of 31 weeks in Dominick's data. Divergence is primarily driven by the assumption of a fixed sale window and by the fact that the filter rules out transitory price increases. Adjusting the parameters of the v-shaped filter yields a trade-off in performance: a small sales window generates many more breaks, but improves on the synchronization in the timing of the breaks found by both methods. For example, setting $J = 3$ weeks generates 360% more breaks than the break test; but 80% of the breaks found by both methods are exactly synchronized. For breaks that are not exactly synchronized, the mean distance between the break points estimated by the two methods is three weeks. Increasing the sales window still generates 55% more breaks, but substantially reduces the method's ability to estimate the timing of breaks: synchronization between the filter and the break test falls from 80% to 58%.

In summary, the v-shaped filter presents a trade-off: a short sale window captures most of the change points identified by the break test with a relatively high degree of precision, but also generates many more additional breaks, leading to an under-estimate of the rigidity of regular prices relative to the break test; a long sale window matches the median duration of regular prices, but misses the timing of breaks.

Reference Price Filter

I next implement the reference price filter proposed by Eichenbaum et al. (2011). They split the data into calendar-based quarters and define the reference price for each quarter as the most frequently quoted price in that quarter. I consider a window length in weeks $W \in \{6, 10, 13\}$.

TABLE A.3
REFERENCE PRICE FILTER PERFORMANCE

Reference window, W (weeks)	6	10	13
Filter duration (median, weeks)	24	41	51
Ratio of breaks (mean, % break test)	168	91	72
Exact synch (mean, % break test)	13	8	5
Gap between methods (median, weeks)	2	3	3

Reference price filter results for different parameterizations on Dominick's data.

Table A.3 presents the results. The median implied duration of reference prices increases from 24 to 51 weeks as I increase the length of the reference window, W. For reference windows above ten weeks, I find that less than 10% of the breaks are synchronized with the break test breaks. This low ratio is entirely due to the reference price filter imposing a fixed minimum cutoff for policy lengths, which largely assumes away the question of identifying the timing of changes in the reference price series. Since I find that the length of policies is highly variable over time, the two methods are likely to overlap exactly only by chance.

In summary, the reference price filter presents a challenge in terms of identifying the timing of policy changes.

Running Mode Price Filter

I implement the running mode filter proposed by Kehoe and Midrigan (2010), which categorizes price changes as either temporary or regular, without requiring that all temporary price changes occur from a rigid *high* price, as does the v-shaped filter. For each product, they define an artificial series called the regular price series, which is a rigid running mode of the series. Every price change that is a deviation from the regular price series is defined as temporary, whereas every price change that coincides with a change in the regular price is defined as regular. In this context, I define a policy change as a change in the regular price.

The algorithm has two key parameters: A, which determines the size of the window over which to compute the modal price, and C, a cutoff used to determine if a change in the regular price has occurred. Specifically, if within a certain window, the fraction of periods in which the price is equal to the modal price is greater than C, then the regular price is updated to be equal to the current modal price; otherwise, the regular price remains unchanged.

TABLE A.4
RUNNING MODE FILTER PERFORMANCE

Rolling window, A (weeks)	6	10	14
Filter duration (median, weeks)	27	38	34
Ratio of breaks (mean, % break test)	144	102	117
Exact synch (mean, % break test)	52	48	42
Gap between methods (median, weeks)	2	2	2

Running mode filter results for different parameterizations on Dominick's data.

Table A.4 presents the results. The running mode filter is much less sensitive to parameter changes compared with the reference or v-shaped filters. The median implied duration ranges from 27 to 34 weeks across parameterizations. This filter also improves on the synchronization of breaks found by the reference price filter: at the preferred parameterization, while exact synchronization with the break test is moderately low, at 48%, the median distance between the breaks found by the filter and those found by the break test is two weeks, indicating that the two methods are fairly close.
In summary, when parameterized to match the duration of policies found by the break test, the running mode filter is largely in agreement with the break test, with small differences in the timing of breaks.

Performance in Simulations

To better understand the performance of the different methods, I apply all methods to simulated data, for which the true location of the breaks is known. For each filter, I use the parameterization that yields the closest match between the filter and the break test (which turns out to be the parameterization that also yields the closest match between the filter and the truth). I use the four simulated processes described above: (i) Single sticky price, (ii) One-to-flex policies, (iii) Downward-flex policies, and (iv) Coarse multiple-price policies.

I report the following statistics: *Ratio of breaks (% truth)*, the number of breaks found by the method as a fraction of the true number of breaks in the simulation; *Exact synch (% truth)*, the number of breaks found by the method that coincide with true breaks, as a fraction of the true number of breaks; *Distance to truth*, the median distance between the break points estimated by the method and the true breaks, excluding exact synchronizations, using a standard nearest-neighbor method; *Length overshoot*, the median number of periods by which the method overestimates the length of policies.

TABLE A.5
FILTER PERFORMANCE IN SIMULATIONS

Method	Break test	V-shaped	Reference	Running
Ratio of breaks (% truth)	93	186	93	94
Exact synch (% of truth)	93	59	17	89
Distance to truth (median, weeks)	2	5	3	2
Length overshoot (median, weeks)	3	−9	3	2

Break test and filter results in simulated data.

