Essays on Price Adjustment and Imperfect Information

Luminita Stevens

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

Columbia University

2012

ABSTRACT

Essays on Price Adjustment and Imperfect Information

Luminita Stevens

Understanding how firms set prices is a key step towards settling classic debates in economics regarding the sources of nominal price rigidities, the mechanisms through which disturbances are transmitted within and across countries, and the effectiveness of monetary policy in dampening business cycle fluctuations. This dissertation examines patterns of price adjustment at the firm level, both empirically and theoretically. The first chapter studies pricing patterns in US grocery store data. Using a novel empirical method, I identify changes in the distribution of product-level prices over time. These changes typically occur every seven months and mark the transition to new pricing regimes. Inside regimes, prices alternate among a small set of prices with high frequency. This evidence motivates a theory of price setting in which firms respond to shocks using multiple-price policies that are simple enough to only specify a small number of prices, and that are updated only on discrete occasions. The second chapter presents a theory of costly information that generates such simple, sticky policies. In order to economize on the costs of acquiring information, the firm designs a pricing policy that is a noisy, coarse representation of market conditions. Moreover, it updates this policy infrequently, based on imprecise signals about the state of the economy. Despite the high volatility of observed prices, the firm responds imperfectly to changes in market conditions. The third chapter, co-authored with Ryan Chahrour, addresses the patterns of adjustment in international relative prices. We develop a two-country model in which retailers have imperfect information and search for producers operating in different regions in the two countries. We demonstrate that frictions at the regional level within countries generate dispersion in international relative prices in the absence of additional frictions at the national border.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am grateful to numerous professors, colleagues and friends for their support. My deepest gratitude goes to my advisor, Mike Woodford. His exceptional insight and generosity have had a profound impact on my work and on my development as an economist. I thank Ricardo Reis for his contagious energy and lessons on how to conduct and present research. I owe special thanks to Jaromir Nosal for his unwavering support and for his steady stream of practical advice.

Thank you to Bruce Preston for his encouragement and friendship and to Steve Zeldes for his piercing questions and insightful comments. Whether they know it or not, Stefania Albanesi and Stephanie Schmitt-Grohé have been inspiring role models. Don Davis deserves the award for biggest impact per word spoken. His pithy words of wisdom will resonate long after I have left Columbia University. My special thanks to Alexei Onatski for his support, especially during the early part of my PhD - without him I may never have dared knock on Mike's door. I also owe a special thank you to Christian Hellwig for everything he has taught me. I am grateful for all the discussions and debates with Edmund Phelps, who pushed me to look at issues from many different angles. I wish to thank Joe Stiglitz, who sparked my interest in problems of imperfect information.

I am indebted to the wonderful faculty at Columbia University, especially Emi Nakamura, Jòn Steinsson, Eric Verhoogen, Bernard Salanié, David Weinstein, Marcelo Moreira, Serena Ng, Martin Uribe, Xavier Sala-i-Martin, Paolo Siconolfi, and Susan Elmes, each of whom has helped to make me a better economist. Thank you to all my undergraduate professors, especially to Nancy Marion for encouraging me to pursue a PhD in economics and to Amitabh Chandra for making me love econometrics, not just macroeconomics.

I wish to thank my colleagues, especially Ryan Chahrour, Cyntia Azevedo, Ozge Akinci, Maria José Prados, Neil Mehrotra and Dmitriy Sergeyev for their ideas, their laughs, and their encouragement. A special thanks to Heriberto Tapia for always being there for me, and to Jody Johnson, our program coordinator, for helping to keep the ball rolling.

Thank you to my family: for all that is, was, and will be.

To David

Introduction

This dissertation examines patterns of price volatility in product-level data, both empirically and theoretically. It contributes to two important questions in the monetary and international macroeconomics literature. First, do prices respond quickly and accurately to changes in market conditions, as in the benchmark flexible-price model, or are they somewhat rigid, as in the benchmark sticky-price model? Second, why do the prices of identical products differ across regions and countries?

Price stickiness is frequently used as a key ingredient in macroeconomic models to generate real effects of monetary policy. The question of how sticky prices should be assumed to be in these models has been at the heart of a long-standing debate in monetary economics, since it has direct implications for the role of monetary policy and its effectiveness in dampening business cycle fluctuations. My approach to this question builds on the recent literature, which has shifted away from analyzing aggregate data, to focus on the properties of prices at the disaggregated, product level.

The current challenge is that the data is not consistent with either of the two benchmark models of price setting. First, consumer prices exhibit high volatility, even under relatively stable macroeconomic conditions. Starting with the seminal paper of Bils and Klenow (2004), recent empirical work[1] has found that the prices of individual products change much more frequently than is typically assumed in standard models with nominal rigidities. Using monthly data, Bils and Klenow (2004) show that prices in the U.S. typically change every four months.[2] Conversely, in order to generate real effects of monetary policy of the magnitude estimated using aggregate data, the typical sticky

[1] Klenow and Malin (2010) provides a comprehensive review of a large body of recent empirical research studying prices in disaggregated data, both in the U.S. and internationally.

[2] Studies at higher frequencies show even larger volatility. For example, using Japanese daily scanner data, over the period 1988 to 2005, Abe and Tonogi (2010) estimate that on average, prices change every three days.

price model assumes that prices remain unchanged for as long as one year. Hence, this evidence challenges traditional sticky price models (e.g., those following Calvo, 1983, or Sheshinski and Weiss, 1977). Moreover, to the extent that all these price changes are responses to changes in market conditions, such high volatility implies fast adjustment, and it suggests that price stickiness is an insufficient force for delivering meaningful real effects of monetary policy.

Based solely on the frequency with which prices change, the empirical evidence would seem to favor the flexible price benchmark. However, starting with Nakamura and Steinsson (2008) recent work has also shown that as much as half of this price volatility is transitory in nature. Importantly, the pattern of prices repeatedly returning to past levels is at odds with both the sticky price and the flexible price models. In both of these models, every price change is the result of a re-optimization, hence there is no reason for past prices to be revisited. Proponents of sticky price models have argued that by eliminating the transitory price changes one can recover the traditional price stickiness in the form of sticky *regular* prices, which may be more relevant in the aggregate. A growing empirical literature - including work by Nakamura and Steinsson (2008), Eichenbaum, Jaimovich and Rebelo (2011), and Kehoe and Midrigan (2010) - has sought to quantity the apparent rigidity underlying the high frequency of adjustment by using different *filters* to identify transitory volatility, and to characterizes the properties of apparently sticky regular prices versus seemingly flexible transitory prices. In parallel, the theoretical literature (for example, the models of Kehoe and Midrigan, 2010, and Guimaraes and Sheedy, 2011) has sought to build models that give firms the incentive to temporarily deviate from rigid price levels, by modeling different constraints for the firm's ability to change regular versus temporary prices. At present, how these transitory price changes should be analyzed empirically, how they should be generated theoretically, and what they imply for the degree of nominal price rigidity remain open questions in the literature.

The first two chapters of this dissertation propose a framework that can simultaneously account for both the high frequency of price adjustment and the rigidity of certain price levels. In this framework, both regular and transitory prices are part of an integrated pricing policy that is infrequently updated. There are no a priori differences between regular and transitory prices, as they are all chosen to be jointly optimal. The first chapter identifies and characterizes such pricing policies in grocery store product-level data, while the second chapter develops a theory that can account for the empirical findings document in the first chapter.

The first chapter employs a novel statistical method that builds on the Kolmogorov-Smirnov test. Specifically, I look for breaks in individual price series, by testing for changes in the distribution of prices over time. Rather than focusing on the properties of regular versus transitory prices, I focus on the properties of pricing regimes, where each regime consists of a distribution of prices.

I find that pricing regimes typically last seven months, a long time relative to the frequency of individual price changes, which in grocery store data is less than one month. Approximately three quarters of products contain regimes in which a small number of prices (typically four) are revisited over the life of the regime. Approximately one quarter of the product-level series consist of regimes in which prices either do not change at all or change very rarely. While the pattern of single sticky prices can be accommodated by existing sticky price theories, the pattern of regimes with multiple rigid prices is inconsistent with existing theories of price-setting. This evidence suggests a new theory of price setting, in which each firm chooses a rigid pricing policy that is sticky and simple: it is updated relatively infrequently, and it only consists of a small number of distinct prices.

The second chapter develops a theory of price setting based on costly information that generates such simple, sticky pricing policies. I present the dynamic price-setting problem of a rationally inattentive firm that cannot observe market conditions for free, and whose acquisition of information is subject to both fixed and variable costs. If it pays a high fixed cost, the firm can obtain extensive information about the state of the world and redesign its policy. In each period between policy reviews, the firm can monitor market conditions, subject to an additional variable cost of information. The firm uses this information to decide which price to charge in each period from the current policy, and also to decide if its current policy has become obsolete, relative to market conditions, such that a new policy needs to be chosen. Both the stickiness and the coarseness of the pricing policy are a result of the firm's need to economize on information costs.

Because it chooses a coarse pricing policy, prices are weakly tied to market conditions. Nevertheless, prices change frequently, and often by large amounts, thereby endogenously generating transitory volatility. The theory breaks the tight link between frequency of price adjustment and responsiveness to disturbances that exists in other models of price setting. Consequently, it has the potential to reconcile conflicting evidence regarding the apparent flexibility of prices at the micro level and the non-neutrality of monetary policy. In this model, rigidity arises not because individual prices change infrequently, but because prices are always drawn from a small set over the life of the policy, and are chosen based on imperfect information about market conditions. The theory brings together different features of the growing literature on imperfect information in price setting. However, it departs from existing work by generating simple pricing policies that consist of a small set of prices. The chapter also provides a novel solution method for problems with rationally inattentive agents.

The third chapter, coauthored with Ryan Chahrour, addresses the literature on deviations from the law of one price in open economies. We propose a model of imperfect information in which buyers, rather than sellers, must pay a cost to acquire information. We develop a model of equilibrium price dispersion via retailer search, and we target well-known empirical facts about the failure of the law of one price across countries. In our model, retailers engage in costly sequential search for the best price among producers in the economy. Retailers know only the distribution of prices, and search in a world of two countries, each with two regions. Segmentation across regions and countries is determined by the extent to which retailers located in a particular region are more likely to sample prices posted by producers located in their home region or home country.

In contrast to recent work, which has underscored segmentation across countries, we show that our model can match the empirical evidence on cross-country price differences with regional segmentation alone. Our findings hold qualitatively whenever there are international differences in the realizations of aggregate shocks or differences in the structural parameters, namely, when the two countries are simply different and not necessarily segmented from each other. This finding implies that countries may not be as segmented at the border as previously thought, which in turn has implications for trade policy. Using data simulated from the model, we also demonstrate some of the difficulties in using popular regression measures of the border cost to infer the degree of market segmentation.

1 Pricing Regimes in Disaggregated Data

1.1 Introduction

This chapter introduces a new statistical method to characterize patterns in product-level prices by identifying breaks in individual price series. I propose to test for changes in the distribution of prices over time using a method that is based on the Kolmogorov-Smirnov test, which determines if two samples are likely to have been drawn from the same distribution. Building on tests that estimate the location of a single break in a series (Carlstein, 1988 and Deshayes and Picard, 1986), I adapt the test to identify an unknown number of breaks at unknown locations. I use simulations to determine the appropriate critical value. An advantage of this approach is its generality, and hence robustness across different data generating processes. This feature is important given the high degree of heterogeneity documented in product-level price series. Moreover, in contrast to recently developed empirical methods, the test sidesteps the need for a priori definitions of transitory versus permanent price changes. This is a desirable feature, since, as will be shown below, important statistics, crucial for informing models of pricing, vary significantly depending on the definition of transitory price changes and the filter implemented to identify such price changes.

I first demonstrate the robustness of the method in simulations. The break test correctly rejects the null of no break 91% of the time across a mixture of different data generating processes; it yields false positives less than 2% of the time. Upon rejection of the null, it finds the true location of the break exactly 94% of the time; in the remaining cases, it is off by two periods, on average. The break test is less likely to reject the null of no break when one of the two samples to be compared is particularly short, thereby providing a less precise estimate of the true distribution. If I restrict the simulated series to contain regimes of at least five periods, the test finds virtually all the breaks.

I apply the break test to weekly prices at the barcode level from Dominick's Finer

Foods, a chain of grocery stores operating in the Chicago area. I find that breaks reflect changes in the set of prices charged. I call such sets "pricing regimes." Pricing regimes typically consist of a small set of prices relative to both the duration of regimes and the frequency of price changes inside regimes. For approximately 90% of all regimes, five or fewer unique price quotes account for more than 90% of the prices inside the regime. The typical regime lasts 31 weeks and contains only four distinct prices despite the fact that in this data prices change at least every four weeks. Figure 1 plots the weekly price of frozen juice, illustrating the pricing regimes identified by the test: within regimes, prices change frequently and by large amounts, yet they alternate among a small set of distinct values.

Based on the finding that regimes typically consist of a small set of prices, I next categorize products in terms of the rigidity of the set of prices observed in each regime. I find that only 5% of all products consist entirely of single-price regimes that would be consistent with sticky price models such as time-dependent models using Taylor (1980) or Calvo (1983) staggered price-setting, or state-dependent models (Sheshinski and Weiss, 1977, Golosov and Lucas, 2007). Regimes last a long time for these products (the median implied regime duration is 45 weeks), and volatility across regimes is low relative to the average (the average price change across regimes for this group of products has a median of less than 6%). These statistics suggest that these products face fairly low volatility in their desired price.

I also uncover evidence against the *one-to-flex* hypothesis that price series are characterized by flexible deviations from a rigid mode. This pattern has been generated in the recent theoretical work of Kehoe and Midrigan (2010) and Guimaraes and Sheedy (2010). I find that 18% of product series are characterized by regimes in which a single rigid price is revisited over the life of the regime. However, these one-to-flex series are not nearly as volatile as the overall sample: for products in this category, 50% of the regimes contain a single price, and the remainder typically only exhibit three price changes. Moreover, the average price change within regimes is typically less than 6%. Hence, large transitory price changes do not appear to be an important part of the pricing policy for these products.

The volatility of the dataset is concentrated in products for which rigidity extends to the *set* of price charged. Specifically, 77% of products contain regimes in which at least two distinct prices are revisited over the life of the regime, as illustrated in figure 1. For these products, the typical regime lasts 31 weeks, and it contains four distinct prices. These products are characterized by high within-regime volatility: prices inside regimes change with a frequency of 29%, and the average size of price changes within regimes is 11%.

Next, I investigate the extent to which the structure of regimes in the data is consistent with the notion of prices being drawn from the same distribution within a regime. I compare the data to artificial series generated in two ways, replacing the observed realizations of prices inside each regime with (1) i.i.d. draws from the realized distribution within each regime, and (2) draws in which all within-regime price changes are clustered at the edges of regimes. Each of these two simulations implies a different approach to modeling product-level price volatility. If different prices are equally likely to be observed over the life of a regime (as in the first simulation), then it would point to a theory of pricing in which the firm chooses a single multiple-price policy that applies over the life of the regime. At the other extreme, if the volatility of prices is confined to relatively short *transitional* periods between single-price regimes (as in the second simulation), then it would suggest a theory of pricing that adapts the existing single-price theories to include a period of experimentation before the firm settles on a new price. I re-run the break test on the artificial series generated in this way, and then compare the location of breaks and the properties of the regimes found in the data with those found in the artificial series. I find strong synchronization between the data and the artificial i.i.d. regimes. Conversely, I find weak evidence in support of the clustering of price changes at the edges of the regimes. Hence, I conclude that the data is best described by regimes in which the volatility of prices is *interior* rather than *transitional*.

This pattern of infrequently updated regimes that consist of a discrete set of prices is difficult to reconcile with most models of price-setting, which cannot generate discreteness in the set of prices charged, unless the underlying shocks are themselves drawn from distributions with mass points. Alternatively, these findings suggest a theory in which the firm chooses a policy that consists of set of discrete prices among which it alternates over the life of the policy, and which it updates relatively infrequently. Such a theory is proposed in the next chapter, where both the infrequent regime changes and the discreteness of prices within regimes are due to imperfect information.

Finally, I compare the break test with three popular filtering methods: a v-shaped sales filter similar to those employed by Nakamura and Steinsson (2008), the reference price filter of Eichenbaum, Jaimovich and Rebelo (2011), and the rolling mode filter proposed by Kehoe and Midrigan (2010). These filters have been proposed as a way to uncover stickiness in product-level pricing data once one filters out transitory price changes that may be orthogonal to aggregate conditions and hence not of direct interest to macroeconomists. One advantage of the break test relative to these 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 regimes. In contrast, v-shaped filters, such as those proposed by Nakamura and Steinsson (2008), identify breaks based on changes in the maximum price, while the reference price/rolling 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: firstly, a non-parametric method is likely to outperform the parameterized filters in data that appears to be characterized by random variation in the duration of both regular and transitory prices; secondly, a method that uses information about the entire distribution of prices is likely to have more accuracy in detecting the *timing* of breaks compared with a method 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-regime 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 within regimes are sensitive to the estimated location of breaks.

Since the break test is robust across different data generating processes, it provides a direct test of the assumptions made by the filters. To the extent that the true data generating process is indeed consistent with the assumptions made by a particular filter, one would expect to find a high degree of alignment between results obtained by the break test and those obtained by the filter. Hence, I apply all methods to price series from Dominick's data. I find that the v-shaped filter is highly synchronized with the break test in terms of the *timing* of breaks; however, it finds many more breaks in the data, reflecting the fact that the maximum price per regime does not always coincide with the mode. Moreover, adjusting the parameters of the v-shaped filter to reduce the number of breaks (and hence increase the duration of regimes) substantially reduces the synchronization between the two methods. Therefore, it is not the case that the break test identifies the same breaks as a v-shaped filter with a given set of parameters. Alternatively, the reference filter is hampered by the use of a fixed window, which essentially assumes away the question of identifying the timing of breaks. Given the heterogeneity of regime lengths identified using the break test, the estimated change points of the two approaches are found to coincide largely by chance. The Kehoe and Midrigan (2010) rolling mode filter is the filter most closely aligned with the break test: when parameterized to match the frequency of breaks identified by the break test, it has higher exact synchronization than the reference price filter, and differences in the timing of breaks are quite small (two periods on average).

Section 1.2 details the statistical method employed to identify multiple potential break points in a given series, building on the basic two-sample Kolmogorov-Smirnov test. Section 1.3 presents the properties of pricing regimes identified by the break test in the Dominick's data. Section 1.4 compares the performance of the break test to that of three different filtering methods. Section 1.5 concludes.

1.2 The Break Test

The *break test* uses the Kolmogorov-Smirnov distance between two sample distributions[3]. In order to identify multiple breaks at unknown locations, I build on tests that estimate the location of a single break in a series, specifically, those proposed by Carlstein (1988) and Deshayes and Picard (1986). I implement an iterative procedure that identifies breaks sequentially: first, I test the null hypothesis of no break in a given sample; upon rejection, I estimate the location of the break. For series that have more than one break, iterations of the test on the resulting sub-samples identify all change points. This approach is similar to that proposed by Bai and Perron (1998) for sequentially estimating multiple breaks in a linear regression model.

To my knowledge, the existing literature on estimating breaks using Kolmogorov-Smirnov focuses on the identification of a single break. Moreover, derivations of the exact critical values for discrete distributions are restricted to the one-sample goodness of fit Kolmogorov-Smirnov test.[4] Finally, I wish to apply the test to potentially discrete data: Nakamura and Steinsson (2008) document the importance of rigid *regular* prices in the micro data underlying the US CPI, and Klenow and Malin (2010) document the

[3]See, for example, Brodsky and Darkhovsky (1993), who provide a discussion of the Kolmogorov-Smirnov test and other non-parametric change point methods.

[4]See Conover (1972) and Wood and Altavela (1978).

"disproportionate importance" of a few price levels over the life of a price series. Hence, based on results from the existing empirical literature on pricing patterns, we can expect that many regimes will contain at least one mass point. Using critical values derived for continuous distributions would render the test conservative. For these reasons, I determine the critical value using simulated data.

The first part of this section defines the test statistic used to test the null hypothesis of no break on a sample of size n, S_n, and, upon rejection, the statistic used to estimate the location of a break, τ_n. The second subsection determines the appropriate critical value for S_n, using simulations of different processes that generate patterns similar to those expected to approximate product-level price series. The last subsection evaluates the overall performance of the test in simulated data, namely its ability to correctly reject the null and to correctly identify the timing of a break given a rejection (the joint performance of the statistics S_n and τ_n).

1.2.1 Test Statistic

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

$$D_n(t) \equiv \sup_p |\widehat{F}_1^t(p) - \widehat{G}_{t+1}^n(p)| \tag{1.1}$$

where $\widehat{F}_1^t$ and $\widehat{G}_{t+1}^n$ are the empirical cumulative distribution functions of the two sub-samples:

$$\widehat{F}_1^t(p) \equiv \frac{1}{t}\sum_{s=1}^{t} \mathbf{1}_{\{p_s^n \leq p\}} \quad \text{and} \quad \widehat{G}_{t+1}^n(p) \equiv \frac{1}{n-t}\sum_{s=t+1}^{n} \mathbf{1}_{\{p_s^n \leq p\}} \tag{1.2}$$

Figure 2 plots the CDFs of two consecutive sub-samples of the price series plotted in figure 1. The distance between the two empirical CDFs is zero at the edges of the $[0, 1]$ interval; under the null hypothesis of no break, it has only random sampling variation in between, and hence the Kolmogorov-Smirnov distance, $D_n(t)$, is not expected to be too large.

I collect the Kolmogorov-Smirnov distances for all potential break points $t \in T_n$, and I define the test statistic for a sample of size n, S_n, following Deshayes and Picard (1986):

$$S_n \equiv \sqrt{n} \max_{t \in T_n} \left[\frac{t}{n} \left(\frac{n-t}{n} \right) D_n(t) \right] \tag{1.3}$$

The normalization factor depends on the sample size and 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 particularly short relative to the other sub-sample, and thus provides 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 in the next subsection), the next step is to estimate the location of the break. The change point estimate τ_n is given by:

$$\tau_n \equiv \arg\max_{t \in T_n} \sqrt{\frac{t(n-t)}{n}} D_n(t) \tag{1.4}$$

Carlstein (1988) provides the strong consistency proof for τ_n/n in the case of independent observations.