Table A.5 reports the synchronization of the methods with the true break points. The v-shaped filter over-estimates the number of breaks, and reparameterizing it to match the frequency of breaks reduces the degree of synchronization with the actual break locations. The reference price filter misses the timing of breaks, and adjusting the parameterization cannot meaningfully improve on this dimension. The running mode filter parameterized to match the frequency of breaks obtained by the break test yields results that are close to the break test, with a high degree of synchronization at 89% versus 93% for the break test.

In summary, of all the filters, the running mode filter proposed by Kehoe and Midrigan (2010) performs best in simulations, especially once it is parameterized to yield a frequency of breaks that is close to the actual frequency in the data or in the simulation.

B Appendix: Proofs

The Firm's Problem Let $\overline{V}_t(\widetilde{\omega}_t)$ denote the maximum attainable value of the firm's continuation value, looking forward from the time of a policy review in an arbitrary state $\widetilde{\omega}_t$ in period t. Let

$$\Pi_{t+\tau}(\omega_{t+\tau}) \equiv \sum_{p\in\mathcal{P}_t} f_{t+\tau}(p|\omega_{t+\tau})\,\pi(p-x_{t+\tau}) - \theta^p I\left(f_{t+\tau}(p|\omega_{t+\tau}), \overline{f}(p)\right)$$

denote the firm's per-period profit in an arbitrary state $\omega_{t+\tau}$, $\tau \geq 0$, (hence after that period's transitory shock, but before receipt of the price signal), expected under the pricing policy in effect in that state, net of the cost of the price signal only, and let

$$\Gamma_{t+\tau}(\widetilde{\omega}_{t+\tau-1}) \equiv \prod_{k=1}^{\tau-1}\left[1-\Lambda_{t+k}(\widetilde{\omega}_{t+k})\right],$$

for $\tau > 1$, with $\Gamma_{t+1}(\widetilde{\omega}_t) \equiv 1$, denote the probability, expected at the time of the review, that the review policy chosen in period t, continues to apply τ periods later, when the history of states is given by $\widetilde{\omega}_{t+\tau-1}$. Under the assumption that an optimal policy will be chosen in all future policy reviews, the firm's continuation value can be expressed in terms of the firm's choices at the time of the review in period t as

$$\overline{V}_t(\widetilde{\omega}_t) = E_t\left\{\Pi_t(\omega_t) + \sum_{\tau=1}^{\infty}\beta^\tau\Gamma_{t+\tau}(\widetilde{\omega}_{t+\tau-1})\,W_{t+\tau}(\omega_{t+\tau})\right\},$$

$$W_\tau(\omega_\tau) \equiv \left[1-\Lambda_\tau(\widetilde{\omega}_\tau)\right]\Pi_\tau(\omega_\tau) + \Lambda_\tau(\widetilde{\omega}_\tau)\left[\overline{V}_\tau(\widetilde{\omega}_\tau) - \kappa\right] - \theta^r I\left(\Lambda_\tau(\widetilde{\omega}_\tau), \overline{\Lambda}_t\right),$$

so that conditional on the current policy surviving all the review decisions leading to a particular state $\widetilde{\omega}_{t+\tau}$, $\tau > 0$, the firm pays the cost of the review signal. It then continues to apply the current policy with probability $1-\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})$, in which case it attains expected profits $\Pi_{t+\tau}(\omega_{t+\tau})$, and it undertakes a policy review with probability $\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})$, in which case it pays the review cost κ and expects the maximum attainable value from that state onward, $\overline{V}_{t+\tau}(\widetilde{\omega}_{t+\tau})$. If a firm undertakes a policy review in an arbitrary state $\widetilde{\omega}_t$ and period t, it chooses a *review policy* that specifies $\overline{\Lambda}_t$ and $\{\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})\}_\tau$ for all periods $t+\tau > t$ and all states $\widetilde{\omega}_{t+\tau}$ until the next review; and a *pricing policy* that specifies $\mathcal{P}_t$, $\overline{f}_t(p)$, and $\{f_{t+\tau}(p|\omega_{t+\tau})\}_\tau$ for all $p \in \mathcal{P}_t$, all periods $t+\tau \geq t$, and all states $\omega_{t+\tau}$ until the next review, to maximize $\overline{V}_t(\widetilde{\omega}_t)$.

Since at the time of a policy review in period t, the firm learns the complete state, $\widetilde{\omega}_t$, the firm's problem can be expressed in terms of the innovations to the state since the last review. Using the normalizations defined in the text, and given the laws of motion for the pre-review and post-review target prices, $\widetilde{x}_t$ and x_t, the normalized variables $\widetilde{y}_\tau$, y_τ, and hence $\widetilde{\varpi}_\tau$, ϖ_τ, are distributed independently of the state $\widetilde{\omega}_t$ at the time of the policy review. Redefined,

$$\Pi_\tau(\varpi_\tau) \equiv \sum_{q\in\mathcal{Q}} f_\tau(q|\varpi_\tau)\,\pi(q-y_\tau) - \theta^p I\left(f_\tau(q|\varpi_\tau), \overline{f}(q)\right),$$

$$\Gamma_\tau(\widetilde{\varpi}_{\tau-1}) \equiv \prod_{k=1}^{\tau-1}\left[1-\Lambda_k(\widetilde{\varpi}_k)\right],\ \forall\tau > 1,$$

and the firm's problem becomes choosing $\overline{\Lambda}_t$, $\{\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})\}_\tau$, $\mathcal{P}_t$, $\overline{f}_t(p)$, and $\{f_{t+\tau}(p|\omega_{t+\tau})\}_\tau$ to solve