To apply this to series that may have multiple breaks at unknown locations, I first test for the existence of a break and estimate its location, following the method described above. I then apply the same process to each of the two resulting sub-series, or branches, and continue until there are no more S_k statistics greater than the critical value.

1.2.2 Critical Value

In theory, the identification method presented above is robust to a wide variety of data generating processes. 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. With this in mind, 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.

It is important to note that given the iterative nature of the test and the fact that at each step, the test statistic is a measure of the *maximum* distance between the two sample CDFs (equation (1.3)), the critical value determines only how soon the algorithm stops in its search for breaks: across different potential critical values, the *path* taken by the test will be the same, stopping sooner given higher critical values, and continuing to shorter sub-samples for lower critical values. Hence varying the critical value iteratively adds new regimes inside the existing regimes, without affecting the location of the existing breaks, which makes it easier to interpret changes in the pricing statistics resulting from variations in the parameter of the test. This gives the break test a degree of robustness relative to existing filtering methods, which do not have this "nesting" property when varying the parameters of each filter.

Simulated Processes

I simulate processes that represent both recent theoretical models of price-setting and the most commonly observed price patterns in micro data. Specifically, I assume that the data is a mixture of the following four processes: 1) sequences of infrequently updated *single prices*, such as those generated by Calvo pricing or simple menu cost models;

2) sequences of *one-to-flex plans*, defined as sticky prices accompanied by frequent, flexible deviations from these rigid modes, which are consistent with the assumptions of a reference price/rolling mode filter and with the price-setting model of Kehoe and Midrigan (2010); 3) sequences of *downward-flex plans*, which consist of sticky prices accompanied by frequent, flexible downward deviations, and which are consistent with the assumptions of a v-shaped sales filter and with models such as the dynamic version of Guimaraes and Sheedy (2011); and 4) sequences similar to the frozen juice series shown in the introduction, labeled *rigid multiple-price plans*, each consisting of a small number of distinct prices that are revisited over the life of the pricing plan, which are consistent with the theory proposed in the next chapter. For simplicity, I assume that the data is an equally-weighted mixture of these processes. Figure 3 shows sample series with multiple regime breaks for each of the four simulated processes.

For process I, the simulated price series is given by:

$$p_{t+1} = B_{t+1} \exp\{\varepsilon_{t+1}\} + (1 - B_{t+1})\, p_t \tag{1.5}$$

where the sequence $\{B_t\}$ is a Bernoulli trial with probability of success β marking the transition to a new price level, and ε_t is a normally distributed i.i.d. innovation. This series also corresponds to the regular price series, p^R_{t+1}, for the multiple-price processes II, III and IV. In each case, the transition to a new regime is marked by $B_t = 1$.

For process II, the simulated price series is given by:

$$p_{t+1} = B_{t+1} \exp\{\varepsilon_{t+1}\} + (1 - B_{t+1}) \left[D_{t+1} p^R_t \exp\left(\varepsilon^T_{t+1}\right) + (1 - D_{t+1})\, p^R_t\right] \tag{1.6}$$

where the sequence $\{D_t\}$ is a Bernoulli trial with probability of success δ, marking the transition to a new transitory price, which is given by an i.i.d. innovation, ε^T_t. I assume that in each period, either a regular price change *or* a transitory price change can occur, as in the model of Kehoe and Midrigan (2010). For process II, I further assume that the mean of the transitory innovations is zero.

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, I 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 (1.6) inside each regime to three bins, each corresponding to a unique price quote, such that each regime consists of only three distinct prices. Table 1 summarizes the specification of the four processes.

The processes used to determine the critical value are parameterized to roughly match the volatility of the Dominick's dataset. I target a mean absolute size of price changes of $10 - 12\%$, with a frequency of price changes of $18 - 20\%$. Prices in the single sticky price process change with a frequency of 3%. However, the performance of the test is robust to variations in volatility.

Break Test Critical Values

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 $t \in T_n$,

$$S_n \to \widetilde{K} \equiv \sup_{u\in[0,1]} \sup_{v\in[0,1]} |B(u,v)| \tag{1.7}$$

where $B(\cdot,\cdot)$ is the two-dimensional Brownian bridge on $[0,1]$.[5] Hence equation (1.7) provides asymptotic critical values for the test of a single break at an unknown location:

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

the null is rejected at level α if $S_n > \widetilde{K}_\alpha$, where $\widetilde{K}_\alpha$ is found from $\mathbf{P}(\widetilde{K} \leq \widetilde{K}_\alpha) = 1 - \alpha$. Deshayes and Picard (1986) provide tables for the single-break test critical values, and they find that these values provide a reasonable approximation for sample sizes of at least 50 observations.

I start from the critical values provided by Deshayes and Picard (1986), and, as noted above, I determine the appropriate critical value using simulations in which I compare the results of the break test with the true location of breaks. For simplicity, I use a single critical value across all sample sizes. The critical value is determined using two statistics: *good_reject* and *bad_reject.* The statistic *good_reject* counts the number of times that the test rejects the null of no change point on a sample that does in fact contain a break; it is reported as a fraction of the number of true breaks in the simulation. Obtaining a low value for *good_reject* implies that the test is not sensitive enough, such that many breaks are not identified. Correcting this requires reducing the critical value used. Conversely, *bad_reject* counts the number of times the test rejects the null of no change point on a sample that does not contain a break; it is reported as a fraction of the number of breaks estimated by the test. A high *bad_reject* implies that the test yields too many false positives, hence the critical value needs to be increased.

The first finding is that the asymptotic critical values provided by Deshayes and Picard are too conservative. This is a result of the discrete nature of the data: given two pairs of samples of equal size, the pair drawn from discrete distributions contains more information about the population CDFs compared with the pair drawn from continuous distributions. I gradually reduce the critical value, balancing the values of *good_reject* and *bad_reject.*

The tension in choosing the appropriate critical value is between process I and the multiple-price processes: for the single sticky price process, a positive critical value arbitrarily close to zero ensures that all the breaks are perfectly identified, since in this case the test statistic S_k is either zero in the case of no change, or greater than zero when there is a change. But using too low of a critical value generates false positives in the multiple-price processes: for process III, in the case when transitory prices can last more than one period, too low of a critical value leads to mislabeling such transitory prices as new regimes; similarly, in process IV, rigidity in all price levels can generate within-regime sequences that look like sequences of single sticky prices but are in fact part of the same regime.

Using the critical value associated with the asymptotic 1% significance level (0.874) provided by Deshayes and Picard (1986) for continuously distributed data, the break test correctly finds 84% of the simulated breaks. The test fails to identify relatively short regimes, underestimating the average regime length by seven periods. Reducing the critical value improves the test's performance: the critical value associated with the asymptotic 5% significance level (0.772) correctly finds nearly 87% of the simulated breaks, with only a marginal increase in the fraction of false positives. The test correctly identifies an increasing fraction of the short regimes in the simulation. I gradually reduce the critical value until the test finds at least 90% of the simulated breaks, but stop before *bad_reject* exceeds 5% for *any* of the simulated processes.

Table 2 reports average statistics across all processes: $K = 0.61$ (in bold) is the lowest critical value for which the average *good_reject* rate exceeds 90%, while keeping the maximum false positive rate for *all* four data generating processes below 5%. Further reducing the critical value generates false positives at an increasing rate, relative to the gain in *good_reject*. For instance, increasing the power to 95% also increases the incidence of false positives to 12%.

At the chosen critical value, the average length of the regimes identified by the break test is longer than the true average length by three periods, reflecting the weak power in identifying regimes that last between two and four periods. Restricting the simulations to regimes lasting at least five periods would ensure the identification of virtually all regimes. It is important to note that the critical value only determines how soon the algorithm stops in its search of break points. As the critical value is reduced, the test finds additional breaks that are added to the set of breaks found by the test with a higher critical value, such that for two critical values $K_2 > K_1$, the corresponding sets of estimated break points satisfy $T_2 \subset T_1$.

1.2.3 Performance

The strength of the break test depends on its ability to correctly reject the null of no break and to correctly identify the timing of a change given a rejection (the joint performance of the statistics S_k and τ_k). As shown in the previous subsection, the break test correctly rejects the null of no break 91% of the time. As a result, it overestimates the average regime length by three periods.

Upon rejection of the null, I find that the change point estimate τ_k coincides exactly with the true change point 94% of the time, and it is otherwise off by two periods, on average. Table 3 shows the performance of the test for each of the four processes, for the critical value selected in the previous subsection: for single sticky price series (process I), 100% of the breaks found by the test are exactly synchronized with the simulated breaks; for the multi-price processes, accuracy is lower, with $91 - 94\%$ of the breaks exactly synchronized with the simulated breaks.

In the case of the *one-to-flex* process III, if transitory sales were restricted to last only one period, false positives would decline to less than 1%, exact synchronization would improve to 95% from 91%, and the mean distance between the estimated breaks and the true breaks among non-synchronized breaks would decline to 1.5 periods.

These statistics are almost entirely driven by the presence of very short regimes in the simulations. For example, for process II, the test identifies virtually none of the regimes that last only two periods; it finds 65% of the regimes that last four periods, and 94% of the regimes that last eight periods. The presence of weakly identified short regimes in turn affects the timing of breaks, reducing the exact synchronization between the estimated breaks and the true breaks. Constraining the simulations to generate only regimes of at least *five* periods results in the identification of virtually all regimes (thus eliminating the bias in the estimated average regime length), and further improves the degree of exact synchronization between the test breaks and the simulated breaks. Hence, to the extent that regimes in actual pricing data last at least four-to-five weeks, and to the extent that the data is well approximated by the processes simulated above, we can expect the break test to perform even better than shown in table 2.

1.3 Regimes in the Data

This section documents a set of new statistics on price adjustment that are computed at the regime level once the break test is applied to product-level price series.

1.3.1 Description of the Data

I apply the break test to price series from Dominick's Finer Foods. Dominick's is a large chain of grocery stores operating in the Chicago area, whose pricing practices are representative of many large US grocery chains. The dataset was built as part of a series of randomized pricing experiments conducted by Dominick's in cooperation with the University of Chicago's Graduate School of Business, and is available online.[6]

The data include weekly prices from September 1989 until May 1997, at 86 store locations, for thousands of barcode-level products, in 29 categories, including various household supplies and non-perishable food items. The prices are the actual transaction prices as recorded by the stores' scanners. Many price series have missing observations. I use only price series that have at least 52 observations and on average at most one missing observation per year. After cleaning the data in this way, I further restrict the sample to the store with the largest number of observations from the group of stores

[6] See Hoch et al (1994) for a description of the experiments.

whose prices were not randomized. Preliminary work suggests virtually identical results across the different stores in the control group, reflecting the fact that Dominick's policy is to set prices at the chain level, such that price changes are very strongly correlated across stores. The final sample contains more than 700,000 observations for 4,275 unique universal product codes (UPCs), in 28 product categories.

This dataset has been used quite extensively in work on price-setting, including in papers by Dutta, Bergen and Levy (2002), Chevalier, Kashyap and Rossi (2003), Burstein and Hellwig (2007), Midrigan (2009), and Kehoe and Midrigan (2010), among others. The wide use of this dataset, despite its fairly narrow product coverage, can be attributed to at least three characteristics. First, Dominick's policy is to change prices only once a week (on Wednesdays), hence we are not missing any intra-week price changes. In contrast, other datasets are built by sampling at some fixed intervals, such as the monthly-sampled BLS data. Second, the dataset has relatively long time series per product compared with other datasets: in my sample, median (average) length of the series is 142 (177) weeks per UPC and the maximum is 400 weeks. While the issue of truncated series is still important, the qualitative implications of my findings are not affected when statistics of interest are computed using interior regimes only. Finally, another advantage of this data is the fact that it contains products whose prices are highly volatile and exhibit precisely the sharp, yet transitory, price swings that have recently come to the forefront of the price dynamics literature. The median implied price duration[7] across all products in the sample is less than four weeks; the median average size of non-zero price changes is 11.9% in absolute value; and the standard deviation of non-zero price changes is 16.3%.

Table 4 documents summary statistics across the different product categories. As is typical with pricing data, there is a significant amount of heterogeneity even within this

[7] I follow the convention in the literature and compute the frequency of regime changes. Assuming a constant hazard rate, the implied duration is computed as $d = -1/\ln(1 - freq)$.

relatively narrowly defined group of products. The implied duration of prices ranges from less than two weeks for soft drinks to more than 17 weeks for cigarettes, with most products in the range of two to seven weeks. Excluding cigarettes, the size of price changes ranges from 6% for canned tuna to 22% for frozen entrees. Additionally, there is a positive correlation between the size and frequency of price changes.

1.3.2 Regime Changes versus Price Changes

Regime changes are estimated to occur infrequently. While the implied price duration for the median product is less than four weeks, the median implied duration of regimes is 31 weeks. There is considerable underlying heterogeneity, both within and across categories, with most regimes lasting anywhere between four months and one year and a half. Figure 4 shows the median implied regime duration for each category, ordered from lowest to highest, as well as the interquartile range.

In principle, the break test can identify any salient changes in both the support and the shape of the distribution of prices over time. In practice, I find that most breaks reflect changes in the set of prices charged: in 50% of consecutive regime pairs, there is no overlap between the sets of prices charged; in 82% of consecutive pairs, at most one price occurs in both regimes; in 94% of pairs, at most two prices occur in both regimes; and in 99% of regimes, at most three prices occur in both regimes.

In terms of within-regime volatility, 18.9% of regimes are single-price regimes, while the remaining 81.1% of regimes contain at least two distinct prices. Despite their relatively long duration, regimes are characterized by a small number of distinct prices: among regimes that contain at least two prices, the median (mean) number of distinct prices per regime is 5 (6), with an interquartile range of $[3, 8]$. Moreover, for approximately 90% of all regimes, five or fewer unique price quotes account for more than 90% of the regime. Nevertheless, inside these regimes, prices change often and by large amounts. The median weekly frequency of within-regime price changes is 29%, and the median size of price changes is 11%. The frequency of price changes per regime is large relative to both the duration of regimes and the number of unique price quotes observed within a regime. This can be viewed as a first indication of rigidity beyond the modal price, which I explore further below.

1.3.3 Within-Regime Pricing Strategies

All products can be grouped into three categories: products characterized exclusively by single sticky price regimes, products consisting entirely of *one-to-flex* regimes, in which a single sticky price is occasionally 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 regime, and finally, *multi-rigid-price* products, which contain regimes in which at least two prices are revisited over the life of the regime. Table 5 summarizes the three strategies, and figure 5 shows a sample price series for each category.

At the product level, products characterized exclusively by single-price regimes represent only 4.9% of series. In this category, I also include products that contain regimes in which I observe a single deviation from the modal price, suggesting that transitory price changes are not a consistent aspect of the firm's pricing policy. For these products, the median regime duration is 45 weeks (versus 31 weeks for all products), and the average size of price changes across regimes has a median of 5.8% across categories (versus 7.5% overall), suggesting that these products face a relatively low volatility of costs and demand.

Motivated by empirical studies that highlight the importance of transitory price changes, recent theoretical work has sought to develop models in which firms have an incentive to temporarily (and flexibly) deviate from a rigid regular price, thereby generating a one-to-flex pattern similar to those simulated by processes II and III above. Kehoe and Midrigan (2010) and the dynamic extension of Guimaraes and Sheedy (2011) generate such one-to-flex patterns in which transitory prices last one period. I find that the one-to-flex pattern accounts for 18.3% of products, of which, for 4.2% of products, transitory prices last only one period, while for the remaining 13.9% of products, transitory price changes last more than one week. For these products, the median implied regime duration is 30 weeks. One-to-flex products have relatively low price volatility both within and across regimes: across all products in this category, 50% of regimes are single-price regimes. As shown in the second column of table 6, for the remaining regimes, within-regime prices change with a frequency of 12.0% and the within-regime absolute size of price changes is 5.6%. Across regimes, the change in the average price per regime has a median of 6.3%. This degree of volatility suggests that for these products, temporary deviations from the rigid mode are not an important aspect of the firm's policy, and it makes these products very similar to the single price products.

Underscoring the presence of rigidity beyond the modal price within each regime, 76.8% of products contain regimes in which at least two prices are revisited over the life of the regime. The median implied duration of regimes for these products is 31 weeks. In contrast to one-to-flex products, multi-rigid products are highly volatile: the median frequency of within-regime price changes across regimes with at least two distinct prices is 28.6%, almost two and a half times that of one-to-flex products. The absolute size of of price changes within regimes is 10.6%, nearly double that of one-to-flex series. Across regimes, all categories show comparable volatility in the average price charged per regime, although multi-rigid products again have somewhat higher volatility: the change in the average price per regime is 7.8% for multi-rigid products, 6.3% for one-to-flex products, and 5.8% for single-sticky products.

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 rigid multi-price plans are not likely to change from one regime to another. Hence, all product series that have at least one *multi-rigid* regime are labeled as pursuing a *multi-rigid* strategy. All product series that have no multi-rigid regimes, but have at least one one-to-flex regime are counted in the one-to-flex category. Finally, all products that consist entirely of single price regimes (where I also include regimes in which I observe a single deviation from the modal price) are counted in the single-sticky category.

The assumption that the firm's type does not change over time implies that single-price and one-to-flex regimes that are part of multi-rigid series should be relatively short-lived, reflecting the fact that a regime change occurred before the full distribution of prices could be realized. Indeed, while the duration of single-price regimes in the purely single-price series is 45 weeks, the average length of *all* single-price regimes is 18 weeks, indicating that single-price regimes included in the one-to-flex and rigid multi-price series are at least half as long as those that are part of the purely single-price series. Finally, rigid multi-price regimes are significantly longer, on average lasting 49 weeks. The fact that rigid multi-price regimes last significantly longer than single-price regimes may be indicative of the fact that the seller uses these multiple prices to adjust to market conditions and can therefore afford to conduct a revision of the entire strategy less frequently.

1.3.4 Distribution of Within-Regime Prices for Rigid Multi-Price Products

How much rigidity is there in the price series belonging to the rigid multi-price group relative to the one-to-flex group, and, along with the differences in within-regime volatility, does it warrant partitioning the data in this way? To investigate this question, I compute the distribution of prices inside each regime that contains multiple rigid prices. The first thing to note is that the raw repetition of multiple price quotes within regimes is quite high. Figure 6 shows the frequency with which the second, third and fourth most common prices are observed in regimes belonging to rigid multi-price products: the second most quoted price occurs at least three times in 75% of regimes; the third most quoted price occurs at least three times in 33% of regimes; and the fourth most quoted price occurs at least three times in 22% of regimes. Repeated values are substantially less common beyond the fourth most commonly quoted price, with the fifth most common price being observed at least three times in only 6% of regimes. Hence a model of multi-rigid price plans would seek to generate up to four-to-five rigid prices for the typical regime.

The frequency with which prices are charged is steeply declining, and this pattern holds across cardinalities, as shown in the panels in figure 7, which plot the frequency with which each top price is observed, as a function of the number of distinct prices observed within the regime, for multi-rigid products: as more prices are added, the frequency of the top price falls, but the pattern of one price dominating holds even for regimes in which there are a relatively large number of distinct prices.

The panels in figure 8 show histograms for the frequency with which top prices are charged across all regimes for all products. The histograms go beyond the medians reported above in illustrating the fact that the data is not aligned with either the single sticky price model or the one-to-flex model: the former would generate a vertical bar reaching $(1, 1)$ in the top left panel; the latter could generate the pattern in that panel given enough heterogeneity in the frequency of the modal price; however, it would not generate meaningful increases in the height of the bars across panels and it would be expected to approach $(1, 1)$ so decisively after only five prices. Conversely, figure 8 shows that while the modal price is indeed important, each subsequent price adds significant mass to the cumulative distribution.

Table 7 reports the median cumulative frequencies for the top five prices, similar to table 8 of Klenow and Malin (2010). I compare the within-regime numbers with the numbers obtained across an entire series, which essentially pick up the "number 1" prices across regimes. For reference, I also include the numbers reported by Klenow and Malin (2010), though these are based on *monthly* data, and they are computed using a rolling mode filter.

1.3.5 Correlation of Regime Length and Regime Cardinality

A strong empirical regularity suggestive of the possibility that prices within a regime are at least partial responses to market conditions is the positive relationship between the length of regimes and the number of unique prices per regime: for rigid multi-price products, the average length of regimes containing two distinct prices is 18.1 weeks. Conversely, rigid multi-price regimes that contain seven or more distinct prices last 44 weeks on average. Table 8 reports the average regime length as a function of the number of distinct prices per regime for both rigid multi-price and one-to-flex series, indicating the relatively wider dispersion in regime lengths for rigid multi-price products.

1.3.6 The Nature of Within-Regime Volatility

What patterns best describe the volatility of prices inside regimes? Are all prices equally likely to be observed over the life of a regime? Or are price series predominantly sequences of single prices interrupted by occasional volatility? The answers to these questions point to different types of theories of pricing. Hence, in this section, I construct artificial series with which to compare the actual data, in order to determine the nature of within-regime volatility.

I compare the location of the breaks and the properties of the regimes found in the data with those found in artificial series generated in two ways, replacing the observed realizations of prices inside each regime with (1) i.i.d. draws from the realized distribution and (2) draws in which all within-regime price changes are clustered at the edges of regimes. Each of these two simulations implies a different approach to modeling product-level price volatility. If different prices are equally likely to be observed over the life of a regime (as in the first simulation), then it would point to a theory of pricing in which the firm chooses a single multiple-price policy that applies over the life of the regime. At the other extreme, if the volatility of prices is confined to relatively short "transitional" periods between single-price regimes (as in the second simulation), then it would suggest a theory of pricing that adapts the existing single-price theories to include a period of experimentation before the firm settles on a new price. Figure 9 contrasts the i.i.d version of two regimes with this single-sticky emphplus volatility version.

First, I explore if the patterns of price changes inside regimes can be approximated by independent draws from the distribution of prices inside each regime. Let r_j, $j = 1, ..., J$, denote the J regimes identified for a particular product series, $\{p_t\}$. For each regime r_j, I sample without replacement from the distribution of prices inside the regime. I concatenate all the simulated regimes into a new artificial series, $\{\widetilde{p}_t\}$. I do this for each product series in the data and build an artificial dataset. I rerun the break test on the artificial dataset to identify the new regimes, $\widetilde{r}_j$, $j = 1, ..., \widetilde{J}$, for each price series. Finally, I repeat the process to generate a second artificial dataset. I compare datasets in terms of the timing of breaks and the resulting statistics at the regime level: I first compare the actual dataset to each artificial dataset, and the compare two artificial datasets, as a measure of how close results should be *expected* to be if within-regime prices were indeed i.i.d.

I find that the i.i.d. simulations are very close to the actual data in terms of identification and location of the breaks: as shown in the first entry of table 9, the test finds essentially the same number of breaks in both datasets. Moreover, 80% of the breaks in the data are found in exactly the same location in the simulation; for 88% of the breaks in the data, the breaks in the simulation are within one week of the breaks in the data. Synchronization in the timing of breaks is slightly lower in the second column of table 9, which reports the comparison between two simulations. This reflects the fact that within-regime volatility is mostly interior, rather than near the edges of each regime, which would affect the timing of breaks. Given the high degree of alignment between the actual data and the i.i.d. data in terms of both the frequency and the timing of breaks, all statistics regarding the types and properties of regimes are largely unchanged.

Next, I expand on the discussion of Klenow and Malin (2010), by directly testing if series are characterized by a sticky (reference) price followed by short-lived volatility before the firm transitions to a new sticky price. As an example of a model that might generate this pattern of singletons alternating with periods of volatility, consider the following adaptation of a menu cost model: suppose that in each period, the firm receives a signal about the value of adjusting its price and compares it to its menu cost. Upon receiving a signal that an adjustment is desirable, the firm enters an experimentation period, during which it "tries out" different prices, and uses them to learn more about the true state, until it settles on its best estimate of the optimal price to be charged until the next time a price review is deemed desirable.

I find that single sticky price regimes alternating with periods of volatility in which the firm experiments with prices before settling on a new price is also a poor fit of the data: as shown in table 10, the break test would identify 54% more breaks, since it would break out the tranquil periods of single sticky prices from the volatile, "experimental" periods.

To measure the distance between the actual and the artificial data, I consider two additional statistics, the share of *uninterrupted* observations and the share of *comeback* prices, with definitions similar to those employed by Klenow and Malin (2010). Consider price p_{tr} in regime r. This price is counted as an *uninterrupted* observation if $p_{tr} = p_{t-1,r}$. This statistic is related to the simulations conducted above: a high rate of uninterrupted price observations could suggest that the volatility of prices does not in fact reflect the existence of multi-price plans. In the data, if there were no multi-price plans, the share of uninterrupted prices would be 93%. At the other extreme, if all regimes were generated via i.i.d. draws from multi-price plans, the share of uninterrupted prices would be 70%. As shown in table 11, the actual data is in the middle, at 80%.

I also consider the number of *comeback* prices as a share of the number of unique prices, where p_{tr} is counted as a comeback price in regime r if 1) it has already occurred inside this regime and 2) $p_{tr} \neq p_{t-1,r}$. For example, in the sequence $\{1; 3; 1; 1\}$, the comeback count would be one and the fraction would be 1/2 (out of two distinct price quotes, one "comes back"). Under the singleton simulation, the comeback share is zero. Again, actual regimes lie in between the single sticky price series and the multi-price i.i.d. series.

In terms of other statistics, such as implied regime durations and break down of regimes by type, the i.i.d. simulation is exactly aligned with the actual data.

1.3.7 Implications for Theories of Pricing

The statistics at the regime level are difficult to reconcile with the most commonly used models of pricing. Full information flexible price models, in which prices are continuously re-optimized, do not generate regimes except to the extent that there are regimes in the underlying shocks, and do not generate mass points in the distribution of prices observed over time, except to the extent that the underlying shocks are themselves drawn from distributions with mass points. By disregarding the substantial rigidity in price *levels* apparent in figure 1, and documented more broadly in above, this approach may overstate the degree of flexibility in the pricing data.

Conversely, sticky price models, such as time-dependent models using Taylor (1980) or Calvo (1983) staggered price-setting or state-dependent models (Sheshinski and Weiss, 1977, Golosov and Lucas, 2007), generate single-price regimes. As in the case of flexible price models, there is no reason for past prices to be revisited once the firm re-optimizes its policy, hence these models cannot explain the discreteness of prices observed in the data. Moreover, sticky price models that abstract from transitory price changes within regimes may overstate the degree of rigidity in the pricing data. As others have documented, a significant portion of firms' revenues is derived from sales at the non-modal prices, which suggests that firms should have a strong incentive to tie transitory prices to concurrent market conditions, at least partially. Klenow and Willis (2007) further document that transitory prices have macro content that does not wash out with aggregation.

The pricing patterns documented in this chapter also suggest that models which assume that only the mode price is rigid, with all other prices being adjusted flexibly, such as the theories proposed by Kehoe and Midrigan (2010) and Guimaraes and Sheedy (2011), may miss an additional source of rigidity in terms of price levels.

The evidence presented is instead consistent with the proposed price plans of Eichenbaum, Jaimovich and Rebelo (2011), according to which firms are assumed to choose from a small set of prices that is updated relatively infrequently, subject to a cost. The theory presented in the next chapter generates such plans endogenously, using fixed and variable costs of information acquisition.

1.4 Break Test versus Filters

I compare the regimes-based method of analyzing price series to existing filtering methods: a v-shaped sales filter similar to those employed by Nakamura and Steinsson (2008), the reference price filter of Eichenbaum, Jaimovich and Rebelo (2011), and the running mode filter of Kehoe and Midrigan (2010), which is similar to that of Chahrour (2011). These filters have been proposed as a way to uncover stickiness in product-level pricing data once one filters out transitory price changes that may be orthogonal to aggregate conditions and hence not of direct interest to macroeconomists.

First, I apply the three filters to product level pricing data from Dominick's Finer Foods stores. I find that standard statistics of interest used to inform theories of price-setting vary significantly across specifications of the different price filters: the estimated median duration of regular prices varies from 10 weeks to 32 weeks across different v-shaped filters and from 26 weeks to 53 weeks for different parametrizations of the rolling mode filter. 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.

Next, I compute the synchronization of the regime test with the different filters. Given the robustness of the break test across different data generating processes, illustrated in section 1.2, the break test offers a direct way to evaluate the assumptions invoked by various filtering methods against the data. For example, to the extent that the true data generating process is indeed consistent with the assumptions made by the v-shaped filter, one would expect to see a high degree of alignment between results obtained by the break test and those obtained by an appropriately parameterized v-shaped filter. Conversely, if results diverge even under the best parameterization of the filter, one can conclude that the data too diverges from the assumptions of the filter.

I find varying degrees of synchronization between the filters and the break test. To understand the sources of discrepancy, I apply the different filters to the simulated processes described in section 1.2.

1.4.1 V-shaped Sales Filters

The application of v-shaped sales filters to product-level pricing data is motivated by the observation that many retailers enact temporary price cuts that may reflect different forms of price discrimination rather than responses to concurrent market conditions. The filters eliminate price cuts that are followed, within a pre-specified window of time, by a price increase to the existing regular price or to a new sticky regular price.

I implement a v-shaped sales filter based on a slight modification of the filter proposed by Nakamura and Steinsson (2008). The algorithm requires choosing four parameters: J, K, L, F. The parameter J is the period of time within which a price cut must return to the regular price in order to be considered a transitory sale. For asymmetric v-shaped sales, in which 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. In the case of asymmetric sales, the parameter F determines whether to associate the sale with the existing regular price or with the new one.

I apply the filter with different parametrizations to Dominick's data, varying the sale window J from three weeks to 12 weeks, K and L from one week (corresponding to the symmetric v-shaped filter) to 12 weeks, and $F \in \{0, 1\}$. The parameter J is the most important determinant of the frequency of regular price changes, hence in table 12 I report statistics for each J, *averaged* across various values of K, L, F. I find that statistics vary significantly with the parameterization, with the median implied duration of regular prices increasing from 13 to 30 weeks as I increase the length of the sale window, J. Increasing J beyond 12 weeks no longer significantly impacts statistics (for example, for $J = 20$ weeks, the median implied duration is 32 weeks). 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.

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 in section 1.2. Hence, I compute the synchronization of the different parametrizations of the v-shaped filter with the regime test: a high degree of synchronization between the two methods would imply robustness of the v-shaped filter. Conversely, if the two methods are not aligned in the data, then we can conclude that the assumptions underlying the implementation of the v-shaped filter are not borne out by the data.

For most parametrizations, the v-shaped method yields shorter regimes compared with the regime test, which yields a median implied duration of 31 weeks. Divergence is primarily driven by the presence of the type II process in the data and by the assumption of a fixed sale window. Moreover, adjusting parameters of the v-shaped filter yields a trade-off in performance: setting a small sales window ensures that the timing of regime breaks is accurately identified, but generates three times more regime breaks in the data. Even a large increase in the sales window still generates 55% more breaks, but substantially reduces the method's ability to estimate the timing of regime changes: synchronization between the filter and the break test falls from 80% to 58%. Hence, it is not the case that the regime test is similar to a v-shaped filter with a longer sale window: the two methods are simply finding different breaks.

Table 13 reports the synchronization between the two methods in Dominick's data as a function of the sale window, J. The first row reports the total number of regime changes found by the break test across all series, while the second row reports the number of change points of the regular price found by the different parametrizations of the v-shaped filter. The filter finds many more breaks, thereby implying a shorter duration of regimes, ranging from 12 weeks to 29 weeks, depending on the sale window. The number of breaks exactly synchronized between the two methods as a fraction of the number of breaks found by the regime test, reported in the fifth row, shows that as the sale window is increased to generate longer regimes, the synchronization between the two methods falls, illustrating the fact that the two methods are finding different break points in the data. The last row of the table reports the median minimum distance between the change points estimated by the two methods, excluding exact synchronizations[8].

For each value of J, the reported results are those for parametrizations of K, L, F that yield the highest degree of synchronization between the v-shaped filter and the regime

[8] I compute the minimum distance between two sets of change points using a standard nearest-neighbor method. For example, for two sets of breaks at locations $\{2; 5; 19\}$ and $\{1; 6; 12; 13; 16\}$, the minimum-distance vector is $\{1; 1; 3\}$ and the resulting mean distance is 1.67.

test. While these parameters do not significantly affect the median implied duration of the regular price, they do affect the timing of breaks, thus affecting synchronization. For example, fixing $J = 3$ while varying the remaining parameters of the v-shaped filter results in a range of exact synchronization from 65% to 80%.

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.

To pinpoint the source of discrepancy between the two methods, I apply the v-shaped filter to the four simulated processes defined in section 1.2. As expected, the filter performs well for series generated from processes I, III and IV, although for the multi-price series it consistently over-estimates the number of breaks. Table 14 reports the results for the parameterization of the v-shaped filter that does the best job of identifying the true change points.

As in Dominick's data, changing the parametrization of the filter to reduce the number of breaks reduces synchronization of the breaks found by the filter with the true breaks from more than 95% to as low as 50%. Since the break test allows the window to vary endogenously, it is likely to outperform a constant-window filter in data that is characterized by random variation in both the duration of regimes and the duration of sales within these regimes.

1.4.2 Reference Price Filters

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 experiment with both a six-week window and a 13−week window. In theory, this method should perform well overall, and should outperform the v-shaped filter for process III. However, the fixed window assumption is quite limiting, and leads to low power and synchronization.

As shown in table 15, only 16% of the breaks identified by the regime-based method are also identified by the reference price method. This low ratio is entirely due to the reference price filter imposing a fixed minimum cutoff for regime 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 regimes is highly variable over time, the two methods are likely to overlap exactly only by chance.

The same pattern, namely very low synchronization, emerges when applying the reference price filter to the simulated data, as shown in table 16.

1.4.3 Rolling Mode Filters

I implement the rolling 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, and without imposing a fixed reference window, as does the reference price filter. For each product, they define an artificial series called the regular price series, which is a rigid rolling mode of the series. Every price change that is a deviation from the regular price series is defined as temporary, and every price change that coincides with a change in the rigid rolling mode price is defined as a regular price change. In this context, I define a regime 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 ($= 2A$), and C, which is 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).

When parameterized to match the number of breaks found by the break test in Dominick's data, Kehoe and Midrigan's (2010) rolling mode filter improves on the synchronization of regime changes found by the reference price filter, and is largely in agreement with the break test, with small differences in the timing of breaks: while exact synchronization with the break test is fairly low, at 55%, the median distance between the breaks found by the filter and those found by the break test is mostly two weeks, indicating that the two methods appear fairly close.

This alignment is confirmed when applying the rolling mode to the simulated processes from section 1.2, as shown in table 17. Overall, the rolling mode filter finds 94% of breaks, with exact synchronization between the filter breaks and the simulated breaks ranging from 100% for the single sticky price series (process I) to 85.4% for the rigid multi-price series (process IV).

1.5 Conclusion

This chapter presents new facts about price adjustment at the micro level using a new statistical method that tests for breaks in the distribution of prices over time. Using weekly grocery store data, I find that regime changes are robustly estimated, with regimes typically lasting 31 weeks. I find strong evidence against both the single-price and the one-to-flex model, and document rigidity in price levels that extends beyond the modal price. Existing models of pricing are at odds with this finding. I also find some evidence that pricing regimes can be viewed as sequences of i.i.d. multi-price plans, with statistics not meaningfully different from simulation-based statistics. Products characterized by multiple rigid prices per regime are also highly volatile, compared with one-to-flex series.

The regime-based approach suggests a new theoretical avenue for modeling pricing behavior, one that is based on infrequently updated multi-price strategies consisting of a small set of prices. A potential interpretation of the findings documented in this chapter is that for products for which frequent transitory adjustment is desirable, the seller finds it optimal to design a pricing strategy that consists of a small number of prices, among which to alternate over the course of the regime. It is conceivable that the resulting persistence of such a model would depend not simply on the frequency with which individual prices change, but rather on the frequency with which strategies are updated, and on the information content of prices inside a regime. Such a theory is presented in the next chapter.

Table 1: Specification of simulated processes.

	I	II	III	IV
Proba(regular Δp)	$\beta \in (0,1)$	$\beta \in (0,1)$	$\beta \in (0,1)$	$\beta \in (0,1)$
Regular shock	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{N}(\mu, \sigma^2)$
Proba(transitory Δp)	$\delta = 0$	$\delta \in (0,1)$	$\delta \in (0,1)$	$\delta \in (0,1)$
Transitory shock	0	$\mathcal{N}(\mu_T, \sigma_T^2)$	$\mathcal{N}(\mu_T, \sigma_T^2)$	$\mathcal{N}(\mu_T, \sigma_T^2)$
Additional constraint	-	$\mu_T = 0$	$\mu - 3\sigma > \mu_T + 3\sigma_T$	$\mu - 3\sigma > \mu_T + 3\sigma_T$
Additional setting	-	-	$0 \leq l_\delta \leq 3$	$n_d = 3$
Initial regular price	$\exp(\varepsilon_0)$	$\exp(\varepsilon_0)$	$\exp(\varepsilon_0)$	$\exp(\varepsilon_0)$
Initial price	p_0^R	p_0^R	p_0^R	p_0^R

Table 2: Determining the critical value for the break test.

Critical value, K	0.874	0.772	0.70	**0.61**	0.60	0.50	0.40
Mean good_reject, % true count	83.9	86.5	88.5	**90.8**	90.9	93.2	95.0
Mean bad_reject, % test count	0.1	0.3	0.7	**1.3**	1.4	4.9	12.2
Max bad_reject, % test count	0.2	0.8	1.8	**4.7**	5.1	10.2	24.1
Mean regime length overshoot	+7	+6	+5	+**3**	+3	−0.2	−5

Table 3: Break test performance for each simulated process ($K = 0.61$).

Process	I	II	III	IV	Average
Good_reject, % true count	90.1	91.4	91.2	90.4	90.8
Bad_reject, % test count	0	0.2	4.7	0.3	1.3
Exact_synch, % good_reject	100	93.8	90.5	91.1	93.9
Mean distance if not exact	0	1.6	3.7	1.6	1.7

Table 4: Summary statistics for Dominick's data.

Category	Code	# Obs	Median # obs/UPC	Median freq(dp)	Median implied duration	Median avg abs size(dp)	Median std(dp)
Full sample	-	690, 578	142	22%	3.9	12%	16%
Analgesics	ana	11,292	133	14%	6.8	11%	14%
Beer	ber	11,310	138	32%	2.5	16%	18%
Bottled Juices	bjc	41,760	132	19%	4.8	9%	13%
Cereals	cer	45,128	192	15%	6.0	14%	23%
Cheeses	che	67,706	190	25%	3.4	9%	14%
Cigarettes	cig	11,532	181	6%	17.6	4%	5%
Cookies	coo	56,327	137	23%	3.8	12%	16%
Crackers	cra	16,960	114	23%	3.9	10%	13%
Canned Soup	cso	47,480	196	16%	5.8	10%	14%
Dish Detergent	did	18,236	98	19%	4.8	9%	13%
Front-end-candies	fec	46,635	164	10%	8.7	16%	21%
Frozen Dinners	frd	12,316	132	37%	2.1	17%	24%
Frozen Entrees	fre	68,580	157	29%	2.9	22%	30%
Frozen Juices	frj	27,316	274	27%	3.2	12%	18%
Fabric Softeners	fsf	18,555	134	17%	5.4	7%	10%
Grooming Products	gro	10,589	107	25%	3.5	11%	13%
Laundry Detergents	lnd	24,236	96	14%	6.2	8%	12%
Oatmeal	oat	9,650	301	17%	5.6	16%	26%
Paper Towels	ptw	7,800	167	23%	3.8	8%	10%
Refrigerated Juices	rfj	23,049	163	32%	2.5	11%	17%
Soft Drinks	sdr	89,117	158	50%	1.4	19%	25%
Shampoos	sha	4,056	66	25%	3.5	21%	26%
Snack Crackers	sna	30,543	177	28%	3.1	11%	15%
Soaps	soa	13,904	112	18%	5.0	7%	9%
Toothbrushes	tbr	5,406	113	19%	4.7	17%	21%
Canned Tuna	tna	14,035	131	17%	5.5	6%	8%
Toothpastes	tpa	16,770	112	26%	3.4	12%	16%
Bathroom Tissues	tti	7,996	185	28%	3.0	9%	13%

Table 5: Breakdown of data by pricing strategy.

	Single-price	One-to-flex	Rigid multi-price
% of products[a]	4.9	18.3	76.8
% of regimes[b]	18.9	30.1	49.6
Regime duration at product level	45	30	31
Avg regime length across all products	18	29	49

[a] The numbers in this row add up to 100% because all the products exhibit some form of rigidity in prices, thereby falling in one of the three categories.

[b] The numbers in this row do not add up to 100% because 1% of regimes are purely flexible in the sense that all price observations in that regime are unique.

Table 6: Breakdown of price volatility by pricing strategy.

	Single-price	One-to-flex	Rigid multi-price
Frequency of within-regime Δp	–	12.0%	28.6%
Abs. size of within-regime Δp	–	5.6%	10.6%
Abs. size of change in avg price per regime	5.8%	6.3%	7.8%

Table 7: Frequency of "top" prices for rigid multi-price regimes.

Median % (Mean %)	Top price	Top 2 prices	Top 3 prices	Top 4 prices	Top 5 prices	Distinct prices
Rigid multi-price	71.4 (67.9)	84.6 (81.8)	90.7 (88.0)	94.2 (91.6)	97.0 (94.0)	4 (4.7)
Dominick's Series	43.3 (47.5)	66.3 (66.5)	78.2 (76.5)	84.8 (82.6)	88.7 (86.6)	14 (17.4)
Klenow & Malin Food[a]	42.4 (47.0)	66.7 (66.7)	79.2 (76.3)	85.5 (81.3)	n.a. (n.a.)	n.a. (n.a.)
Klenow & Malin All[a]	31.4 (37.6)	50.9 (53.2)	62.7 (61.3)	70.1 (66.2)	n.a. (n.a.)	n.a. (n.a.)

[a] Klenow & Malin (2010) numbers are for monthly data.

Table 8: Relationship between regime length and regime cardinality.

N(distinct)	Rigid multi-price (weeks)	One-to-flex (weeks)
1	15.7	23.9
2	18.1	27.0
3	22.2	32.2
4	28.1	35.1
5	31.5	37.3
6	36.7	45.4
7	44.3	47.1

Table 9: Simulation results for i.i.d. rigid multi-price plans.

	1 = data 2 = simulation	1 = simulation 2 = simulation
Number of breaks (% of sample 1)	99.6	99.9
Synch'd breaks, exact (% of sample 1)	80.4	70.6
Synch'd breaks,+/-1 week (% of sample 1)	88.2	83.7

Table 10: Simulation results for experimental rigid multi-price plans.

	1 = data 2 = simulation
Number of breaks (% of sample 1)	153.8
Synch'd breaks, exact (% of sample 1)	75.3
Synch'd breaks,+/-1 week (% of sample 1)	85.0

Table 11: Comparison of data with i.i.d. and experimental simulations.

(Inside regimes)	Multi-price plans (iid)	Data	Experimental volatility
% Uninterrupted prices	69.5	80.0	93.1
% Comeback prices	66.7	33.3	0.0
Median implied regime duration (weeks)	31	31	19
Single-price regimes (% of all regimes)[a]	19.0	19.0	35.6
Rigid multi-price regimes (% of all regimes)[b]	49.8	49.6	1.2

[a] Single-price regimes are defined as regimes consisting of exactly one price.