$\overline{V} = \max E\left[\Pi_0\left(\varpi_0\right) + \sum_{\tau=1}^{\infty} \beta^{\tau} \Gamma_{\tau}\left(\widetilde{\varpi}_{\tau-1}\right) W_{\tau}\left(\varpi_{\tau}\right)\right],$

$W_{\tau}\left(\varpi_{\tau}\right) \equiv \left(1 - \Lambda_{\tau}\left(\widetilde{\varpi}_{\tau}\right)\right) \Pi_{\tau}\left(\varpi_{\tau}\right) + \Lambda_{\tau}\left(\widetilde{\varpi}_{\tau}\right)\left(\overline{V} - \kappa\right) - \theta^{r} I\left(\Lambda_{\tau}\left(\widetilde{\varpi}_{\tau}\right), \overline{\Lambda}\right).$

I obtain the solution to the firm's problem in steps, deriving each element of the optimal policy taking the other elements as given.

The Conditional Distribution of Prices The firm's choice of an optimal pricing policy for a given review policy is reduced to the maximization of the term that directly depends on the pricing policy in the firm's objective,

$E\left\{\sum_{\tau=0}^{\infty} \beta^{\tau} \Gamma_{\tau+1}\left(\widetilde{\varpi}_{\tau}\right) \Pi_{\tau}\left(\varpi_{\tau}\right)\right\}.$

Consider the subproblem of choosing the optimal sequence of conditional price distributions, $\left\{f_{\tau}\left(q|\varpi_{\tau}\right)\right\}_{\tau}$, for a given review policy, and further taking as given the set of normalized prices, Q, and the anticipated frequency with which each price is charged, $\overline{f}\left(q\right) > 0$ for all $q \in Q$. For each τ and each possible news state ϖ_{τ} under the current review policy, the firm chooses the conditional distribution of normalized prices $f_{\tau}\left(q|\varpi_{\tau}\right)$ that solves

$\max_{f_{\tau}\left(q|\varpi_{\tau}\right)} \Pi_{\tau}\left(\varpi_{\tau}\right) \quad \text{s.t. } \sum_{q \in Q} f_{\tau}\left(q|\varpi_{\tau}\right) = 1 \text{ and } f_{\tau}\left(q|\varpi_{\tau}\right) \geq 0, \forall q \in Q.$

Let the Lagrangean multipliers on the constraints be denoted by μ and $\eta\left(q\right)$. For $f_{\tau}\left(q|\varpi\right) > 0$, such that $\eta\left(q\right) = 0$, differentiating with respect to $f_{\tau}\left(q|\varpi\right)$, yields

$\pi(q - y_{\tau}) - \theta^{p}\left[\log f_{\tau}\left(q|\varpi_{\tau}\right) - \log \overline{f}\left(q\right)\right] - \left(\theta^{p} + \mu\right) = 0.$

Rearranging, and letting $\phi \equiv \exp\left\{1 + \frac{\mu}{\theta^{p}}\right\}$ yields

$f_{\tau}\left(q|\varpi_{\tau}\right) = \frac{1}{\phi} \overline{f}\left(q\right) \exp\left\{\frac{1}{\theta^{p}} \pi(q - y_{\tau})\right\}.$

Summing over q yields

$f_{\tau}\left(q|\varpi_{\tau}\right) = \overline{f}\left(q\right) \frac{\exp\left\{\frac{1}{\theta^{p}} \pi(q - y_{\tau})\right\}}{\sum_{\widehat{q} \in Q} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^{p}} \pi(\widehat{q} - y_{\tau})\right\}}.$

Finally, note that if $\overline{f}\left(q\right) > 0$, then $f_{\tau}\left(q|\varpi\right) > 0$, such that the multiplier $\eta\left(q\right)$ is indeed zero for all q, as was assumed above.

The conditional distribution, $f_{\tau}\left(q|\varpi_{\tau}\right)$, only depends on ϖ_{τ} through its dependence on the normalized post-review state, y_{τ}. Moreover, it depends only on the time-invariant profit function, π, and on the invariant distribution, $\overline{f}$. Hence, we can write it directly as $f\left(q|y_{\tau}\right)$, for all $\tau \geq 0$, and for each normalized target price y_{τ} in each state ϖ_{τ}.

The Hazard Function for Reviews Consider next the firm's choice of an optimal sequence of hazard functions $\{\Lambda_\tau(\widetilde{\varpi}_\tau)\}_\tau$ for a given pricing policy, and further taking $\overline{\Lambda}$ as given. This problem can be given a recursive formulation by noting that the choice of the sequence $\{\Lambda_{\tau'}(\widetilde{\varpi}_{\tau'})\}_{\tau'}$ for all $\tau' > \tau$, looking forward from an arbitrary state $\widetilde{\varpi}_\tau$, is independent of the choices made for periods prior to τ, or for news states $\widetilde{\varpi}_{\tau'}$ that are not successors of $\widetilde{\varpi}_\tau$. Let $V_\tau(\widetilde{\varpi}_\tau)$ be the maximum attainable value of the firm's objective, from some period τ onwards. From the solution to the firm's optimal choice for the conditional distribution of prices, $f(q|y)$, the firm's per-period profit net of the cost of the price signal is an invariant function, $\Pi(y)$, for all y. The value $V_\tau(\widetilde{\varpi}_\tau)$ depends on $\widetilde{\varpi}_\tau$ only through the dependence of the expected profit $\Pi(y_\tau)$ on the value of y_τ. Since $\widetilde{y}_\tau$ is a random walk and $y_\tau = \widetilde{y}_\tau + \nu_\tau$, where ν_τ is i.i.d, then for any $\tau' \geq \tau$, the probability distributions for realizations of $\widetilde{y}_{\tau'}$ and $y_{\tau'}$ conditional on $\widetilde{\varpi}_\tau$ depend only on the value of $\widetilde{y}_\tau$. Hence, the maximum attainable value $V_\tau(\widetilde{\varpi}_\tau)$ only depends on the value of $\widetilde{y}_\tau$, $V_\tau(\widetilde{\varpi}_\tau) = V(\widetilde{y}_\tau)$, for some invariant function $V(\widetilde{y})$. The firm's choice of an optimal sequence of hazard functions has the recursive form