[b] Multi-price regimes are all regimes with more than two distinct prices.

Table 12: Different parametrizations of the v-shaped filter on Dominick's data.

J	3	5	7	9	12
Mean frequency	0.083	0.053	0.043	0.039	0.037
Median frequency	0.076	0.050	0.040	0.035	0.033
Duration(mean)	11.6	18.3	22.6	25.3	26.8
Duration(median)	12.7	19.6	24.8	28.4	29.9

Table 13: Synchronization of v-shaped filter with break test in Dominick's data.

J	3	5	7	9	12
Number of breaks, break test	18,412	18,412	18,412	18,412	18,412
Number of breaks, v-shaped	66,424	39,659	32,138	28,852	26,596
Median duration, v-shaped	12.4	19.0	23.9	27.5	28.5
Average v-shaped breaks fraction	360	214	177	164	155
Average exact synch fraction	80	68	64	61	58
Median distance b/w breaks	4.0	5.0	6.0	8.0	9.0

Table 14: Simulation results for the v-shaped filter versus the break test.

# breaks/# true breaks	V-shaped	Break Test
I Single sticky price process	100%	90%
II Flexible deviations from rigid mode	450%	94%
III Fexible downward deviations from rigid mode	115%	96%
IV Rigid multi-price process	122%	91%

Table 15: Synchronization of reference price filter with break test in Dominick's data.

W[a]	6	13
Regime duration	21 weeks	41 weeks
Synch'd, exact	$2,883$ (16%)	$1,170$ (7%)
Regimes breaks only	$15,529$ (84%)	$17,242$ (93%)
Filter breaks only	$26,943$ (146%)	$12,813$ (70%)

[a] W is the length of the fixed window in the reference price filter.

Table 16: Simulation results for the reference price filter versus the break test.

	Reference Price	Break Test
# breaks/# true breaks	93%	93%
Exact synch with true breaks	17%	94%

Table 17: Simulation results for the rolling mode filter versus the break test.

	Rolling Mode ($W = 10$)	Break Test
# breaks/# true breaks	94%	93%
Exact synch with true breaks	94%	94%
Mean distance if not synch'd	2	2
Overshoot of median reg length	+2.3	+3

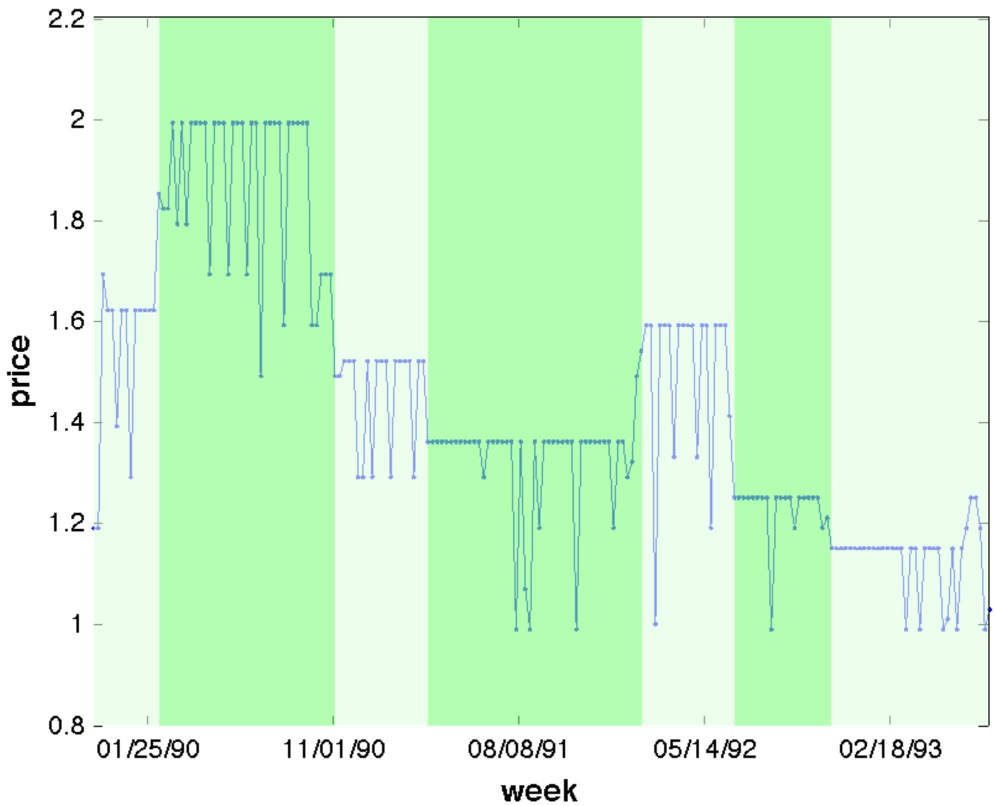


Figure 1: Sample price series (frozen juice) from Dominick's data. The shading marks the identified pricing regimes.

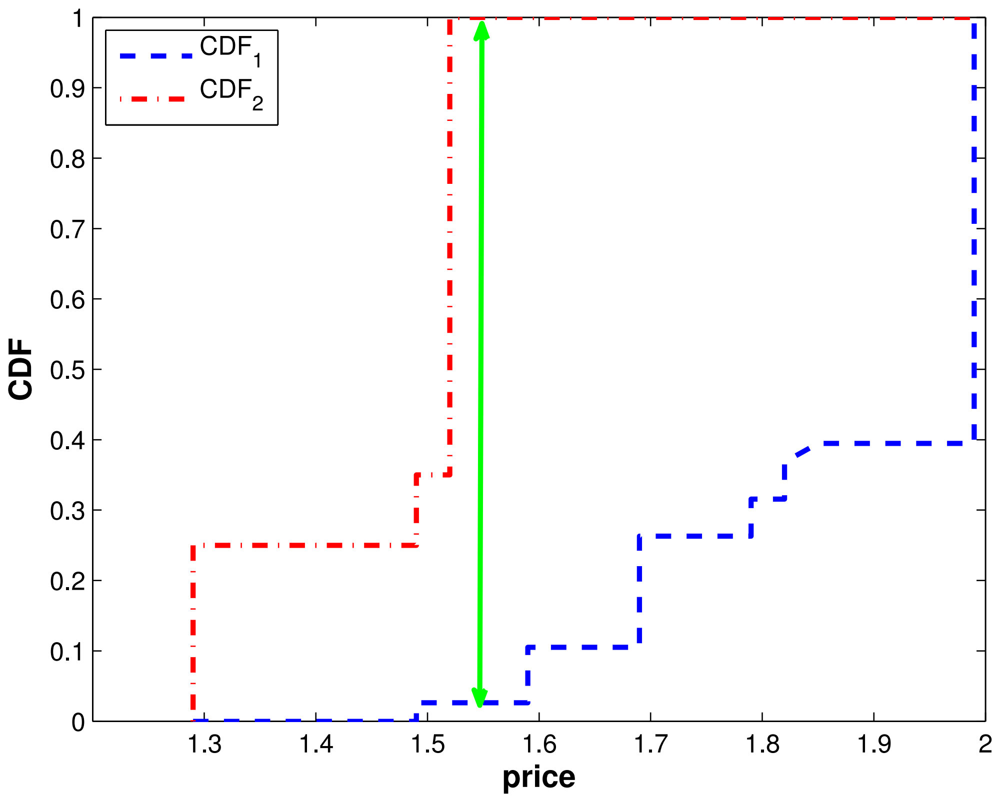


Figure 2: The Kolmogorov-Smirnov distance (maximum marked by vertical line).

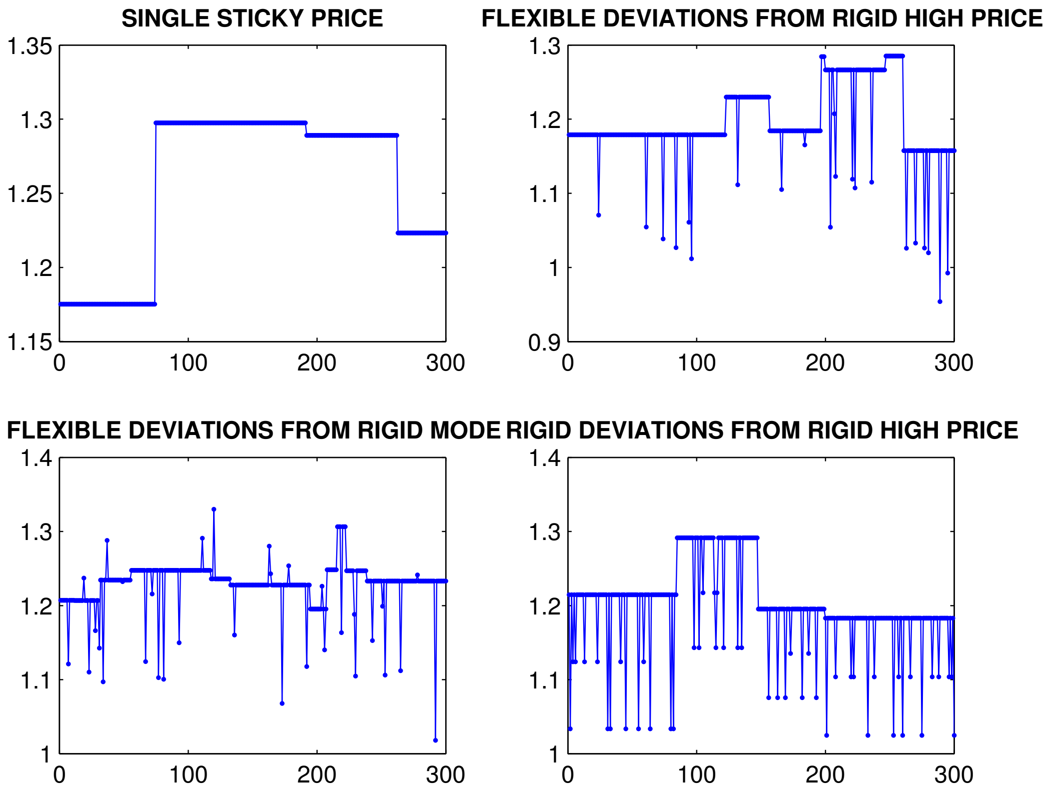


Figure 3: Sample series for four simulated processes.

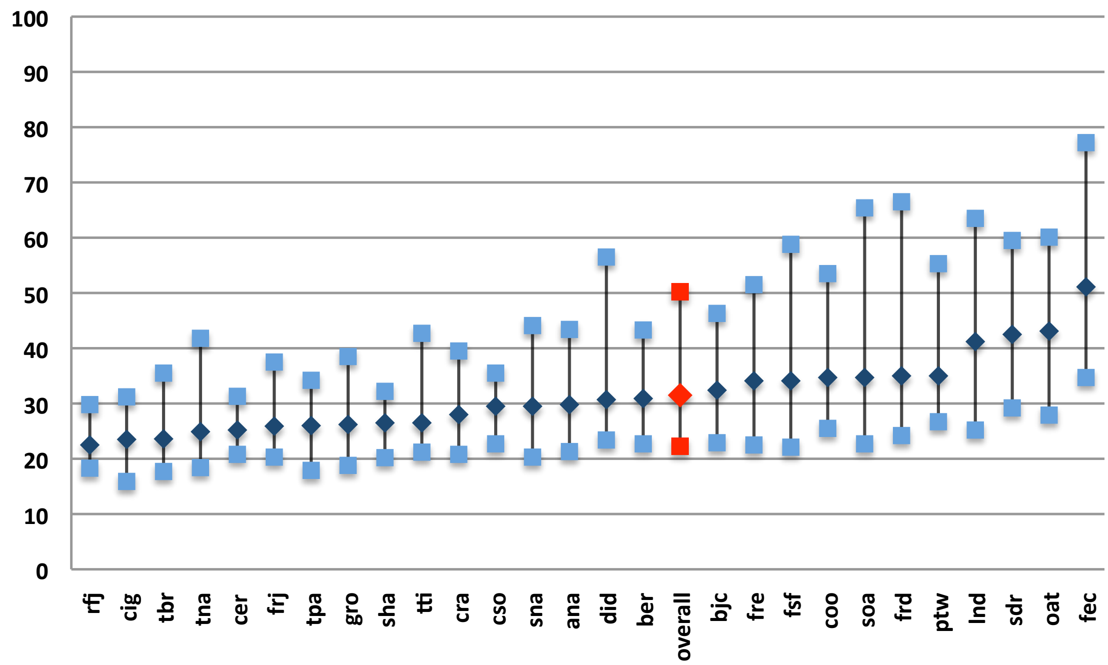


Figure 4: Regime durations in Dominick's data (medians and interquartile ranges by product category).

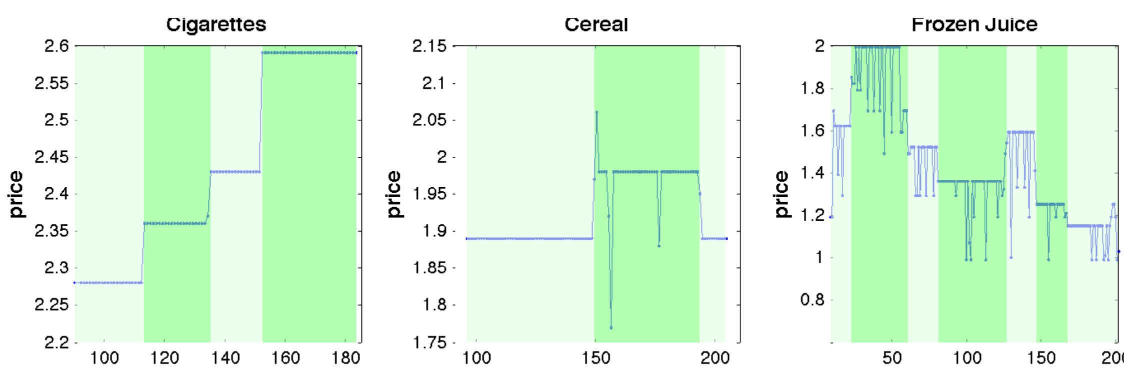


Figure 5: Sample pricing policies in Dominick's data.

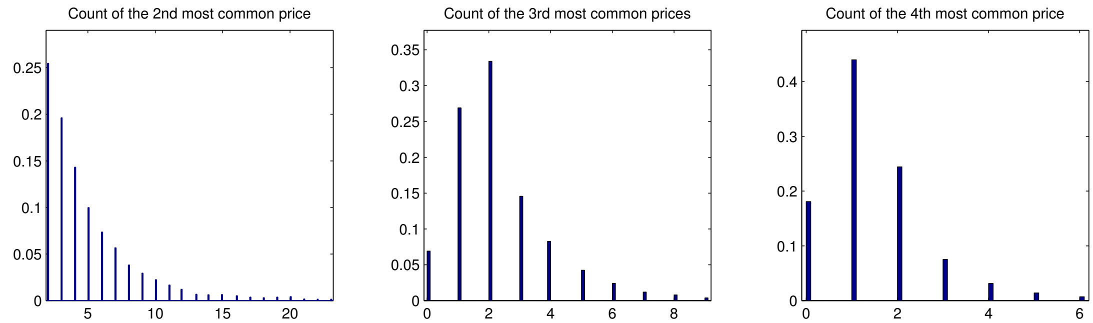


Figure 6: Rigidity of non-modal prices in rigid multi-price policies (count of top prices across regimes).

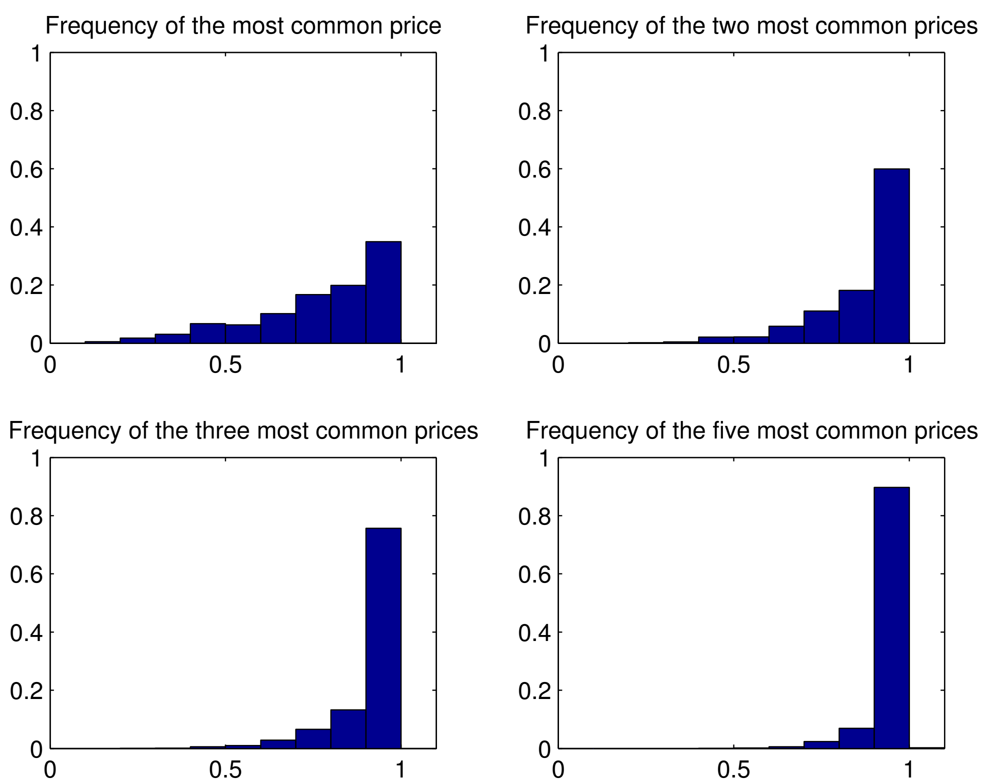


Figure 7: Frequency of top prices across policies.

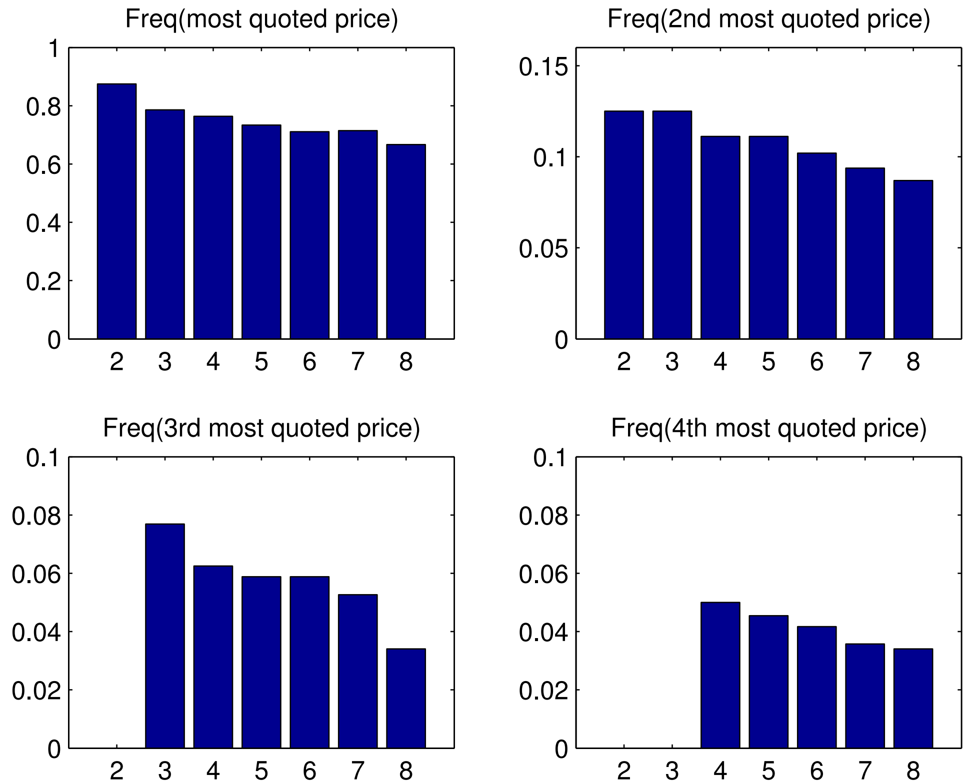


Figure 8: Frequency of top prices by regime cardinality.

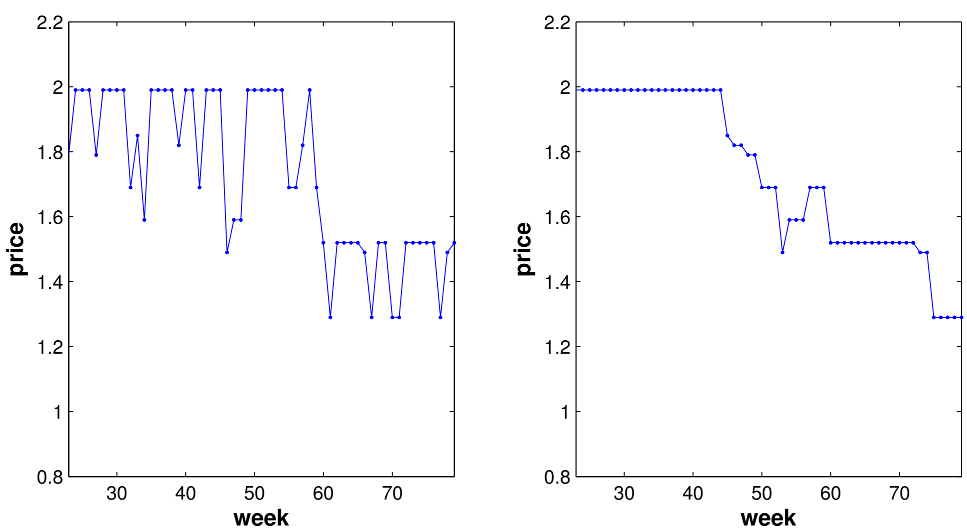


Figure 9: Simulations of i.i.d. and experimental regimes.

2 Discrete Price Adjustment in a Model with Multiple-Price Policies

2.1 Introduction

This chapter develops a theory of price setting based on imperfect information that yields pricing policies that are sticky and simple, namely, they are updated infrequently and consist of a small set of prices. Both the stickiness and the coarseness of the pricing policy are a result of the firm's need to economize on information costs.

I consider the problem of a monopolistically competitive firm that sets prices subject to uncertainty in its demand and its production technology. Obtaining any information about the state of the world is costly in two ways. First, both the firm's prices and its acquisition of information are determined by a policy that can be reviewed subject to a fixed cost. As in Reis (2006), payment of this cost enables the firm to collect complete information about the state of the world at the time of the review. Second, in every period between policy reviews, the firm acquires additional information, based on which it makes two decisions: whether or not to review its policy and, if the policy consists of a menu of prices, which price to charge. The additional information acquired between policy reviews is subject to a cost per unit of information, which captures the cost of monitoring market conditions. The measurement of the amount of information acquired for each decision follows the rational inattention literature (Sims, 2003). The signals that the firm chooses to receive compress the state into a simpler representation, given the firm's objective, the fixed and variable costs of information, and the market conditions that the firm expects to encounter under the current policy, until the next review. For each decision, the firm has access to no other information except that received through the corresponding signal: the review signal and the price signal act as the only interface between the firm and its environment at the time of each decision.

I first show that the firm's optimal policy consists of three elements: 1) a single hazard function that specifies the probability of conducting a policy review conditional on the current state, for all states and periods between reviews, 2) a set of prices, and 3) a single conditional distribution that specifies which price to charge conditional on the current state, for all states and periods between reviews. Together with the evolution of market conditions, these elements determine the frequency with which the firm undertakes reviews and the frequency with which it charges different prices between reviews. The optimal policy has the same form for all periods until the next review. Moreover, each review generates a shift in the optimal distribution of prices. Hence, every policy review starts a new regime, and every regime is identified by a shift in the distribution of prices.

Prices vary stochastically with the state, as in other rational inattention pricing models (e.g., Matejka, 2011), and policy reviews are stochastically state-dependent and independent of the time elapsed since the last review, as in Woodford (2009). 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. Obtaining more precise signals requires purchasing a larger quantity of information in each period. Hence, the firm faces a trade-off between economizing on information expenditure and pricing accuracy. The degree to which prices respond to concurrent market conditions depends on this trade-off.

For a given review policy, I characterize how the firm's pricing policy depends on the cost of the price signal. I present conditions that can be used to determine the optimal support of the distribution of prices charged between reviews, and I use these conditions to determine numerically if the optimal pricing policy is discrete. I show that depending on parameter values, either a single-price or a multiple-price policy may be optimal. In particular, I establish a positive 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. Numerical examples illustrate the optimality of a single-price policy for high enough (though still finite) information costs. For lower information costs, I illustrate how the number of prices in the support increases as the cost of information is decreased. These results are generated with an efficient algorithm that builds on existing work in the information theory literature.

Calibrations of the model qualitatively match the features of price series in the Dominick's data set, documented in Stevens (2011). Depending on parameter values, and consistent with the empirical evidence, the model can generate both single-price and multiple-price regimes that are updated relatively infrequently. For the case of multiple-price policies, regimes consist of a small number of distinct prices, but are nevertheless characterized by frequent and large within-regime price changes. Hence, the model endogenously generates transitory volatility to and from discrete price levels. Figure 11 shows a sample price series simulated from the model. The shading marks the timing of policy reviews.

Relation to the Literature

This theory contributes to the existing literature by providing a framework that generates pricing regimes that consist of a small set of prices. Moreover, consistent with the data, prices within regimes deviate frequently and by large amounts from seemingly sticky levels. Hence, the model can reconcile large transitory volatility with apparent rigidity in price *levels*.

Full information flexible price models, in which prices are continuously re-optimized, do not generate regimes except to the extent that there are regimes in the underlying shocks, and do not generate mass points in the distribution of prices observed over time, except to the extent that the underlying shocks are themselves drawn from distributions with mass points. By disregarding the substantial rigidity in price *levels* documented in Stevens (2011), these models may overstate the degree of flexibility in the pricing data.

Sticky price models, such as time-dependent models (Taylor, 1980 or Calvo, 1983) or state-dependent models (Sheshinski and Weiss, 1977, Golosov and Lucas, 2007), generate single-price regimes. As in the case of flexible price models, there is no reason for past prices to be revisited once the firm re-optimizes its policy, hence these models cannot explain the discreteness of prices observed in the data. Moreover, sticky price models that abstract from transitory price changes within regimes may overstate the degree of rigidity in the pricing data. As others have documented, a significant portion of firms' revenues is derived from sales at the non-modal prices, which suggests that firms should have a strong incentive to tie transitory prices to concurrent market conditions, at least partially. Klenow and Willis (2007) further document that transitory prices have macro content that does not wash out with aggregation.

It is important to note that in the model proposed here there are no physical costs of price adjustment; in fact, prices can change all the time in this model. Rigidity arises because they are always drawn from a fixed set of prices over the life of the regime, and are based on noisy information about market conditions. There are also no a priori constraints on the firm's ability to change "regular" versus "temporary" prices, thus distinguishing this model from those proposed by Kehoe and Midrigan (2010) and Guimaraes and Sheedy (2011).

The model brings together different features of the growing literature on imperfect information in price setting. In particular, the introduction of both fixed and variable costs of information combines two competing approaches to modeling information acquisition. However, the model departs from both literatures by generating *simple* pricing policies that consist of a small set of prices.

First, as in the inattentiveness model of Reis (2006), the strategy that is used to decide when to conduct the next review is itself part of the policy that is chosen at the time of a review. In the model of Reis (2006), the policy specifies the *path* of prices to be charged until the next review, and the *date* of the next review. Between reviews, the firm cannot obtain any information about market conditions, other than information regarding the passage of time, which is available for free. In contrast, I allow the firm to acquire information between reviews, but all information, including knowledge about the number of periods since the last review, is subject to the same cost per unit of information. The resulting timing of reviews and the price charged in each period are stochastically state-dependent rather than time-dependent. In this model, a perfectly precise review signal would generate the triggers in an Ss model of policy reviews, as in the model of Burstein (2006). Conversely, if the firm acquired no information through its review signal, the timing of policy reviews would be completely random, as in the model of Mankiw and Reis (2002).

Second, as in the rational inattention literature, the acquisition of information between reviews is subject to a cost per unit of information, using entropy as a measure of information. Allowing the firm to occasionally review its policy, subject to a cost, generates regime changes, distinguishing this setup from other rational inattention papers, such as those of Sims (2003, 2006), Mackowiak and Wiederholt (2009), or Matejka (2011). In those models, the firm specifies the optimal policy once and for all at some initial date, and then receives signals in accordance with that policy. In contrast, I model both the decision to change the price and the decision to change the overall policy, and hence to move to a new regime. The fact that the firm can occasionally review its policy means that it can implement simple policies between reviews.

Moreover, other rational inattention models assume that the cost per unit of information applies to current market conditions, while the full history of past signals is available for free. In contrast, I assume that all information, including memory of past events and knowledge of the number of periods elapsed since the last review, is subject to the same cost per unit of information. This assumption identifies the information friction directly with the limited attention of the *decision-maker* processing the information from a particular signal. This assumption is critical in generating regimes that are identified by a single distribution of prices: without it, the firm would charge prices from a different pricing policy in every period; moreover, the optimal policy would not generate a discrete distribution of prices, except in the special case of i.i.d. variations in market conditions, as assumed in Matejka (2011).

This treatment of time and memory is the same as in Woodford (2009), who also models policy reviews that are subject to a fixed cost and whose timing is determined by a stochastically state-dependent hazard function. The present model differs from Woodford (2009) along two dimensions. Firstly, I relax that model's assumption that between policy reviews the firm charges a single price. Introducing the price signal generates price volatility between policy reviews, consistent with the empirical evidence of multiple-price regimes documented in Stevens (2011). Secondly, I allow the firm to redesign its signals at each review, whereas in Woodford (2009), the firm's information acquisition policy is chosen once and for all at some initial date.

Section 2.2 presents the setup and introduces the information costs, starting from the full-information frictionless benchmark. Section 2.3 presents the acquisition of information between reviews and defines the firm's problem. Section 2.4 derives and discusses the optimal policy. Section 2.5 discusses the model's ability to generate price patterns that match the empirical evidence. Section 2.6 concludes.

2.2 Setup

A monopolistic firm producing a non-durable good must choose the price to charge for its output in every period, subject to a demand function and a production technology that vary stochastically. The firm's per-period profit in units of marginal utility, $\pi(p-x)$, is a function of the firm's actual log-price, p, and its target log-price, x. The profit function is a smooth real-valued function with a unique global maximum at $p = x$.

All the information about firm-specific and aggregate market conditions that the firm needs in order to choose its optimal price is summarized in the target price, x_t.

This target is a linear combination of the exogenous disturbances in the economy, both transitory and permanent. It evolves over time according to:

$$x_t = \widetilde{x}_t + \upsilon_t, \tag{2.1}$$

$$\widetilde{x}_t = \widetilde{x}_{t-1} + \widetilde{\upsilon}_t, \tag{2.2}$$

where the permanent and transitory innovations, $\widetilde{\upsilon}_t$ and υ_t, are drawn independently from known distributions $h_{\widetilde{\upsilon}}$ and h_{υ}. After both $\widetilde{\upsilon}_t$ and υ_t have been realized, the period t price is set and orders are fulfilled.

Section 2.5 maps a standard monopolistic competition model with Dixit-Stiglitz preferences into this specification.

2.2.1 Full Information

In the frictionless benchmark, the firm chooses a pricing policy that specifies what price to charge in each period and state of the world, to maximize its discounted profit stream,

$$E_0 \sum_{t=0}^{\infty} \beta^t \pi(p_t - x_t), \tag{2.3}$$

where $\beta \in (0,1)$ is the discount factor.

In the absence of information costs, the firm perfectly observes the realization of x_t in each period. If there are no other frictions, such as physical costs of price adjustment, the firm's optimal policy is to charge

$$p_t = x_t, \ \ \forall t. \tag{2.4}$$

2.2.2 Costly Information

I depart from the frictionless benchmark by assuming that although complete information about the state of the economy is available in principle, the firm must expend resources to receive any information in order to make its pricing decision in each period. The measurement of information is based on the literature on rational inattention (Sims, 1998, 2003, 2006). In this setting, acquiring a larger quantity of information leads to higher precision in tracking market conditions. Higher precision in turn implies that the information-constrained firm sets a price that is closer to the frictionless optimal price. Hence, the firm faces a trade-off between economizing on information costs and setting prices that are close to the frictionless optimum.

As in other dynamic models of rational inattention (e.g. Sims, 2003, or Mackowiak and Wiederholt, 2009), both the firm's prices and its acquisition of information are determined by an endogenously chosen policy. The quantity and type of information that the firm chooses to acquire depend on the distribution of shocks in the economy, how sensitive profits are to deviations of the price from the frictionless optimum, and how costly it is for the firm to acquire and process information.

In an important departure from these models, I assume that the firm's policy can be occasionally reviewed, subject to a cost, as in Woodford (2009). The firm chooses to review its policy when it receives information that suggests its current policy has become obsolete relative to the evolution of market conditions since the last review. Therefore, in every period, the firm must acquire information not only to decide what price to charge, but also to decide whether or not it should review its policy.

When conducting a review, the firm chooses 1) a *review policy* that specifies the acquisition of information for the review decision, and the rule for deciding, in each period, whether or not to conduct a review, based on this information; and 2) a *pricing policy* that specifies the acquisition of information for the pricing decision, and the rule for setting prices, in each period, based on this information.

The acquisition of information for each of the two decisions is subject to a variable cost of monitoring market conditions. Letting I_t^r denote the quantity of information acquired for the review decision in period t, the information expenditure associated with this decision is $\theta^r I_t^r$. Similarly, the expenditure on acquiring information for the pricing decision in period t is $\theta^p I_t^p$. The two monitoring costs, θ^r and θ^p, are not necessarily equal. For instance, it may be the case that two individuals with different costs of acquiring information make the two decisions within the firm. For each decision-maker, the unit cost determines the information processing capacity that the decision-maker allocates to this problem. The quantities of information acquired for each of the two decisions are defined in the next section.

The fixed cost of conducting a policy review, denoted by κ, is a different type of information cost. It represents the managerial resources associated with acquisition of the information necessary to design a new policy, and with the decision-making and communication of the new policy. As documented by Zbaracki et al (2004), firms spend a significant amount of resources acquiring information and deciding what type of policy to implement. Payment of this cost allows the firm to acquire extensive information about the state of the world, on the basis of which it designs its new policy. For simplicity, I assume that it enables the firm to receive *complete* information about the state of the world at the time of the review. The assumption that the cost is fixed can be rationalized via economies of scale in the review technology. This assumption follows Reis (2006).[9]

The firm's objective under costly information is to maximize

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\pi(p_t - x_t) - \theta^r I_t^r - \kappa \delta_t^r - \theta^p I_t^p \right], \tag{2.5}$$

[9] The assumption of a fixed cost of policy reviews is also similar to that of Burstein (2006), except for the fact that in that model, the firm has full information at all times for free, and the fixed cost represents only the resources required to design and communicate the new policy.

where $\theta^r I_t^r$ is the cost expended to make the review decision in period t, κ is the fixed cost of a policy review, δ_t^r is an indicator function that is equal to 1 if the firm reviews its policy in period t and 0 otherwise, and $\theta^p I_t^p$ is the cost expended to make the pricing decision in period t.

A key assumption embedded in the objective defined in equation (2.5) is that for each decision-maker, the quantity of information required for this particular problem is small relative to that decision-maker's total information processing capacity. As a result, the two costs per unit of information may be taken as fixed. Hence, this same unit cost applies to all types of information that may be relevant[10] for that manager's problem, regardless of their degrees of complexity. This assumption follows Woodford (2009).

It is important to underscore that in this setting, as in Woodford (2009), no potentially relevant information is available for free. In particular, information that might be stored in memory (such as the history of past signals and decisions) is equally costly to access as is information available externally. Hence, the unit costs may be interpreted as the effort on the part of the decision-maker expended to *process* one unit of information, regardless of where that information may be stored when not in use.

In contrast, in the model of Sims (2003) and in other dynamic rational inattention papers, the entire history of past signals is available to the decision-maker for free in each period, prior to acquiring the information for that period.[11] The availability of that side information makes those models stationary. However, it is not required in the current setup, given the firm's ability to occasionally review its policy.

[10]The types of information that are potentially relevant to each decision, and are therefore included in the state variable in each period, include information about the current market conditions, the history of signals previously received and prices charged, and the number of periods that have elapsed since the last review.

[11]Note that if given this side information, the decision-maker can track market conditions more precisely, for a given level of expenditure on information.

2.2.3 Sequence of Events

The firm's policy specifies the acquisition of information in the form of signals that *compress* the state of the world into a simpler representation. The sequence of events that occur in each period t is as follows:

1. The value of the permanent innovation, $\widetilde{\upsilon}_t$, is realized.

2. The firm receives the *review signal*, based on which it decides whether or not to undertake a review, in accordance with its current policy:

   (a) if it decides to undertake a review, it pays κ, obtains complete information about the current state of the world, and chooses a new policy that consists of a strategy for its review decision, to be implemented starting in period $t+1$, and a strategy for its pricing decision, to be used starting in period t;

   (b) otherwise, the existing policy is maintained.

3. The value of the transitory innovation, υ_t, is realized.

4. The firm receives the *price signal*, based on which it decides what price to charge in the current period, in accordance with its current policy.

5. Period-t demand is met and profits are realized.

The assumption that in each period the firm makes its review decision before that period's transitory shock is realized is a simplification that reduces the state space relevant for this decision, while only having small quantitative implications. If, instead, all shocks were realized at the beginning of the period, the review decision would depend on both types of shocks. However, the extent to which the transitory shock would impact the review decision would be small: only particularly large transitory shocks would justify triggering a review, despite the transient character of the shock. The timing assumption abstracts from this complication by eliminating the possibility of such an effect.

2.3 The Firm's Problem

This section reformulates the firm's objective in terms of the choices that the firm makes each time it undertakes a review, and it defines the firm's complete optimization problem.

First, I define the signals that inform the firm's two decisions, and the quantities of information required by each signal, starting from general definitions for each signal. A crucial determinant of the optimal signals is the way in which one measures the quantity of information conveyed by each signal, I_t^r and I_t^p. Following the rational inattention literature, I use a measure derived from information theory (Shannon, 1948), which quantifies the reduction in the agent's uncertainty about the state of the world at the time of the receipt of the signal.[12] The cost of information is linear in this quantity.

I then simplify the definition of each policy, using some preliminary results that exploit the information theoretic framework. In particular, the most efficient policy for each of the two decisions is shown to generate signals that directly specify the decision that the firm should make. In the case of the review decision, the review signal directly indicates whether or not the firm should undertake a review in the current state. In the case of the pricing decision, the price signal directly tells the firm what price to charge conditional on the current state.

These results ensure that the firm implements the most efficient signal structure, namely one that does not entail acquiring any superfluous information. Moreover, they allow me to abstract from any implementation details: it is not necessary to specify what data the firm monitors in order to make each of its two decisions. All that is needed is a mapping between the state and the final decision. Finally, these results allow me to redefine the firm's objective in terms of a tractable set of firm choices.

[12]See Cover and Thomas (2006) for an introduction to information theory.

2.3.1 The Review Policy

Let $\widetilde{\omega}_t$ denote the complete state of the world at the time of the receipt of the review signal in period t. It includes the realization of the permanent shock in the current period, $\widetilde{\upsilon}_t$, and the full history of shocks, signals, and decisions through the end of period $t-1$. Suppose that the firm reviews its policy in this period. The new review policy is implemented starting in period $t+1$. I begin with a general definition of the review policy, which specifies the set of possible review signals, the probability of receiving each signal in each period and in each state of the world until the next review, and an action rule that maps each signal into the firm's decision of whether or not to undertake a new review.

Definition 1. *A general* 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 that will be received until the next review;*
2. $\{\rho_{t+\tau}(r|\widetilde{\omega}_{t+\tau})\}_\tau$*, the sequence of conditional probabilities of receiving the review signal* r*, for all* $r \in \mathcal{R}_t$*, all* $\tau > 0$*, and all* $\widetilde{\omega}_{t+\tau}$ *that follow the policy review, until the next review;*
3. $\overline{\rho}_t(r)$*, the overall frequency with which the decision-maker anticipates receiving each review signal, until the next review, for all* $r \in \mathcal{R}_t$*;*
4. $\lambda_t(r) : \mathcal{R}_t \to [0,1]$*, the decision rule for conducting a policy review, which specifies the probability of conducting a policy review when the review signal* r *is received, for all* $r \in \mathcal{R}_t$*.*

The first three elements of the review policy can be thought of as the interface between the decision-maker and his environment, while the last element maps the information received through this interface into the decision-maker's actions. The first element of the review policy, the set of possible signals, $\mathcal{R}_t$, can be arbitrarily large. It can include any type of indicator that may be useful for determining whether or not the policy currently in effect has become obsolete, relative to the evolution of the state, $\widetilde{\omega}_t$. The second element, the conditional probability of receiving a particular signal, can be related in an arbitrary way to the complete state, or to any part of the state in each period, and this relationship can vary with each future period, $t+\tau$, until the next review. Before discussing the specification and role of the final two elements in the design of the policy, it is necessary to derive the quantity of information that is acquired and used by the policy.

The Cost of the Review Policy

The quantity of information transmitted by an optimally-designed policy plays a dual role in this setup. On the one hand, it measures the amount of information that is *acquired* by the decision-maker, quantifying how much of the decision-maker's uncertainty about the state has been removed upon receipt of the signal. Given the unit cost, θ^r, it yields the cost of the given signalling mechanism. On the other hand, it measures the amount of information that is *used* by the decision-maker, quantifying the reduction in uncertainty that is reflected in his actions.

I begin by deriving the quantity of information acquired by a decision-maker who uses the policy specified in definition 1 using the definition of mutual information between the signal and the state. Following Shannon (1948), the entropy of a random variable represents the amount of information that the decision-maker would have to receive in order to eliminate all uncertainty about that random variable. The residual uncertainty, conditional on receipt of a signal, is given by the conditional entropy. Hence, the quantity of information acquired through the signal is the relative entropy, namely, the difference between entropy and conditional entropy.

For convenience, for both the review signal and the price signal specified in the next section, I employ an equivalent definition, namely the average amount by which uncertainty about the optimal signal would be reduced if the firm could observe the state. These two definitions are equivalent since the information about the state contained in the signal is equal to the information about the signal contained in the state. Exploiting this symmetry simplifies the exposition.

The entropy of a signal with density $\overline{\rho}$ is defined as

$$H(\overline{\rho}) \equiv \sum_{r \in \mathcal{R}} \overline{\rho}(r) \log \frac{1}{\overline{\rho}(r)}. \tag{2.6}$$

For expository purposes, $\mathcal{R}$ is a countable set, though the definition can be modified to allow for continuous signal distributions. It will be established below that the optimal set of review signals is not only countable, but finite.

Let the entropy of the signal conditional on the state be denoted by $H_{\widetilde{\omega}}(\rho)$, with

$$H_{\widetilde{\omega}}(\rho) \equiv E\left\{\sum_{r \in \mathcal{R}} \rho(r|\widetilde{\omega}) \log \frac{1}{\rho(r|\widetilde{\omega})}\right\}. \tag{2.7}$$

This leads to a generic definition for the quantity of acquired information as the difference between entropy and conditional entropy.