$$V(\widetilde{y}_\tau) = \max_{\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})} E_\tau \left\{ \Pi(y_\tau) + \beta \left[\begin{array}{c} (1-\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}))V(\widetilde{y}_{\tau+1}) + \Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})\left[\overline{V}-\kappa\right] \\ -\theta^r I\left(\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}), \overline{\Lambda}\right) \end{array} \right] \right\},$$

where E_τ integrates over all possible innovations to the state, $\widetilde{\varpi}_{\tau+1}$, that follow $\widetilde{\varpi}_\tau$ under the current review policy. For each state $\widetilde{\varpi}_{\tau+1}$, the hazard function $\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})$ is then chosen to maximize $(1-\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}))V(\widetilde{y}_{\tau+1}) + \Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})\left[\overline{V}-\kappa\right] - \theta^r I\left(\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}), \overline{\Lambda}\right)$.

This problem depends only on the value of $V(\widetilde{y}_{\tau+1})$ and is otherwise independent of the time elapsed since the last review, $\tau+1$, and of the particular history of past signals in $\widetilde{\varpi}_{\tau+1}$. Therefore, the solution is of the form $\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}) = \Lambda(\widetilde{y}_{\tau+1})$, where $\Lambda(\widetilde{y})$ is a time-invariant function. Differentiating with respect to $\Lambda(\widetilde{y}_{\tau+1})$ yields

$\overline{V} - \kappa - V(\widetilde{y}_{\tau+1}) - \theta^r \frac{\partial I\left(\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}), \overline{\Lambda}\right)}{\partial \Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})} = 0$, where

$\frac{\partial I\left(\Lambda, \overline{\Lambda}\right)}{\partial \Lambda} = \log \frac{\Lambda}{1-\Lambda} - \log \frac{\overline{\Lambda}}{1-\overline{\Lambda}}$.

Hence

$\frac{\Lambda(\widetilde{y})}{1-\Lambda(\widetilde{y})} = \frac{\overline{\Lambda}}{1-\overline{\Lambda}} \exp\left\{\frac{1}{\theta^r}\left[\overline{V} - \kappa - V(\widetilde{y})\right]\right\}$,

The maximum attainable value under the current policy can now be seen to satisfy the fixed point equation

$$V(\widetilde{y}) = E\left\{\Pi(y) + \beta\left[(1-\Lambda(\widetilde{y}'))V(\widetilde{y}') + \Lambda(\widetilde{y}')\left[\overline{V}-\kappa\right] - \theta^r I\left(\Lambda(\widetilde{y}'), \overline{\Lambda}\right)\right]\right\},$$

where E denotes expectations over all possible values $\widetilde{y}' = \widetilde{y} + \widetilde{\nu}$ and $y' = \widetilde{y} + \nu$, conditional on $\widetilde{y}$.Finally, the continuation value upon conducting a review is $\overline{V} = V(0)$.

The Frequency of Reviews For a given pricing policy, and a given hazard function for policy reviews, and using the previous two results, the optimal frequency of reviews, $\overline{\Lambda}$, is chosen to maximize

$$E\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau-1}\right)\left[\left(1-\Lambda\left(\widetilde{y}_{\tau}\right)\right)\Pi\left(y_{\tau}\right)+\Lambda\left(\widetilde{y}_{\tau}\right)\left[\overline{V}-\kappa\right]-\theta^{r}I\left(\Lambda\left(\widetilde{y}_{\tau}\right),\overline{\Lambda}\right)\right],$$

where $\Gamma\left(\widetilde{y}^{\tau-1}\right)\equiv\prod_{k=1}^{\tau-1}\left[1-\Lambda\left(\widetilde{y}_{k}\right)\right]$ for $\tau>1$, with $\Gamma\left(0\right)\equiv1$, is the policy's survival probability to period τ, which depends on the history of the pre-review target prices, $\widetilde{y}^{\tau-1}$. Holding fixed the pricing policy, the value of $\overline{V}$, and the hazard function $\Lambda\left(\widetilde{y}_{\tau}\right)$, this problem is reduced to minimizing the cost of the review signal over the expected life of the policy. Specifically, $\overline{\Lambda}$ solves

$$\min_{\overline{\Lambda}}E\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau-1}\right)I\left(\Lambda\left(\widetilde{y}_{\tau}\right),\overline{\Lambda}\right),$$

where the quantity of information acquired in each period for the review decision is given by

$$I\left(\Lambda,\overline{\Lambda}\right)\equiv\Lambda\left[\log\Lambda-\log\overline{\Lambda}\right]+\left(1-\Lambda\right)\left[\log\left(1-\Lambda\right)-\log\left(1-\overline{\Lambda}\right)\right].$$

This minimizationproblem is equivalent to maximizing

$$E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau-1}\right)\left[\Lambda\left(\widetilde{y}_{\tau}\right)\log\overline{\Lambda}+\left(1-\Lambda\left(\widetilde{y}_{\tau}\right)\log\left(1-\overline{\Lambda}\right)\right)\right]\right\}.$$

The first order condition yields

$$\overline{\Lambda}=\frac{E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau-1}\right)\Lambda\left(\widetilde{y}_{\tau}\right)\right\}}{E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau-1}\right)\right\}}.$$

The Frequency of Prices Given the results above, the firm's pricing policy maximizes $E\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau}\right)\Pi\left(y\right)$.

Holding fixed the review policy, the support of the price signal, and the conditional price distribution, the problem of choosing the optimal anticipated frequency of prices is reduced to minimizing the total cost of the price signal over the expected life of the policy. Specifically, $\overline{f}\left(q\right)>0$ solves

$$\min_{\overline{f}(q)}E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau}\right)\left[\sum_{q\in Q}f\left(q|y\right)\left[\log f\left(q|y\right)-\log\overline{f}\left(q\right)\right]\right]\right\}$$

subject to $\sum_{q\in Q}\overline{f}\left(q\right)=1$, just as the frequency of reviews, $\overline{\Lambda}$, was shown to minimize the cost of the review signal. Forming the Lagrangian with multiplier μ, the first order condition for each q charged with positive probability yields

$E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau}\right)\frac{f(q|y)}{\overline{f}(q)}\right\}=\mu$. Summing over q yields

$\mu=E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau}\right)\right\}$. Hence,

$$\overline{f}\left(q\right)=\frac{E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau}\right)f\left(q|y\right)\right\}}{E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma\left(\widetilde{y}^{\tau}\right)\right\}}.$$

Finally, the proof that $\overline{f}$ and f specify the unique optimal pricing policy among all pricing policies with support Q follows from the strict concavity of $\int G\left(y\right)\Pi\left(y\right)dy$ in f and $\overline{f}$. See also Csiszar (1974) in the information theory literature.

The Optimal Support Consider the firm's pricing objective after substituting in the optimal conditional distribution, $f(q|y)$, for a given marginal, $\overline{f}(q)$,

$$\mathcal{F}\left(\overline{f}\right) \equiv \int G(y) \log\left[\sum_{q\in Q} \overline{f}(q) \exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}\right] dy.$$

The distribution $\overline{f}(q)$ must maximize this objective subject to $\sum_{q\in Q} \overline{f}(q) = 1$ and $\overline{f}(q) \geq 0$, $\forall q$. Forming the Lagrangian with multipliers μ and $\eta(q)$ on the constraints, and differentiating with respect to $\overline{f}(q)$ yields

$$\int G(y) \frac{\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in Q} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}} dy - \mu + \eta(q) = 0.$$

For $\overline{f}(q) > 0$, such that $\eta(q) = 0$, multiplying by $\overline{f}(q)$ yields

$$\int G(y) \frac{\overline{f}(q)\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in Q} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}} dy = \mu \overline{f}(q),$$

and summing over $q \in Q$ yields the Lagrange multiplier $\mu = 1$. Hence,

$$\int G(y) \frac{\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in Q} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}} dy \leq 1,$$

with equality for each q such that $\overline{f}(q) > 0$. Finally, the fixed point equation for $\overline{f}(q)$ is obtained by integrating equation (**??**) over y.

Threshold Information Cost Following Rose (1994), the points of support must satisfy the following necessary conditions:

$$\int G(y|q) \frac{\partial \pi(q-y)}{\partial q} dy = 0,$$

$$\int G(y|q) \left[\frac{\partial^2 \pi(q-y)}{\partial q^2} + \frac{1}{\theta^p}\left(\frac{\partial \pi(q-y)}{\partial q}\right)^2\right] dy \leq 0,$$

These necessary conditions imply that the single-price policy, if optimal, is defined by the price

$$\overline{q} = \arg\max_q \int G(y)\, \pi(q-y)\, dy.$$

and the threshold cost of the price signal that is sufficiently low such that the single-price policy is not optimal is given by

$$\overline{\theta}^p \equiv \frac{\int G(y)\left(\frac{\partial}{\partial q}\pi(q-y)\right)^2 dy}{\int G(y)\left(\frac{\partial^2}{\partial q^2}\pi(q-y)\right) dy}, \text{ where the derivatives are evaluated at } \overline{q}.$$

C Appendix: Algorithm

This appendix describes the numerical algorithm that solves the firm's optimal policy.

Optimal Review Algorithm For a Given Pricing Algorithm

1. Given a distribution for the permanent shock $\tilde{\nu}$, discretize $\tilde{y}$ in ny points and compute the transition probability matrix $\tilde{\pi}(\tilde{y}'|\tilde{y})$ using the Tauchen method.

2. Guess a hazard function for policy reviews $\Lambda(\tilde{y})$.

3. Compute a finite approximation to the discounted distribution of pre-review target prices over the life of the policy $\tilde{G}(\tilde{y})$.

4. Find the implied $\overline{\Lambda} = \int \Lambda(\tilde{y})\, \tilde{G}(\tilde{y}) d\tilde{y}$.

5. Compute a finite approximation to the discounted distribution of post-review target prices over the life of the policy $G(y)$

6. Find the optimal pricing-policy following the algorithm described in the next section. This returns a vector of prices q^* with associated marginal and conditional distributions $\bar{f}(q^*)$ and $f(q^*|y)$.