Definition 2. *The quantity of information that is acquired in order to implement a signalling mechanism defined by* $\{\mathcal{R}, \rho(r|\widetilde{\omega}), \overline{\rho}(r)\}$ *is* $E\{I^r(\rho(r|\widetilde{\omega}), \overline{\rho}(r))\}$, *where*

$$I^r(\rho, \overline{\rho}) \equiv \sum_{r \in \mathcal{R}} \rho(r|\widetilde{\omega}) \left[\log \rho(r|\widetilde{\omega}) - \log \overline{\rho}(r)\right]. \tag{2.8}$$

The quantity of information that is expected to be acquired in a particular state $\widetilde{\omega}$, $I^r(\rho(r|\widetilde{\omega}), \overline{\rho}(r))$, is a function of the frequency $\overline{\rho}(r)$ that the firm anticipates prior to receiving the signal. More precise information about $\widetilde{\omega}$ implies a bigger difference between the conditional and the unconditional distributions.[13]

Using this definition, I now define the quantity of information acquired through the review policy specified in definition 1.

Definition 3. *The quantity of information expected, at the time of the review in an arbitrary state $\widetilde{\omega}_t$ and period t, to be* acquired *in the implementation of the review policy specified in definition 1, in each period $t+\tau$, $\tau > 0$, over all states $\widetilde{\omega}_{t+\tau}$ that follow the policy review, until the next review, is given by*

$$I^r_{t+\tau} \equiv E_t \left\{ I^r \left(\rho_{t+\tau} \left(r|\widetilde{\omega}_{t+\tau} \right), \overline{\rho}_t \left(r \right) \right) \right\}, \tag{2.9}$$

where $I^r \left(\rho_{t+\tau} \left(r|\widetilde{\omega}_{t+\tau} \right), \overline{\rho}_t \left(r \right) \right)$ is defined in equation (2.8), and $E_t \left\{ \cdot \right\}$ 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.

Each conditional distribution, $\rho_{t+\tau} \left(r|\widetilde{\omega}_{t+\tau} \right)$, together with the distribution of the state in that period under the chosen review policy, implies a period-specific frequency of reviews, $\overline{\rho}_{t+\tau} \left(r \right)$. However, the sequence of these marginal distributions is not relevant for the design of the policy at review time. Prior to receiving the signal in each period, the decision-maker has no side information, including no knowledge of τ or of past signals. Hence, he does not know which frequency $\overline{\rho}_{t+\tau} \left(r \right)$ to anticipate. He only knows that over the life of the policy he can anticipate review signals to occur with a frequency $\overline{\rho}_t \left(r \right)$. As a result, the quantity of information acquired in each period is measured as the reduction in uncertainty relative to this common marginal, $\overline{\rho}_t \left(r \right)$. Therefore, the firm designs a single information structure that generates signals from multiple potential

[13] The quantity of information can also be seen to be equal to the Kullback-Leibler distance between the probabilies of r and $\widetilde{\omega}$: a low quantity of information means that the joint distribution of r and $\widetilde{\omega}$ is close to the product of the marginals. If the two random variables are independent, the joint distribution is equal to the product of the marginals, and the quantity of information conveyed by the signal is zero.

sources, without the decision-maker knowing which source is "active" at any point in time.[14] If the sources are not identical, this information structure will necessarily entail acquiring a larger quantity of information, relative to the case in which a separate signalling mechanism is designed specifically for each source.

The Precision of the Review Policy

The amount of information that is *used* by the decision-maker employing the policy specified by definition 1 measures the reduction in uncertainty that is reflected in the final binary decision (review or do not review), given the state. I begin by defining this quantity for a static review policy, $\{\mathcal{R},\, \rho\left(r|\widetilde{\omega}\right),\, \overline{\rho}\left(r\right),\, \lambda\left(r\right)\}$. Under such a policy, the firm expects to undertake a review in state $\widetilde{\omega}$, with probability given by

$$\Lambda\left(\widetilde{\omega}\right) \equiv \sum_{r\in\mathcal{R}} \lambda\left(r\right) \rho\left(r|\widetilde{\omega}\right), \tag{2.10}$$

and expects to retain the existing policy with probability $1-\Lambda\left(\widetilde{\omega}\right)$. The probability with which the decision-maker anticipates undertaking a policy review across all states $\widetilde{\omega}$ is

$$\overline{\Lambda} \equiv \sum_{r\in\mathcal{R}} \lambda\left(r\right) \overline{\rho}\left(r\right). \tag{2.11}$$

This leads to the definition of the quantity of information that is used by the decision-maker employing this policy.

Definition 4. *The quantity of information that is used by a review policy defined by* $\{\mathcal{R},\, \rho\left(r|\widetilde{\omega}\right),\, \overline{\rho}\left(r\right),\, \lambda\left(r\right)\}$ *is* $E\left\{I^r\left(\Lambda\left(\widetilde{\omega}\right),\overline{\Lambda}\right)\right\}$, *where*

$$I^r\left(\Lambda,\overline{\Lambda}\right) \equiv \Lambda\left[\log\Lambda - \log\overline{\Lambda}\right] + \left(1-\Lambda\right)\left[\log\left(1-\Lambda\right) - \log\left(1-\overline{\Lambda}\right)\right], \tag{2.12}$$

[14] Thank you to Mike Woodford for this insight.

is the relative entropy between two binary random variables for which the probabilities of observing the signal $r = 1$ *are* $\Lambda(\widetilde{\omega})$ *and* $\overline{\Lambda}$*, respectively, where* $\Lambda(\widetilde{\omega})$ *and* $\overline{\Lambda}$ *are defined in equations (2.10) and (2.11), respectively.*

Using this definition, I now define the quantity of information expected, at the time of the review, to be used by decision-maker employing the policy given in definition 1.

Definition 5. *The quantity of information that the decision-maker expects to* use *when implementing the review policy given in definition 1, in each period* $t+\tau$*,* $\tau > 0$*, over all states* $\widetilde{\omega}_{t+\tau}$ *that follow the policy review, until the next review, is given by*

$$J^r_{t+\tau} \equiv E_t\left\{I^r\left(\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right), \overline{\Lambda}_t\right)\right\}, \tag{2.13}$$

where

$$\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right) \equiv \sum_{r\in\mathcal{R}} \lambda_t(r)\,\rho_{t+\tau}\left(r|\widetilde{\omega}_{t+\tau}\right), \tag{2.14}$$

$$\overline{\Lambda}_t \equiv \sum_{r\in\mathcal{R}} \lambda_t(r)\,\overline{\rho}_t(r), \tag{2.15}$$

$I^r\left(\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right), \overline{\Lambda}_t\right)$ *is defined in equation (2.12).*

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

Recall that before the receipt of the review signal in each period, the firm has no additional information (including knowledge of τ), except for knowledge of the state at the last review, $\widetilde{\omega}_t$, and of the policy chosen at that review. Therefore, at the time of the review, the firm must choose a single decision rule, $\lambda_t(r)$, which it can then use in every subsequent period, $t+\tau$, to convert the signal into a review decision.

Conversely, if the firm had independent knowledge of τ *before* receiving the signal in each period, in principle, it could design a policy that specified a different decision rule

$\lambda_{t+\tau}(r)$ for each $\tau > 0$.

As noted above, the optimal review policy is one in which the quantity of information acquired is equal to the quantity of information used by the decision-maker, lest the decision-maker purchases superfluous information. This leads to a preliminary result from Woodford (2008).

Lemma 1 (Woodford, 2008)**.** *In the implementation of the policy specified in definition 1, the quantity of information acquired is weakly greater than the quantity of information used,*

$$J^r_{t+\tau} \leq I^r_{t+\tau} \tag{2.16}$$

Proof. See Appendix B.1. □

The quantity of information $J^r_{t+\tau}$ is the lowest quantity that the decision-maker can acquire and still make exactly the same decisions as when acquiring $I^r_{t+\tau}$. This lemma leads immediately to a redefinition of the review policy.

Corollary 1. *The most efficient* review 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 the review signal $r = 1$ (conducting a policy review) for all $\tau > 0$ and all $\widetilde{\omega}_{t+\tau}$ that follow the policy review, until the next review;*

2. *$\overline{\Lambda}_t$, the overall frequency with which the decision-maker anticipates receiving the review signal $r = 1$, until the next review.*

The quantity of information $I^r_{t+\tau}$ acquired through this signalling mechanism in each period $t + \tau$, $\tau > 0$, until the next review is equal to the quantity of information used by the decision-maker in each period, which is given in equation (2.13).

Proof. See Appendix B.1. □

Hence, the optimal review signal directly specifies whether or not the firm should review its policy, conditional on the state. This result is not only intuitive, but also formally defines the cheapest review policy that the firm can employ in order to make its review decision. Any other signal structure would require a quantity of information weakly greater than the quantity defined in equation (2.13). Reformulating the signalling mechanism in this way also leads to a simplification in solving for the firm's review decision: rather than choosing the four objects defined in definition 1, the firm chooses the sequence of hazard functions, $\{\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})\}_{\tau}$, and the anticipated frequency of policy reviews, $\overline{\Lambda}_t$.

Critical to this result is the assumption that the decision-maker can arrange to receive any type of signals from a completely unrestricted set. It may be reasonable to argue that the set of possible signals available to economic agents is at least partially restricted. Hence, the signal structure specified in corollary 1 implies a conservative estimate of the frictions generated by imperfect information. In practice, if agents are constrained in their ability to arrange signals on the state of the economy, they will obtain lower precision for a given level of expenditure on information acquisition than the precision implied by the unconstrained signalling mechanism.

2.3.2 The Pricing Policy

In each period, the price signal is received after the review signal and the associated review decision, and after the realization of the transitory shock, υ_t. As above, suppose that the firm conducts a policy review in an arbitrary state $\widetilde{\omega}_t$ in period t. The pricing policy applies starting in the same period. For any $\tau \geq 0$, let $\omega_{t+\tau} \equiv \{\widetilde{\omega}_{t+\tau}, r_{t+\tau}, \upsilon_{t+\tau}\}$ denote the state of the world at the time of the receipt of the price signal in period $t+\tau$.

As in the case of the review policy, I begin with a general definition of the pricing policy, which specifies the set of possible price signals, the probability of receiving each signal in each period and each state of the world until the next review, and an action rule that maps each signal into the price to be charged. The firm's choices for the optimal pricing policy are then simplified by showing that the most efficient price signal directly specifies the price that the firm should charge in each period. The derivation parallels that of the optimal review signal.

Definition 6. *A general* 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 that will be received until the next review;*

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

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

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

The set of possible price signals, $\mathcal{S}_t$, can also be arbitrarily large, including any variable that may be useful for determining the price to be charged conditional on the state $\omega_{t+\tau}$. For generality, I allow the decision rule for price-setting, $\alpha_t(p|s)$, to be a potentially random function of the signal. As in the case of the review signal discussed above, a single decision rule is chosen to apply across all periods; and the frequency with which the decision-maker anticipates receiving each price signal until the next review, $\overline{\phi}_t(s)$, is the relevant density that determines the quantity of information processed in every period under this policy.

The Cost of the Pricing Policy

The definition for the quantity of information that is acquired in order to implement a particular pricing policy emulates the definition in the previous section.

Definition 7. *The quantity of information that is acquired in order to implement a signalling mechanism defined by* $\left\{\mathcal{S},\ \phi\left(s|\omega\right),\ \overline{\phi}\left(s\right)\right\}$ *is* $E\left\{I^{p}\left(\phi\left(s|\omega\right),\overline{\phi}\left(s\right)\right)\right\}$, *where*

$$I^{p}\left(\phi,\overline{\phi}\right)\equiv\sum_{s\in\mathcal{S}}\phi\left(s|\omega\right)\left[\log\phi\left(s|\omega\right)-\log\overline{\phi}\left(s\right)\right]. \tag{2.17}$$

For expository purposes, $\mathcal{S}$ is a countable set, although the definition can be extended to allow for continuous signal distributions.

Using this generic definition, I next define the quantity of information that is acquired through the pricing policy specified in definition 6.

Definition 8. *The quantity of information expected, at the time of the review in an arbitrary state* $\widetilde{\omega}_{t}$ *and period* t*, to be* acquired *in the implementation of the pricing policy specified in definition 6, in each period* $t+\tau$, $\tau\geq 0$, *over all states* $\omega_{t+\tau}$ *that follow the policy review, until the next review, is given by*

$$I_{t+\tau}^{p}\equiv E_{t}\left\{I^{p}\left(\phi_{t+\tau}\left(s|\omega_{t+\tau}\right),\overline{\phi}_{t}\left(s\right)\right)\right\}, \tag{2.18}$$

where $I^{p}\left(\phi_{t+\tau}\left(s|\omega_{t+\tau}\right),\overline{\phi}_{t}\left(s\right)\right)$ *is defined in equation (2.17).*

Here, as above, $E_{t}\left\{\cdot\right\}$ denotes expectations conditional on the state $\widetilde{\omega}_{t}$, on a policy review having taken place in that state, and on the review policy implemented at that time.

The Precision of the Pricing Policy

The amount of information that is *used* by the decision-maker employing the pricing policy specified above measures the reduction in uncertainty that is reflected in the firm's prices. As in the case of the review policy, I begin by defining this quantity for a static policy, $\left\{\mathcal{S},\, \phi\left(s|\omega\right),\, \overline{\phi}\left(s\right),\, \alpha\left(p|s\right)\right\}$. Let the set of prices implied by this policy be denoted by $\mathcal{P}$. The probability that the firm charges price p in state ω is given by

$$f\left(p|\omega\right) \equiv \sum_{s\in\mathcal{S}} \alpha\left(p|s\right)\phi\left(s|\omega\right), \tag{2.19}$$

for each $p \in \mathcal{P}$.

The probability with which the firm anticipates charging price p across all states ω until the next review is

$$\overline{f}\left(p\right) \equiv \sum_{s\in\mathcal{S}} \alpha\left(p|s\right)\ \overline{\phi}\left(s\right), \tag{2.20}$$

for each $p \in \mathcal{P}$. These definitions lead to the quantity of information that is used by the decision-maker employing this pricing policy.

Definition 9. *The quantity of information that is used by a pricing policy defined by* $\{\mathcal{S},\, \phi\left(s|\omega\right),\, \overline{\phi}\left(s\right),\, \alpha\left(p|s\right)\}$ *is* $E\left\{I^p\left(f\left(p|\omega\right), \overline{f}\left(p\right)\right)\right\}$, *where*

$$I^p\left(f\left(p|\omega\right), \overline{f}\left(p\right)\right) \equiv \sum_{p\in\mathcal{P}} f\left(p|\omega\right)\left[\log f\left(p|\omega\right) - \log \overline{f}\left(p\right)\right] \tag{2.21}$$

is the relative entropy between $f\left(p|\omega\right)$ *and* $\overline{f}\left(p\right)$, *which are defined in equations (2.19) and (2.20), respectively.*

Using this definition, I now define the quantity of information expected, at the time of the review, to be used by decision-maker employing the policy given in definition 6.

Definition 10. *The quantity of information that the decision-maker expects to* use *when implementing the pricing policy given in definition 6, in each period* $t+\tau$, $\tau \geq 0$, *over* *all states $\omega_{t+\tau}$ that follow the policy review, until the next review, is given by*

$$J^p_{t+\tau} \equiv E_t \left\{ I^p \left(f_{t+\tau} \left(p | \omega_{t+\tau} \right), \overline{f_t} \left(p \right) \right) \right\}, \tag{2.22}$$

where

$$f_{t+\tau} \left(p | \omega_{t+\tau} \right) \equiv \sum_{s \in \mathcal{S}} \alpha_t \left(p | s \right) \phi_{t+\tau} \left(s | \omega_{t+\tau} \right), \tag{2.23}$$

$$\overline{f_t} \left(p \right) \equiv \sum_{s \in \mathcal{S}} \alpha_t \left(p | s \right) \overline{\phi}_t \left(s \right), \tag{2.24}$$

and $I^p \left(f_{t+\tau} \left(p | \omega_{t+\tau} \right), \overline{f_t} \left(p \right) \right)$ is defined in equation (2.21).

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

The requirement that the quantity of information acquired is equal to the quantity of information used results in a new specification of the pricing policy. First, note that, as in the case of the review policy, the general policy specified in definition 6 is suboptimal.

Lemma 2. *In the implementation of the pricing policy specified in definition 6, the quantity of information acquired is weakly greater than the quantity of information used,*

$$J^p_{t+\tau} \leq I^p_{t+\tau}. \tag{2.25}$$

Proof. See Appendix B.1. □

The quantity of information $J^p_{t+\tau}$ is the lowest quantity that the decision-maker can acquire and still choose prices according to the same rule and based on the same information as when acquiring $I^p_{t+\tau}$. As in the case of the review signal, lemma 2 leads immediately to a redefinition of the pricing policy.

Corollary 2. *The most efficient* pricing policy*, implemented following a policy review in an arbitrary state* $\widetilde{\omega}_t$ *in period* t*, specifies*

1. $\mathcal{P}_t$*, the set of prices charged until the next review;*

2. $\{f_{t+\tau}(p|\omega_{t+\tau})\}_\tau$*, the sequence of conditional probabilities of charging price* p *for all* $p \in \mathcal{P}_t$*, all* $\tau \geq 0$*, and all* $\omega_{t+\tau}$ *that follow the policy review, until the next review;*

3. $\overline{f}_t(p)$*, the anticipated frequency with which each price is charged over all states and periods until the next review, for all* $p \in \mathcal{P}_t$*.*

The quantity of information $I^p_{t+\tau}$ *acquired through this signalling mechanism in each period* $t+\tau$*,* $\tau \geq 0$*, is equal to the quantity of information used by the decision-maker in each period, which is given in equation (2.22).*

Proof. See Appendix B.1. □

Hence, the firm's optimal pricing policy directly specifies which price the firm should charge, conditional on the state.

Following a policy review in period t, the firm's choices when solving for the optimal policy are therefore reduced to choosing five objects: the sequence of hazard functions, $\{\Lambda_{t+\tau}(\widetilde{\omega}_{t+\tau})\}_\tau$, which govern the probability of a policy review in each state and in each period; the overall anticipated probability of a review, $\overline{\Lambda}_t$, the set of prices $\mathcal{P}_t$, the sequence of conditional probabilities,$\{f_{t+\tau}(p|\omega_{t+\tau})\}$, which govern the probability of charging each price in each state and period, and the overall anticipated frequency of prices, $\overline{f}_t(p)$. Any other signal structures would entail acquiring quantities of information weakly greater than the quantities given in equations (2.13) and (2.22).

2.3.3 The Firm's Problem

The choices defined above provide a concrete formulation of the firm's policy. Using these choices, the continuation value of the objective defined in equation (2.5), looking forward from the time of a policy review in an arbitrary state $\widetilde{\omega}_t$ in period t, is given by

$$E_t \left\{ \sum_{\tau=0}^{\infty} \beta^{\tau} \left[\begin{array}{c} \sum_{p \in \mathcal{P}_{t+\tau}} \pi(p - x_{t+\tau}) f_{t+\tau}\left(p|\omega_{t+\tau}\right) \\ -\beta\theta^r I^r \left(\Lambda_{t+\tau+1}\left(\widetilde{\omega}_{t+\tau+1}\right), \overline{\Lambda}_{t+\tau+1}\right) - \beta\kappa\Lambda_{t+\tau+1}\left(\widetilde{\omega}_{t+\tau+1}\right) \\ -\theta^p I^p \left(f_{t+\tau}\left(p|\omega_{t+\tau}\right), \overline{f}_{t+\tau}\left(p\right)\right) \end{array} \right] \right\}, \tag{2.26}$$

where, as above, $E_t\left\{\cdot\right\}$ denotes expectations conditional on state $\widetilde{\omega}_t$, on a review having taken place in that state, and on the review policy implemented at that time. Here, $\overline{\Lambda}_{t+\tau}$, $\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)$, $\mathcal{P}_{t+\tau}$, $\overline{f}_{t+\tau}\left(p\right)$, and $f_{t+\tau}\left(p|\omega_{t+\tau}\right)$ are the policy choices that are in effect in each future period $t+\tau$ and in each state of the world, regardless of whether they were adopted at the time of the review in period t or in some subsequent policy review. Hence, in this equation, I make no explicit reference to the period and state in which the policy that applies in each $t+\tau$ was chosen.

It is convenient to collect all of the terms in the objective that depend on the pricing policy in effect in a particular period. Let $\Pi_t\left(\omega_t\right)$ denote the firm's per-period profit in an arbitrary state ω_t (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,

$$\Pi_t\left(\omega_t\right) \equiv \sum_{p \in \mathcal{P}_t} f_t\left(p|\omega_t\right) \left\{ \pi(p - x_t) - \theta^p \left[\log f_t\left(p|\omega_t\right) - \log \overline{f}_t\left(p\right)\right] \right\}. \tag{2.27}$$

The firm's continuation value can be written more compactly, as

$$E_t \left\{ \sum_{\tau=0}^{\infty} \beta^{\tau} \left[\Pi_{t+\tau}\left(\omega_{t+\tau}\right) - \beta\theta^r I^r \left(\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right), \overline{\Lambda}_{t+\tau}\right) - \beta\kappa\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right) \right] \right\}. \tag{2.28}$$

This formulation will prove helpful later on, since it separates the firm's optimal pricing policy, which only depends on the first term.

The Recursive Formulation

The firm's review policy determines the probability that the policy chosen in period t continues to apply in period $t+\tau$, as a function of the history of states. Redefining the continuation value in terms of this survival probability is a first step towards formulating the firm's objective recursively.

Let $\Gamma_{t+\tau}\left(\widetilde{\omega}_{t+\tau-1}\right)$ 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 $\widetilde{\omega}_{t+\tau-1}$. Since there can only be one review per period, $\Gamma_{t+1}\left(\widetilde{\omega}_{t}\right) \equiv 1$ for all $\widetilde{\omega}_{t}$, and

$$\Gamma_{t+\tau}\left(\widetilde{\omega}_{t+\tau-1}\right) \equiv \prod_{k=1}^{\tau-1}\left[1-\Lambda_{t+k}\left(\widetilde{\omega}_{t+k}\right)\right], \tag{2.29}$$

for $\tau > 1$.