7. Compute the expected profit function $\Pi(q - y|\tilde{y})$.

8. Iterate until convergence on the value function

$$V(q, \tilde{y}) = \Pi(q - y|\tilde{y}) + \beta \sum_{\tilde{y}'} V(q, \tilde{y}')\tilde{\pi}(\tilde{y}', \tilde{y}) \forall \tilde{y}$$

9. Compute the new hazard function,

$$\Lambda(\tilde{y})^{new} = \frac{\frac{\overline{\Lambda}}{1-\overline{\Lambda}} e^{\left\{\frac{1}{\theta^r}(V(q,0) - \kappa - V(q,\tilde{y})\right\}}}{1 + \frac{\overline{\Lambda}}{1-\overline{\Lambda}} e^{\left\{\frac{1}{\theta^r}(V(q,0) - \kappa - V(q,\tilde{y})\right\}}}$$

10. If the maximum difference between $\Lambda(\tilde{y})^{new}$ and $\Lambda(\tilde{y})$ is small enough, stop. Otherwise, update $\Lambda(\tilde{y})$ as follows and go back to step 3:

$$\Lambda(\tilde{y}) = \delta\Lambda(\tilde{y}) + (1 - \delta)\Lambda(\tilde{y})^{new}, 0 < \delta \leq 1$$

Optimal Pricing Algorithm For a Given Review Policy

1. Define nq as the number of prices in the pricing policy, and $q^*_{\{nq\}}$ as the optimal pricing policy with nq different prices.

2. Find the single price policy (q^{*spp}) using the algorithm described in the next section.

3. Initialize the pricing policy. $q^*_{\{1\}} = q^{*spp}$.

4. Create a dense grid of prices q^{out}, with M equally spaced prices between $\tilde{y}^{min}$ and $\tilde{y}^{max}$, which are the minimum and maximum values for $\tilde{y}$ in the grid. Define w^{out} as the space between prices in q^{out}, and add to this grid the vector of prices $q^*_{\{nq\}}$.

5. Compute the function Z^{out} for each price $\tilde{q}$ in q^{out}:

$$Z^{out}(\tilde{q}) = \int G(y) \frac{e^{\left\{\frac{1}{\theta^p}\pi(\tilde{q},y)\right\}}}{\sum_q \bar{f}(q) e^{\left\{\frac{1}{\theta^p}\pi(q,y)\right\}}} dy$$

6. Find $\tilde{q}^*$ such that:

$$\tilde{q}^* = \arg\max_{\tilde{q}} \left\{Z^{out}(\tilde{q})\right\}$$

7. Find the closest price to $\tilde{q}^*$ in the vector $q^*_{\{nq\}}$. Call that price q^{close}

8. If the distance between q^{close} and $\tilde{q}^*$ is less than w^{out}, stop and conclude that there are no more prices in the pricing policy. Otherwise, conclude that there is another price in the pricing policy q^*, and continue to the next step.

9. Increase in one unit nq, namely $nq = nq + 1$.

10. Given nq, find the optimal pricing policy $q^*_{\{nq\}}$, $\bar{f}(q^*_{\{nq\}})$ as follows:

    (a) Given a guess for $q^*_{\{nq\}} = q^{\{n\}}$, compute the optimal marginal distributions $\bar{f}(q^{\{n\}})$ using the Blahut-Arimoto algorithm described in the last section of this appendix.

    (b) Compute:

$$W(q^{\{n\}}) = \int G(y|q^{\{n\}})\pi(q^{\{n\}} - y)dy$$

$$W'(q^{\{n\}}) = \int G(y|q^{\{n\}}) \frac{\partial \pi(q^{\{n\}} - y)}{\partial q} dy$$

$$W''(q^{\{n\}}) = \int G(y|q^{\{n\}}) \left[\frac{\partial^2 \pi(q^{\{n\}} - y)}{\partial q^2} + \frac{1}{\theta^p}\left(\frac{\partial \pi(q^{\{n\}} - y)}{\partial q}\right)^2\right] dy$$

    (c) Update your guess for $q^*_{\{nq\}}$ following Newton's algorithm:

$$q^{\{n+1\}} = q^{\{n\}} - \left[W''\left(q^{\{n\}}\right)\right]^{-1} W'\left(q^{\{n\}}\right), n \geq 1$$

    (d) If the difference between $q^{\{n+1\}}$ and $q^{\{n\}}$ is small, define $q^*_{\{nq\}} = q^{\{n+1\}}$ and stop. Otherwise, go back to step 10a.

11. Go back to step 4.

Single Price Algorithm For a Given Review Policy

This algorithm assumes that the distribution $G(y)$ is known and exploits the following facts that: (i) the value function $V(q,0)$ is single peaked, and (ii) the optimal price q^* is between $[\tilde{y}^{min}, \tilde{y}^{max}]$ which are the minimum an maximum values in the grid for $\tilde{y}$.

1. Given $q^{range} = [q^{min}, q^{max}]$, define $\bar{q}$ as the mid point of q^{range}.
2. Compute the function $W(\bar{q}) = \int \pi(q-y)G(y)dy$
3. Compute the derivative $W'(\bar{q}) = \frac{\partial W(\bar{q})}{\partial q} = \int \frac{\partial \pi(q-y)}{\partial q} G(y)dy$
4. If the difference between q^{max} and q^{min}, or W', is small, $q^* = \bar{q}$. Otherwise, update q^{range} as follows and go back to step 1:

$$\begin{aligned} q^{range} &= [q^{min}, \bar{q}] \qquad if \quad W'(\bar{q}) < 0 \\ q^{range} &= [\bar{q}, q^{max}] \qquad if \quad W'(\bar{q}) > 0 \end{aligned}$$

The Blahut-Arimoto Algorithm

For a given support, the optimal marginal distribution is found by iterating on

$$\overline{f}(q) = \overline{f}(q) \int \frac{\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}} G(y)\, dy.$$

For a given $\overline{f}(q)$, the conditional distribution is then given by

$$f(q|y) = \overline{f}(q) \frac{\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}}.$$

For a proof of convergence, see Csiszar (1974).

For a given grid $Q = \{q_j\}$ of size n, the algorithm proceeds as follows:

1. Initialize $\overline{f}_j^{(0)} = 1/n$, $j = 1,..,n$.
2. Compute the $n_y \times n$ matrix d whose $(ij)^{th}$ entry is given by

$$d_{ij} = \exp\left\{\frac{1}{\theta^p}\pi(q_j - y_i)\right\}.$$

3. Compute

$$D_i = \sum_{j=1}^{n} \overline{f}_j^{(k)} d_{ij}, \quad i = 1,..,n_y;$$

4. Compute

$$Z_j^{(k)} = \sum_{i=1}^{n_y} G_i \frac{d_{ij}}{D_i}, \quad j = 1, .., n;$$

$$\overline{f}_j^{(k+1)} = \overline{f}_j^{(k)} Z_j^{(k)}, \quad j = 1, .., n.$$

5. Compute

$$TU = -\sum_{j=1}^{n} \overline{f}_j^{(k+1)} \ln Z_j^{(k)}; TL = -\max_j \ln Z_j^{(k)}.$$

   If $TU - TL$ exceeds a prescribed tolerance level, go back to the beginning of step 3.

6. Compute the resulting conditional and marginal, and the associated expected profit Π and information flow I

$$f_{jk} = \overline{f}_k \frac{d_{jk}}{D_j}; \overline{f}_k = \sum_{j=1}^{n_y} f_{jk} G_j;$$

$$\Pi = \sum_{j=1}^{n_y} \sum_{k=1}^{n} \pi(q_k - y_j) f_{jk} G_j;$$

$$I = \frac{1}{\theta^p} \Pi - \sum_{j=1}^{n_y} G_j \log D_j.$$

The upper and lower triggers, TU and TL, generate, via successive iterations, a decreasing and an increasing sequence respectively, which converge to the information flow I for a given expected profit, Π, and hence information cost, θ^p (see discussion in Blahut, 1972).

D Appendix: Model of Price Setting

This appendix maps a standard monopolistically competitive economy with a cash-in-advance constraint into the setup presented in the main body of the paper. The economy has three types of agents: a representative household, a continuum of monopolistically competitive producers of differentiated goods, and a government that follows an exogenous policy.

Households The problem of the representative household is standard. The household is perfectly informed and supplies differentiated labor $H_t(i)$ to each firm i in the economy. Each period is divided into two subperiods: a period in which asset markets open and financial exchange occurs, and a period in which goods markets are open and the goods exchange occurs. The household can finance its current money and bond holdings using current nominal income sources, and using remaining cash balances and bond income from the prior period, after financing that period's consumption. It chooses paths for consumption, hours, money and bond holdings to solve

$$\max_{\{C_t, C_t(i), H_t(i), M_t, B_t\}} E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{1}{1+\nu} \int_0^1 H_t(i)^{1+\nu} di \right]$$

subject to the budget constraint for the financial exchange,

$$M_t + B_t \leq \int_0^1 W_t(i) H_t(i) di + \int_0^1 \Pi_t(i) di + T_t + M_{t-1} + (1 + i_{t-1}) B_{t-1} - P_{t-1} C_{t-1},$$

the cash-in-advance constraint in the goods market,

$$M_t \geq P_t C_t,$$

and the definition of the consumption basket,

$$C_t \equiv \left[\int_0^1 \left[A_t(i) C_t(i) \right]^{\frac{\varepsilon-1}{\varepsilon}} di \right]^{\frac{\varepsilon}{\varepsilon-1}},$$

where $\beta \in (0, 1)$ is the discount factor, C_t is the consumption basket, with elasticity of substitution $\varepsilon > 1$ and good-specific preference shocks $A_t(i)$, $\sigma > 1$ is the constant relative risk aversion parameter, $\nu \geq 0$ is the inverse of the Frisch elasticity of labor supply, $W_t(i)$ is the nominal hourly wage of firm i, $\Pi_t(i)$ is the dividend received from firm i, T_t is the net monetary transfer received from the government, B_t is the amount of risk-free nominal bonds held in the period, i_t is the risk-free nominal interest rate on these bonds, M_t is money holdings, and P_t is the aggregate price index for the consumption basket C_t,

$$P_t \equiv \left[\int_0^1 \left(\frac{P_t(i)}{A_t(i)} \right)^{1-\varepsilon} di \right]^{\frac{1}{1-\varepsilon}}.$$

Inter-temporal consumer optimization yields the standard first order conditions:

$$W_t(i) = H_t(i)^{\nu} C_t^{\sigma} P_t \qquad \text{and} \qquad \frac{1}{1+i_t} = \beta E_t\left[\frac{C_{t+1}^{-\sigma} P_t}{C_t^{-\sigma} P_{t+1}}\right].$$