Let $\overline{V}_{t}\left(\widetilde{\omega}_{t}\right)$ denote the maximum attainable value of the firm's continuation value defined in equation (2.28). Under the assumption that an optimal policy will be chosen in all future policy reviews, this continuation value can be expressed in terms of the firm's choices at the time of the review in period t as

$$E_{t}\left\{\Pi_{t}\left(\omega_{t}\right)+\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma_{t+\tau}\left(\widetilde{\omega}_{t+\tau-1}\right)\left[\begin{array}{c}\left(1-\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)\right)\Pi_{t+\tau}\left(\omega_{t+\tau}\right)\\ +\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)\left[\overline{V}_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)-\kappa\right]\\ -\theta^{r}I^{r}\left(\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right),\overline{\Lambda}_{t}\right)\end{array}\right]\right\}. \tag{2.30}$$

where the continuation value now explicitly incorporates the firm's review policy, and hence $E_{t}\left\{\cdot\right\}$ now denotes expectations conditional on state $\widetilde{\omega}_{t}$ and on a review having taken place in that state. 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, $\theta^{r}I^{r}\left(\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right),\overline{\Lambda}_{t}\right)$. It then continues to apply the current policy with probability $1-\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)$, and it undertakes a policy review with probability $\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)$, in which case it pays the review cost κ and expects the maximum attainable value from that state onward, $\overline{V}_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)$.

Problem 1. *If the firm undertakes a policy review in an arbitrary state $\widetilde{\omega}_t$ and period t, it chooses*

1. *a* review policy *that specifies $\overline{\Lambda}_t$ and $\left\{\Lambda_{t+\tau}\left(\widetilde{\omega}_{t+\tau}\right)\right\}_\tau$ for all $\tau > 0$ and all $\widetilde{\omega}_{t+\tau}$ that follow the policy review, until the next review, and*
2. *a* pricing policy *that specifies $\mathcal{P}_t$, $\overline{f}_t\left(p\right)$, and $\left\{f_{t+\tau}\left(p|\omega_{t+\tau}\right)\right\}_\tau$ for all $p \in \mathcal{P}_t$, all $\tau \geq 0$, and all $\omega_{t+\tau}$ that follow the policy review, until the next review.*

The two policies are chosen to maximize the objective defined in equation (2.30).

The Stationary Formulation

At the time of a policy review in period t, the firm learns the complete state, $\widetilde{\omega}_t$. Hence, the firm's problem can be expressed in terms of the innovations to the state since the last review. Using this normalization, I formulate the firm's objective independent of the time and state in which a policy review is conducted.

First, for any state $\widetilde{\omega}_{t+\tau}$, occurring before the review decision in each period, let $\widetilde{\varpi}_\tau$ denote the innovations in $\widetilde{\omega}_{t+\tau}$ since $\widetilde{\omega}_t$. Recall that the complete state $\widetilde{\omega}_{t+\tau}$ contains the history of prior signals and decisions, through period $t+\tau-1$, and the history of permanent pre-review states, $\widetilde{x}^{t+\tau}$, that occur before each review decision, through period $t+\tau$.

Similarly, for any state $\omega_{t+\tau}$, let ϖ_τ denote the part of the state that is news since $\widetilde{\omega}_t$. The state $\omega_{t+\tau}$ contains $\widetilde{\omega}_{t+\tau}$, the review decision in period $t+\tau$, and the post-review state, $x_{t+\tau}$.

Let $\widetilde{y}_\tau$ denote the normalized pre-review state, and let y_τ denote the normalized post-review state,

$$\widetilde{y}_\tau \equiv \widetilde{x}_{t+\tau} - \widetilde{x}_t, \tag{2.31}$$

$$y_\tau \equiv x_{t+\tau} - \widetilde{x}_t. \tag{2.32}$$

Given the laws of motion in equations (2.1) and (2.2), 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 in period t. Hence, we can express the firm's problem in terms of these normalized variables, without any reference to either the date t or the state $\widetilde{\omega}_t$ in which the review takes place.

Letting q denote the normalized price,

$$q \equiv p - \widetilde{x}_t, \tag{2.33}$$

the firm's optimal *pricing policy* is expressed in terms of the set of normalized prices $q \in \mathcal{Q}$, anticipated to occur with frequencies $\overline{f}(q)$, and the sequence of conditional distributions $\{f_\tau(q|\varpi_\tau)\}_\tau$. The firm's profit function becomes $\pi(q - y_\tau)$, and y_τ is the normalized target price that the firm would charge in a frictionless environment. And $\Pi_{t+\tau}(\omega_{t+\tau})$, the expected per-period profit under the current pricing policy, net of the cost of the price signal, is replaced by its normalized counterpart $\Pi_\tau(\varpi_\tau)$,

$$\Pi_\tau(\varpi_\tau) \equiv \sum_{q \in \mathcal{Q}} f_\tau(q|\varpi_\tau) \left\{ \pi(q - y_\tau) - \theta^p \left[\log f_\tau(q|\varpi_\tau) - \log \overline{f}(q) \right] \right\}. \tag{2.34}$$

The optimal *review policy* can also be written in normalized terms as the choice of a sequence of hazard functions $\{\Lambda_\tau(\widetilde{\varpi}_\tau)\}_\tau$ and an anticipated frequency of reviews $\overline{\Lambda}$.

The survival probability becomes

$$\Gamma_{\tau+1}\left(\widetilde{\varpi}_{\tau}\right) \equiv \prod_{k=1}^{\tau}\left[1-\Lambda_{k}\left(\widetilde{\varpi}_{k}\right)\right], \tag{2.35}$$

for $\tau > 0$, with $\Gamma_1(\cdot) \equiv 1$.

These steps lead to the stationary formulation of the firm's problem, defined below.

Problem 2. *If the firm undertakes a policy review in any arbitrary state and period, it chooses*

1. *a* review policy *that specifies* $\overline{\Lambda}$ *and* $\{\Lambda_\tau(\widetilde{\varpi}_\tau)\}_\tau$ *for all* $\tau > 0$ *and all news states* $\widetilde{\varpi}_\tau$ *until the next review, and*

2. *a* pricing policy *that specifies* $\mathcal{Q}$, $\overline{f}(q)$, *and* $\{f_\tau(q|\varpi_\tau)\}_\tau$ *for all normalized prices* $q \in \mathcal{Q}$, *all* $\tau \geq 0$, *and all news states* ϖ_τ *until the next review.*

The two policies are chosen to maximize

$$E\left\{\Pi_0\left(\varpi_0\right)+\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma_{\tau}\left(\widetilde{\varpi}_{\tau-1}\right)\left[\begin{array}{c}\left(1-\Lambda_{\tau}\left(\widetilde{\varpi}_{\tau}\right)\right)\Pi_{\tau}\left(\varpi_{\tau}\right)\\ +\Lambda_{\tau}\left(\widetilde{\varpi}_{\tau}\right)\left[\overline{V}-\kappa\right]\\ -\theta^{r}I^{r}\left(\Lambda_{\tau}\left(\widetilde{\varpi}_{\tau}\right),\overline{\Lambda}\right)\end{array}\right]\right\}, \tag{2.36}$$

where $\overline{V}$ *is the maximized value of the objective defined in (2.36).*

2.4 Optimal Policy

I obtain the solution to Problem 2 in steps, deriving each element of the optimal policy taking the other elements as given.

Proposition 1. *The policy that maximizes the objective defined in equation (2.36) specifies*

1. *a* review policy *given by a scalar,* $\overline{\Lambda}$*, that denotes the frequency of reviews, and a single real-valued function,* $\Lambda(\widetilde{y})$*, that is defined for all possible normalized pre-review target prices* $\widetilde{y}$*, and that determines the probability of a policy review in each state and period* τ*, and*

2. *a* pricing policy *given by the set of normalized prices* $\mathcal{Q}$*, the distribution* $\overline{f}(q)$*, for all* $q \in \mathcal{Q}$*, and a single conditional distribution,* $f(q|y)$*, that is defined for all possible normalized post-review target prices* y*, and that determines the probability of charging the normalized price* q *in each state and period* τ*.*

The hazard function for policy reviews, $\Lambda(\widetilde{y})$*, is given by*

$$\frac{\Lambda(\widetilde{y})}{1-\Lambda(\widetilde{y})} = \frac{\overline{\Lambda}}{1-\overline{\Lambda}} \exp\left\{\frac{1}{\theta^r}\left[\overline{V} - \kappa - V(\widetilde{y})\right]\right\}, \tag{2.37}$$

for each $\widetilde{y}$*, where* $V(\widetilde{y})$ *is the firm's continuation value under the firm's current policy, and* $\overline{V} = V(0)$ *is the firm's continuation value upon conducting a policy review.*

The frequency of policy reviews, $\overline{\Lambda}$*, implied by the hazard function, is given by*

$$\overline{\Lambda} = \frac{E\left\{\sum_{\tau=1}^{\infty} \beta^\tau \Gamma_\tau\left(\widetilde{y}^{\tau-1}\right)\Lambda\left(\widetilde{y}_\tau\right)\right\}}{E\left\{\sum_{\tau=1}^{\infty} \beta^\tau \Gamma_\tau\left(\widetilde{y}^{\tau-1}\right)\right\}}. \tag{2.38}$$

The conditional distribution of prices, $f(q|y)$*, is given by*

$$f(q|y) = \overline{f}(q)\frac{\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}} \overline{f}(\widehat{q})\exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}}, \tag{2.39}$$

for each y*.*

The unconditional distribution of prices, $\overline{f}(q)$*, is given by*

$$\overline{f}(q) = \frac{E\left\{\sum_{\tau=0}^{\infty} \beta^\tau \Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right) f(q|y)\right\}}{E\left\{\sum_{\tau=0}^{\infty} \beta^\tau \Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)\right\}}, \tag{2.40}$$

The support of the distribution of prices, $\mathcal{Q}$, is determined from the pair of conditions,

$$Z\left(q;\overline{f}\right)=1,\ \forall q\in\mathcal{Q}, \tag{2.41}$$

$$Z\left(q;\overline{f}\right)\leq 1,\quad \forall q \tag{2.42}$$

where

$$Z\left(q;\overline{f}\right)\equiv\int\frac{\exp\left\{\frac{1}{\theta^{p}}\pi\left(q-y\right)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}}\overline{f}\left(\widehat{q}\right)\exp\left\{\frac{1}{\theta^{p}}\pi\left(\widehat{q}-y\right)\right\}}G\left(y\right)dy. \tag{2.43}$$

Here, $G\left(y\right)$ is the distribution of normalized target prices, y, under the firm's review policy, and given the laws of motion in equations (2.31) and (2.32). This distribution will be derived explicitly below.

The first thing to note is that the optimal policy specifies both a review policy and a pricing policy that condition only on the normalized target prices $\widetilde{y}_{\tau}$ and y_{τ}. Although I allow the firm to condition each component of its policy on the complete state ($\widetilde{\varpi}_{\tau}$, for the review decision, and ϖ_{τ}, for the pricing decision), the firm allocates all the information capacity to learning about changes in market conditions since the last review, rather than paying any attention to past events, past signals, or the passage of time. This outcome is a result of the setup that implies that all types of information have equal cost. Given this fixed cost, and since the firm would like to have information regarding past events or regarding the passage of time only insofar as it is informative about the current normalized state, the firm chooses to learn directly about the target price that directly affects its profit function.

The second thing to note 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 hazard functions and conditional price distributions that are indexed by time. This outcome is a direct consequence of the first point discussed above. Since the firm chooses to directly learn about the normalized target price ($\widetilde{y}_{\tau}$, for the review decision, and y_{τ}, for the pricing decision), its signal extraction 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 decision is expected to be made over the life of the policy.

The firm's complete policy is derived in the following subsections, which present the optimization problem that each element of the policy solves.

2.4.1 The Conditional Distribution of Prices

For the purposes of showing that the firm allocates its entire attention to monitoring only innovations in target price, it convenient to begin by discussing the firm's pricing policy, taking the review policy as given.

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. Inspection of equation (2.36) reveals that, excluding the terms that depend only on the review policy, the pricing objective is the maximization of

$$E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma_{\tau+1}\left(\widetilde{\varpi}_{\tau}\right)\Pi_{\tau}\left(\varpi_{\tau}\right)\right\}. \tag{2.44}$$

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, $\mathcal{Q}$, and the anticipated frequency with which each price is charged, $\overline{f}\left(q\right)>0$ for all $q\in\mathcal{Q}$. The objective specified in equation (2.44) is additively separable across dates and states. Hence, 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

$$\max_{f_\tau(q|\varpi_\tau)} \Pi_\tau\left(\varpi_\tau\right) \tag{2.45}$$

$$s.t. \sum_{q\in\mathcal{Q}} f_\tau\left(q|\varpi_\tau\right) = 1, \tag{2.46}$$

$$f_\tau\left(q|\varpi_\tau\right) \geq 0,\ \forall q \in \mathcal{Q}. \tag{2.47}$$

Lemma 3. *Let the review policy, the set of normalized prices $\mathcal{Q}$, and the frequency with which the firm anticipates charging each price until the next review, $\overline{f}\left(q\right)$ for all $q \in \mathcal{Q}$, be fixed. If $\theta^p > 0$, the optimal conditional price distribution $f_\tau\left(q|\varpi_\tau\right)$ that solves the objective defined in equation (2.45) subject to the constraints specified in equations (2.46) and (2.47), has the same form for all $\tau \geq 0$, and depends only on the normalized target price, y_τ, for each ϖ_τ. Letting $\mathcal{Y}$ denote the set of all possible values of y_τ under the current review policy, the optimal conditional price distribution is given by*

$$f\left(q|y\right) = \overline{f}\left(q\right)\frac{\exp\left\{\frac{1}{\theta^p}\pi\left(q-y\right)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}}\overline{f}\left(\widehat{q}\right)\exp\left\{\frac{1}{\theta^p}\pi\left(\widehat{q}-y\right)\right\}}, \tag{2.48}$$

for all $y \in \mathcal{Y}$, and all $q \in \mathcal{Q}$.

Proof. See Appendix B.1. □

Note that since the expected profit in each period $\Pi_\tau\left(\varpi_\tau\right)$ depends on ϖ_τ and τ only through the dependence of the profit function $\pi(q - y_\tau)$ on the target price y_τ, the optimal pricing policy conditions prices only on y_τ. The resulting conditional distribution, $f\left(q|y_\tau\right)$, indicates the probability of charging a normalized price q when the target price is y_τ, and is otherwise independent of the number of periods elapsed since the last review and of all other aspects of ϖ_τ. Intuitively, since the firm faces the same unit cost of processing information about all aspects of the complete state ϖ_τ, it chooses to allocate its entire attention to monitoring changes in its target price directly, and it does so in the same way in each period until the next review. Furthermore, note that the formula for the optimal conditional distribution in this dynamic setting is of the same form as that first derived by Shannon (1959) for a static rate-distortion function.

The resulting expected profit net of the cost of the price signal is also a time-invariant function of the target price in each period, $\Pi(y_\tau)$, where

$$\Pi(y) \equiv \sum_{q \in \mathcal{Q}} f(q|y) \left\{ \pi(q-y) - \theta^p \left[\log f(q|y) - \log \overline{f}(q) \right] \right\}. \tag{2.49}$$

Equation (2.48) illustrates the sense in which the price signal is optimally designed, given the firm's objective function. For a given target price y, the conditional probability of charging a particular price q that is closer to the target is relatively higher, since the profit under that price is high relative to the average profit that the firm can expect in this state across all normalized prices in the set $\mathcal{Q}$. However, the relationship between the state and the price signal is noisy: the signal places positive mass on all prices in the support of the distribution, for each target price y. This randomness reflects the need to economize on the information cost associated with receiving the price signal in each period.

If, before receiving the price signal in each period, the firm had independent knowledge of the number of periods elapsed since the last review, it would have more precise information about the states of the world that are more likely in a particular period. For example, it would have less uncertainty about the state soon after a review. It would use this knowledge to design a signalling mechanism that specified different anticipated frequencies for the normalized prices for each period, $\overline{f}_\tau(q)$, and hence different conditional distributions, $f_\tau(q|y)$. Similarly, if the firm had access to the sequence of past price signals for free, a separate signalling mechanism would also be chosen for each history of prior signals. Here, instead, the only information that the firm has, prior to the receipt of the price signal, is the information obtained at the last policy review. Hence, the optimal pricing policy is characterized by a single conditional distribution that is optimal across all states and periods in which the current policy is expected to apply.

2.4.2 The Hazard Function for Reviews

I consider next the firm's choice of an optimal sequence of hazard functions 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, defined in equation (2.36), from some period τ onwards,

$$E_\tau \left\{ \Pi_\tau(\varpi_\tau) + \sum_{\tau'=\tau+1}^{\infty} \beta^{\tau'-\tau} \Gamma_{\tau,\tau'}(\widetilde{\varpi}_{\tau'-1}) \left[\begin{array}{c} (1-\Lambda_{\tau'}(\widetilde{\varpi}_{\tau'}))\,\Pi_{\tau'}(\varpi_{\tau'}) \\ +\Lambda_{\tau'}(\widetilde{\varpi}_{\tau'})\left[\overline{V}-\kappa\right] \\ -\theta^r I^r\left(\Lambda_{\tau'}(\widetilde{\varpi}_{\tau'}),\overline{\Lambda}\right) \end{array} \right] \right\}, \tag{2.50}$$

where $E_\tau\{\cdot\}$ denotes expectations over all possible histories for dates $\tau' \geq \tau$, conditional on reaching state $\widetilde{\varpi}_\tau$, and, using equation (2.35), the survival probability between periods τ and τ' is given by

$$\Gamma_{\tau,\tau'}(\widetilde{\varpi}_{\tau'-1}) \equiv \prod_{k=\tau+1}^{\tau'-1} \left[1-\Lambda_k(\widetilde{\varpi}_k)\right], \tag{2.51}$$

for $\tau' > \tau+1$, with $\Gamma_{\tau,\tau+1}(\cdot) \equiv 1$.

The optimal sequence of hazard functions $\{\Lambda_{\tau'}(\widetilde{\varpi}_{\tau'})\}_{\tau'}$ maximizes the continuation value defined in equation (2.50), given $\overline{\Lambda}$ and the firm's pricing policy.

Lemma 4. *Let the pricing policy be fixed, and let the conditional price distribution in each period be of the form specified in equation (2.48). Let the anticipated frequency of reviews be fixed at some $\overline{\Lambda} \in (0,1)$. For $\theta^r > 0$, the hazard functions in the optimal sequence that maximizes equation (2.50) have the same form for all $\tau > 0$ and depend* *only on the target price at the time of receipt of the review signal,* $\widetilde{y}_\tau$*, for each* $\widetilde{\varpi}_\tau$*. Furthermore, the maximum attainable value of the objective defined in equation (2.50) is also a time-invariant function,* $V(\widetilde{y})$*.*

Letting $\widetilde{\mathcal{Y}}$ *denote the set of all possible values of* $\widetilde{y}_\tau$ *under the current review policy, the optimal hazard function is given by*

$$\frac{\Lambda(\widetilde{y})}{1-\Lambda(\widetilde{y})} = \frac{\overline{\Lambda}}{1-\overline{\Lambda}} \exp\left\{\frac{1}{\theta^r}\left[\overline{V} - \kappa - V(\widetilde{y})\right]\right\}, \tag{2.52}$$

for all $\widetilde{y} \in \widetilde{\mathcal{Y}}$*, where the function* $V(\widetilde{y})$ *satisfies the fixed point equation*

$$V(\widetilde{y}) = E\left\{\Pi(y) + \beta\left[(1-\Lambda(\widetilde{y}'))V(\widetilde{y}') + \Lambda(\widetilde{y}')\left(\overline{V}-\kappa\right) - \theta^r I^r\left(\Lambda(\widetilde{y}'), \overline{\Lambda}\right)\right]\right\}, \tag{2.53}$$

and where the continuation value upon conducting a policy review satisfies $\overline{V} = V(0)$*. Here,* $E\{\cdot\}$ *denotes expectations over the possible values of* $y_\tau = y$ *and* $\widetilde{y}_{\tau+1} = \widetilde{y}'$ *conditional on* $\widetilde{y}_\tau = \widetilde{y}$*.*

Proof. See Appendix B.1. □

The fact that the hazard function for a policy review only depends on the target price results from the fact that, as derived in the previous section, the firm's pricing policy specifies a time-invariant conditional distribution that itself only conditions on the target price. Lemmas 3 and 4 establish that for both the review decision and the pricing decision, the firm chooses to acquire information only about the innovations to the target price between the last review and the receipt of the signal for each decision, and it allocates no capacity to learning about past events. Moreover, these lemmas show that the optimal policy has the same form for all periods between reviews, with a single conditional distribution of prices and a single hazard function characterizing the review decision and the pricing decision in each state and period between reviews. Finally, expressed in terms of the normalized variables, q and y, the same policy is chosen at each review. At each review, the distribution of actual prices charged, p, is shifted by the innovation to the target price between reviews.

Expression (2.52) is of the same form as the optimal hazard function derived in Woodford (2008) for the case in which the pricing policy is reduced to a single price. The hazard function is monotonically increasing in the value of the exponent: a higher value of adjustment relative to keeping the policy unchanged is associated with a higher probability of receiving a signal that the policy should be reviewed. However, the relationship between the state and the review decision is noisy. For any overall frequency of policy reviews, $\overline{\Lambda} \in (0,1)$, the hazard function satisfies $\Lambda(\widetilde{y}) \in (0,1)$. In order to economize on information costs, the optimal review signal never indicates a review with certainty.[15]

The hazard function implies a survival probability that depends only on the history of the pre-review target prices, $\widetilde{y}^{\tau-1}$,

$$\Gamma_\tau\left(\widetilde{y}^{\tau-1}\right) = \prod_{k=1}^{\tau-1}\left[1 - \Lambda\left(\widetilde{y}_k\right)\right], \tag{2.54}$$

for $\tau > 1$, with $\Gamma_1(0) \equiv 1$. I shall use this survival probability in the next section, to determine the optimal frequency of reviews, $\overline{\Lambda}$.