Intra-temporal expenditure minimization yields a demand function for each variety i,

$$C_t(i) = A_t(i)^{\varepsilon-1} P_t(i)^{-\varepsilon} P_t^{\varepsilon} C_t.$$

Firms Each firm produces a differentiated good i using a production function given by

$$Y_t(i) = \frac{H_t(i)^{\frac{1}{\gamma}}}{A_t(i)},$$

where $\gamma \geq 1$ denotes decreasing returns to scale in production, $H_t(i)$ is the differentiated labor input, and $A_t(i)$ denotes a firm-specific quality shock that increases both the utility from consuming the product and the effort required to produce it. The assumption that this shock shifts both the household's demand for the good and the cost of producing the good implies that the firm's profit is shifted in the same way by the aggregate nominal shock and by the idiosyncratic shock. This assumption enables a reduction in the state space of the problem, increasing tractability. The same assumption is made by Midrigan (2011) and Woodford (2009). The quality shock contains independently distributed transitory and permanent components. In logs,[2] $a_t(i) = z_t(i) + \zeta_t(i)$ and $z_t(i) = z_{t-1}(i) + \xi_t(i)$, with $\xi_t(i) \overset{i.i.d.}{\sim} h_{\xi}$, and $\zeta_t(i) \overset{i.i.d.}{\sim} h_{\zeta}$. The firm's nominal profit each period, excluding the resources spent to acquire information about market conditions, is

$$\Pi_t(i) = P_t(i)Y_t(i) - W_t(i)H_t(i).$$

Government and Market Clearing For simplicity, the government pursues an exogenous policy. The net monetary transfer in each period is equal to the change in money supply, $T_t = M_t^s - M_{t-1}^s$, where the log of money supply evolves exogenously, according to $m_t = m_{t-1} + \mu_t$, $\mu_t \overset{i.i.d.}{\sim} h_{\mu}$.

Market clearing requires $C_t = Y_t$, $C_t(i) = Y_t(i)\ \forall i$, $H_t = \int_0^1 H_t(i)\,di$, $M_t = M_t^s$, $B_t = 0$.

Full Information Solution Substituting the household's optimality conditions and market clearing in the firm's profit function, profit in units of marginal utility becomes

$$\pi_t(i) = P_t(i)^{1-\varepsilon} A_t(i)^{\varepsilon-1} P_t^{\varepsilon-1} Y_t^{-\sigma} - P_t(i)^{-\varepsilon\gamma(\nu+1)} A_t(i)^{\varepsilon\gamma(\nu+1)} P_t^{\varepsilon\gamma(\nu+1)} Y_t^{\gamma(\nu+1)}.$$

The first order condition with respect to $P_t(i)$ yields the flexible price solution. The full-information optimal log-price, denoted by $x_t(i)$, is a linear combination of all the shocks in the economy:

$$x_t(i) = m_t + z_t(i) + \zeta_t(i) - \log(Y^*),$$

[2] I use lower-case letters to denote logs of upper-case variables.

where the natural rate of output Y^* is given by

$$Y^* = \left[\frac{\varepsilon - 1}{\varepsilon\gamma(\nu+1)}\right]^{\frac{1}{\gamma(\nu+1)+\sigma-1}}, \quad \forall t.$$

Partial Equilibrium I assume that all aggregate variables evolve according to the flexible price, full information equilibrium. A set of firms of measure zero are information-constrained. Substituting the full-information equilibrium outcomes, the profit of a constrained firm is proportional to $\pi(p_t(i) - x_t(i))$[3], where $p_t(i)$ is the log-price charged by the information-constrained firm, $x_t(i)$ is the optimal full-information log-price and

$$\pi(p-x) = e^{(1-\varepsilon)(p-x)} - \frac{\varepsilon - 1}{\varepsilon\gamma(\nu+1)} e^{-\varepsilon\gamma(\nu+1)(p-x)}.$$

This equation defines the profit function introduced in the setup of the model. Note that the profit function is maximized at $p_t(i) = x_t(i)$, hence $x_t(i)$ is also the current profit-maximizing price for the information-constrained firm in the static problem, excluding information costs. Therefore, the rationally inattentive firm would like to set a price that is as close as possible to the target full-information price, $x_t(i)$, subject to the costs of acquiring information about the evolution of this target.

The mapping into the notation used in the main body of the paper is $\widetilde{x}_t(i) \equiv m_t + z_t(i)$, $\widetilde{v}_t(i) \equiv \mu_t + \xi_t(i)$, and $v_t(i) \equiv \zeta_t(i)$. The normalized target prices in the stationary formulation, τ periods after a review has occurred, are $\widetilde{y}_0(i) = 0$, $\widetilde{y}_\tau(i) = \sum_{j=1}^{\tau}(\mu_j + \xi_j(i))$, and $y_\tau(i) \equiv \widetilde{y}_\tau(i) + \zeta_\tau(i)$. Finally, conditional on a review in period t, the information-constrained price in period $t+\tau$ is $p_{t+\tau}(i) = \widetilde{x}_t(i) + q_\tau(i)$. The per-period profit function $\pi(p_t(i) - x_t(i))$ is replaced by $\pi(q_\tau(i) - y_\tau(i))$, a function of the normalized price and the normalized state.

[3] I omit a term that does not affect optimization.