2.4.3 The Frequency of Reviews

For a given pricing policy, and a given hazard function for policy reviews, the optimal frequency of reviews, $\overline{\Lambda}$, is chosen to maximize the objective specified in equation (2.36). Using the results of the previous two sections, and excluding the first term, which is

[15] See Woodford (2008) for a proof of the optimality of noise in the review signal in the case of a single-price policy.

independent of the review policy, $\overline{\Lambda}$ is chosen to maximize

$$E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma_{\tau}\left(\widetilde{y}^{\tau-1}\right)\left[\begin{array}{c}\left(1-\Lambda\left(\widetilde{y}_{\tau}\right)\right)\Pi\left(y_{\tau}\right)\\ +\Lambda\left(\widetilde{y}_{\tau}\right)\left[\overline{V}-\kappa\right]\\ -\theta^{r}I^{r}\left(\Lambda\left(\widetilde{y}_{\tau}\right),\overline{\Lambda}\right)\end{array}\right]\right\}. \tag{2.55}$$

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

$$\min_{\overline{\Lambda}} E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma_{\tau}\left(\widetilde{y}^{\tau-1}\right)I^{r}\left(\Lambda\left(\widetilde{y}_{\tau}\right),\overline{\Lambda}\right)\right\}, \tag{2.56}$$

where the quantity of information acquired in each period, $I^{r}\left(\Lambda\left(\widetilde{y}_{\tau}\right),\overline{\Lambda}\right)$, is given by the function defined in equation (2.12).

Lemma 5. *Let the firm's pricing policy be fixed. Let the hazard function for policy reviews, $\Lambda\left(\widetilde{y}\right)$, be fixed and of the form given in lemma 4. The anticipated frequency of policy reviews, $\overline{\Lambda}\in(0,1)$, that minimizes the cost of the review policy, given in equation (2.56), is*

$$\overline{\Lambda}=\frac{E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma_{\tau}\left(\widetilde{y}^{\tau-1}\right)\Lambda\left(\widetilde{y}_{\tau}\right)\right\}}{E\left\{\sum_{\tau=1}^{\infty}\beta^{\tau}\Gamma_{\tau}\left(\widetilde{y}^{\tau-1}\right)\right\}}. \tag{2.57}$$

Proof. See Appendix B.1. □

Equation (2.57) shows that the optimal anticipated frequency of reviews is equal to the (discounted) weighted average of the conditional probabilities of reviews across all the pre-review target prices $\widetilde{y}_{\tau}$ that the firm expects to encounter over the life of the policy. Recall that before the receipt of the review signal in each period, the firm anticipates undertaking a review with this "default" probability. The amount of information acquired in each period is the relative entropy between the conditional probability of a review, $\Lambda\left(\widetilde{y}\right)$, and this anticipated probability, $\overline{\Lambda}$.

The hazard function $\Lambda(\widetilde{y})$, together with the laws of motion for the innovations in the pre-review state $\widetilde{y}_\tau$, determine the distribution of pre-review states that the firm expects to encounter under the current review policy, which yields the following additional result.

Lemma 6. *Let the hazard function for policy reviews, $\Lambda(\widetilde{y})$, be fixed and of the form given in lemma 4. Let the distribution of pre-review target prices in period τ, under this hazard function, be denoted by $\widetilde{g}_\tau(\widetilde{y})$, given by $h_{\widetilde{\nu}}(\widetilde{y})$ for $\tau = 1$, and by*

$$\widetilde{g}_\tau(\widetilde{y}) = \int [1 - \Lambda(\widetilde{y} - \widetilde{\nu})]\, \widetilde{g}_{\tau-1}(\widetilde{y} - \widetilde{\nu})\, h_{\widetilde{\nu}}(\widetilde{\nu})\, d\widetilde{\nu}, \tag{2.58}$$

normalized to sum to one, for $\tau > 1$, where $h_{\widetilde{\nu}}$ is the distribution of the permanent innovation, $\widetilde{\nu}$. Let the discounted distribution of pre-review target prices over the life of the policy, $\widetilde{y} \in \widetilde{\mathcal{Y}}$, be denoted by $\widetilde{G}(\widetilde{y})$,

$$\widetilde{G}(\widetilde{y}) = (1-\beta) \sum_{\tau=1}^{\infty} \beta^\tau \widetilde{g}_\tau(\widetilde{y}). \tag{2.59}$$

Then, the optimal frequency with which the decision-maker anticipates undertaking reviews at the time of a policy review is equal to the frequency of reviews induced by $\Lambda(\widetilde{y})$,

$$\overline{\Lambda} = \int \Lambda(\widetilde{y})\, \widetilde{G}(\widetilde{y})\, d\widetilde{y}. \tag{2.60}$$

Proof. See Appendix B.1. □

Lemmas 4 and 5 provide a complete characterization of the firm's *review policy*, for a given pricing policy. Lemma 6 establishes that the review signal thus specified is rational in that the anticipated frequency of reviews coincides with the realized frequency of reviews.

I next characterize the remaining elements of the firm's policy, the anticipated frequency of prices, $\overline{f}(q)$, for all $q \in \mathcal{Q}$, and the optimal support $\mathcal{Q}$.

2.4.4 The Frequency of Prices

The firm's pricing policy maximizes (2.44), which, given the results in the previous sections, becomes

$$E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)\Pi\left(y\right)\right\}, \tag{2.61}$$

where, the expected per-period profit net of the cost of the price signal is

$$\Pi\left(y\right)\equiv\sum_{q\in\mathcal{Q}}f\left(q|y\right)\left\{\pi(q-y)-\theta^{p}\left[\log f\left(q|y\right)-\log\overline{f}\left(q\right)\right]\right\}, \tag{2.62}$$

reproduced here from an earlier section, for clarity.

Holding fixed the review policy, the support of the price signal, $\mathcal{Q}$, and the conditional price distribution $f\left(q|y\right)$, 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, $\overline{f}\left(q\right)$ solves

$$\min_{\overline{f}(q)}E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)\left[\sum_{q\in\mathcal{Q}}f\left(q|y\right)\left[\log f\left(q|y\right)-\log\overline{f}\left(q\right)\right]\right]\right\} \tag{2.63}$$

$$s.t.\sum_{q\in\mathcal{Q}}\overline{f}\left(q\right)=1, \tag{2.64}$$

just as the frequency of reviews, $\overline{\Lambda}$, was shown to minimize the cost of the review signal.

Lemma 7. *Let the review policy and the set of normalized prices, $\mathcal{Q}$, be fixed. Let the conditional probability of charging each price, $f\left(q|y\right)$, and the hazard function for policy reviews, $\Lambda\left(\widetilde{y}\right)$, be of the form given in lemmas 3 and 4, respectively. The anticipated frequency of normalized prices, $\overline{f}\left(q\right)$, that minimizes the cost of the pricing policy, given in equation (2.63), subject to the constraint specified in equation (2.64) is*

$$\overline{f}\left(q\right)=\frac{E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)f\left(q|y\right)\right\}}{E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)\right\}}, \tag{2.65}$$

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

Proof. See Appendix B.1. □

The optimal anticipated frequency of prices is equal to the (discounted) weighted average of the conditional price distribution over all 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. Before receiving the price signal in each period, the firm anticipates a signal drawn from this default distribution. Hence, the amount of information that is received in each period is equal to the relative entropy between this distribution and the conditional distribution, $f(q|y)$.

As in the case of the review signal, the frequency with which the firm anticipates to charge each price $q \in \mathcal{Q}$ is equal to the realized frequency, integrating over the distribution of possible target prices that the firm can expect under its review policy.

Lemma 8. *Let the hazard function for policy reviews, $\Lambda(\widetilde{y})$, and the associated distribution of pre-review target prices in each period, $\widetilde{g}_\tau(\widetilde{y})$, be fixed and of the form given in lemmas 4 and 6. Let the distribution of post-review target prices in period τ be denoted by $g_\tau(y)$, given by $h_\nu(y)$ for $\tau = 0$, and by*

$$g_\tau(y) = \int [1 - \Lambda(y-\nu)]\, \widetilde{g}_\tau(y-\nu)\, h_\nu(\nu)\, d\nu, \tag{2.66}$$

normalized to sum to one, for $\tau > 0$, where h_ν is the distribution of the transitory innovation, ν. Let the discounted distribution of post-review target prices over the life of the policy, $y \in \mathcal{Y}$, be denoted by $G(y)$,

$$G(y) = (1-\beta) \sum_{\tau=0}^{\infty} \beta^\tau g_\tau(y). \tag{2.67}$$

Then, the optimal frequency with which the decision-maker anticipates charging each price over the life of the policy at the time of a policy review is the marginal distribution *corresponding to* $f(q|y)$,

$$\overline{f}(q) = \int f(q|y)\, G(y)\, dy, \tag{2.68}$$

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

Proof. See Appendix B.1. □

Hence, before the receipt of the price signal in any period and state of the world, the firm anticipates receiving a signal from this "default" distribution, and this distribution coincides with the realized distribution of prices over the life of the policy.

2.4.5 The Support of the Price Distribution

In this section, I consider the choice of the optimal support, given the form of the pricing policy obtained above. I first establish the necessary and sufficient conditions that determine if a given support is optimal. I then show how to *find* the support using 1) sufficient conditions for the points of support for a given cardinality, and 2) necessary and sufficient conditions that determine the cardinality. In the derivations that follow, I refer to results from the information theory literature, in particular, Shannon (1959), Blahut (1972), Fix (1978), and Rose (1994).

Using lemma 8, the part of the objective that depends on the firm's pricing policy can be written directly in terms of the distribution of normalized target prices, $G(y)$, as

$$\int G(y)\, \Pi(y)\, dy. \tag{2.69}$$

Note that through this formulation, the dynamic problem presented in equation (2.44) has been transformed into a static rational inattention problem for a distribution of states given by $G(y)$. Equations (2.48) and (2.68), which characterize $f(q|y)$ and $\overline{f}(q)$ for a given support, have the same form as the equations that characterize the solution to the static rate distortion problem for a memoryless source (Shannon, 1959).

Given the definition of $\Pi(y)$, the pricing objective specified in equation (2.69) is strictly concave in both $f(q|y)$ and $\overline{f}(q)$. Therefore, equations (2.48) and (2.68) describe the optimal policy on a fixed support, Q.

Lemma 9. *Let the distribution of states, $G(y)$, and the set $\mathcal{Q}$ be fixed. Then if $f(q|y)$ and $\overline{f}(q)$ are probability distributions such that $\overline{f}(q) > 0$ for all $q \in \mathcal{Q}$, and such that equations (2.48) and (2.68) are satisfied for all $y \in \mathcal{Y}$ and $q \in \mathcal{Q}$, these distributions specify the unique optimal pricing policy among all pricing policies with support $\mathcal{Q}$.*

Proof. See Appendix B.1. □

Conditions (2.48) and (2.68) cannot rule out the existence of some other price, $\widehat{q} \notin \mathcal{Q}$, that would be charged with positive probability under the optimal policy. This is because in order to derive these conditions, it has been convenient thus far to temporarily ignore the constraint $\overline{f}(q) \geq 0$.

In order to derive the set of necessary and sufficient conditions for the optimal support, consider the firm's pricing objective after substituting in the optimal conditional distribution, $f(q|y)$, for a given marginal, $\overline{f}(q)$. Using the solution given in equation (2.48), this objective becomes a function of $\overline{f}(q)$,

$$\mathcal{F}\left(\overline{f}\right) \equiv \int G(y) \log \left[\sum_{q \in \mathcal{Q}} \overline{f}(q) \exp \left\{ \frac{1}{\theta^p} \pi (q - y) \right\} \right] dy. \tag{2.70}$$

The distribution $\overline{f}(q)$ must then maximize this objective subject to

$$\sum_{q \in \mathcal{Q}} \overline{f}(q) = 1, \tag{2.71}$$

$$\overline{f}(q) \geq 0, \quad \forall q. \tag{2.72}$$

This maximization problem yields the following result (see also results for the static rate distortion problem in the information theory literature, e.g., Blahut, 1972).

Lemma 10. *Let the distribution of states,* $G(y)$*, be fixed, and let the probability distributions* $f(q|y)$ *and* $\overline{f}(q)$ *satisfy (2.48) and (2.68) for all* $y \in \mathcal{Y}$ *and* $q \in \mathcal{Q}$*. Let the functional* $Z\left(q;\overline{f}\right)$ *be defined as*

$$Z\left(q;\overline{f}\right) \equiv \int G(y) \frac{\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}} dy. \tag{2.73}$$

The set $\mathcal{Q}$ *is the optimal support of the pricing policy if and only if*

$$Z\left(q;\overline{f}\right) \begin{cases} = 1 & \text{if} \quad q \in \mathcal{Q}, \\ \leq 1 & \text{if} \quad q \notin \mathcal{Q}. \end{cases} \tag{2.74}$$

Proof. See Appendix B.1. □

If one can find a set of prices $\mathcal{Q}$ that satisfy the conditions of lemma 10, then this set characterizes the uniquely optimal solution at the information cost θ^p. Before discussing how to find $\mathcal{Q}$, it is useful to consider what type of solution one might expect from lemma 10. Here I present a result from Fix (1978), extended to apply to the setup in this paper.

Lemma 11 (Fix, 1978)**.** *The optimal support* $\mathcal{Q}$ *is either the entire real line or a discrete set of prices.*

Proof. See Appendix B.1. □

The solution is continuous if and only if $Z\left(q;\overline{f}\right) = 1$ for all $q \in \mathbb{R}$, namely, $\mathcal{Q} = \mathbb{R}$. In this case, equations (2.48) and (2.68) 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 satisfies equations (2.48) and (2.68), but which yields either $\overline{f}(q) = 0$ or $Z\left(q;\overline{f}\right) < 1$

for any point in this set. This is the key insight that will allow me to prove that for certain parametrizations, the solution is discrete.

Finally, note that the solution cannot consist of disjoint intervals. Intuitively, there cannot be "holes" in the support of the signal unless the support is discrete, because the firm's average profits 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. Matejka and Sims (2010) independently prove a similar result through a slightly different approach. [16]

Following Rose (1994), one can establish a pair of useful necessary conditions.

Lemma 12 (necessary conditions)**.** *Let $f(q|y)$ and $\overline{f}(q)$ satisfy the optimality conditions in equations (2.48) and (2.68). The points of support must satisfy*

$$\int G(y|q) \frac{\partial \pi(q-y)}{\partial q} dy = 0, \tag{2.75}$$

$$\int G(y|q) \left[\frac{\partial^2 \pi(q-y)}{\partial q^2} + \frac{1}{\theta^p} \left(\frac{\partial \pi(q-y)}{\partial q} \right)^2 \right] dy \leq 0, \tag{2.76}$$

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

Proof. See Appendix B.1. □

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 $\mathcal{Q}$.

[16] Note that if $G(y)$ is a distribution with bounded support, then the support $\mathcal{Q}$ will necessarily be bounded as well (it is not efficient to arrange to receive a signal outside the support of the state). Fix's result then implies that the support $\mathcal{Q}$ is discrete. Fix draws out this implication, and Matejka and Sims (2010) also show the discreteness of the solution for problems with bounded support.

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 (2.76), then the size of the set $\mathcal{Q}$ must be increased. In practice, I shall use the necessary and sufficient condition in lemma 10 directly in order to determine the optimal cardinality of the solution. However, equation (2.76) does provide a way to verify if, for a given information cost, the solution must necessarily involve more than one price, as discussed further below.

The method for finding the optimal pricing policy can be summarized in three steps:

1. Initialize the cardinality of $\mathcal{Q}$;

2. Iterate to convergence between equation (2.75), given $f(q|y)$, and equations (2.48), (2.68), given $\mathcal{Q}$;

3. Check the conditions in lemma 10: if $\overline{f}(q) = 0$ or $Z\left(q; \overline{f}\right) < 1$ for any $q \in \mathcal{Q}$, decrease the cardinality of $\mathcal{Q}$ and return to step 2; otherwise, if $Z\left(q; \overline{f}\right) = 1$ for all $q \in \mathcal{Q}$ and $Z\left(q; \overline{f}\right) > 1$ for some $q \notin \mathcal{Q}$, increase the cardinality of $\mathcal{Q}$ and return to step 2.

Appendix B.2 discusses the numerical implementation in detail.

2.4.6 Evolution of the Optimal Support

In this section, I illustrate how the firm's pricing policy evolves as a function of the cost of the price signal, θ^p, keeping the review policy fixed. Specifically, I show the optimality of single-price and multiple-price policies for different ranges of θ^p, and illustrate how the cardinality of the solution increases as the cost of information is decreased. The approach echoes the discussions in Fix (1978) and Rose (1994).

The numerical results are generated for a profit function and a distribution $G(y)$ that are discussed in section 2.5. Table 18 summarizes the different stages and types of pricing policies that are optimal for different values of θ^p, given the particular shape of the profit function and the particular shape and dispersion of the distribution of possible states.

Single-Price Policy

A single-price policy, if optimal, is defined by the price

$$\overline{q} = \arg\max_q \int G(y)\, \pi(q-y)\, dy. \tag{2.77}$$

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

$$\overline{\theta}^p \equiv \frac{\int G(y) \left(\frac{\partial}{\partial q} \pi(q-y) \right)^2 dy}{\int G(y) \left(\frac{\partial^2}{\partial q^2} \pi(q-y) \right) dy}, \tag{2.78}$$

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

Per lemma 10, the single price policy is optimal for a given θ^p if and only if

$$Z(q) \equiv \int G(y) \exp\left\{ \frac{1}{\theta^p} \left[\pi(q-y) - \pi(\overline{q}-y) \right] \right\} dy \leq 1 \tag{2.79}$$

for all $q \neq \overline{q}$.[17] The function $Z(q)$ is decreasing in θ^p, therefore, if the inequality is satisfied for some $\widehat{\theta}^p$, it is satisfied for any $\theta^p \geq \widehat{\theta}^p$.

Since both $\overline{q}$ and $\overline{\theta}^p$ are determined by the shape of the firm's profit function and the distribution $G(y)$, which are both given for now, I shall take these values as fixed, and discuss the firm's pricing policy in terms of prices and information costs that are expressed relative to these values.

[17]Note that $Z(\overline{q}) = 1$ is trivially satisfied.

Solution (SPP). *For the profit function $\pi(q-y)$ and the distribution of states $G(y)$ specified in section 2.5, the single-price policy $\mathcal{Q}=\{\overline{q}\}$ is optimal for all $\theta^p \geq 1.66\,\overline{\theta}^p$. At $\theta^p = 1.65\,\overline{\theta}^p$ a single new mass point has reached critical mass such that the single-price policy is no longer optimal.*

Consider first the optimal pricing policy for an arbitrary high value of the information cost, $\theta^p = 2\,\overline{\theta}^p$. Solving equations (2.75),(2.48) and (2.68) for an arbitrary set of prices around $\overline{q}$, the solution converges to $\mathcal{Q}=\{\overline{q}\}$. The first panel in figure 12 plots the function $Z(q)-1$ for a range of values of q at this high level of the cost of the price signal. The function is below zero everywhere except at $\overline{q}$. Hence at $\theta^p = 2\,\overline{\theta}^p$ not only is the solution discrete, but it is in fact a single-price solution. The single-price policy is verified to remain optimal until $\theta^p = 1.66\,\overline{\theta}^p$.

At $\theta^p = 1.65\,\overline{\theta}^p$, a new mass point has reached critical mass, such that condition (2.79) is no longer satisfied. The panels in figure 12 show the growth of this new mass point between $\theta^p = 2\,\overline{\theta}^p$ and $\theta^p = 1.65\,\overline{\theta}^p$. Note how as the information cost falls, the function $Z(q)-1$ 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. Moreover, I verify that there is no other fast-growing area over the entire possible range of q, so the transition from the single-price policy to the multiple-price policy occurs with the growth a *single* new mass point.

Multiple-Price Policies

A two-price policy, if it exists, is characterized by the set $\mathcal{Q}=\{q_1,q_2\}$, with the associated distributions determined in equations (2.48) and (2.68), $f(q_i|y)$, $\overline{f}(q_i)$, for $i=1,2$. Each price is given by

$$q_i = \arg\max_q \int G(y|q_i)\,\pi(q-y)\,dy. \tag{2.80}$$

The threshold cost of the price signal that is sufficiently low such that a policy consisting of only two prices is no longer optimal is given by

$$\overline{\theta}^{2p} \equiv \max\left\{\theta_1^p, \theta_2^p\right\}, \tag{2.81}$$

where

$$\theta_i^p \equiv \frac{\int G\left(y|q_i\right)\left(\frac{\partial}{\partial q}\pi\left(q-y\right)\right)^2 dy}{\int G\left(y|q_i\right)\left(\frac{\partial^2}{\partial q^2}\pi\left(q-y\right)\right) dy}, \tag{2.82}$$

with the derivatives evaluated at q_i, for each $i = 1, 2$.

This policy is optimal at a given θ^p if and only if (1) $Z\left(q_i; \overline{f}\right) = 1$ for each $i = 1, 2$, and (2) $Z\left(q; \overline{f}\right) \leq 1$ for all $q \neq q_i$.

Solution (2PP). *For the profit function $\pi\left(q-y\right)$ and the distribution of states $G\left(y\right)$ specified in section 2.5, the two-price policy $\mathcal{Q} = \left\{q_1, q_2\right\}$ is optimal for all $1.65\,\overline{\theta}^p \leq \theta^p \leq 0.87\,\overline{\theta}^p$. For $\theta^p \geq 1.66\,\overline{\theta}^p$, no two-price policy price can be found that satisfies condition (1). For $\theta^p \leq 0.865\,\overline{\theta}^p$, no two-price policy can be found that satisfies condition (2); at $\theta^p = 0.865\,\overline{\theta}^p$, a new mass point has reached critical mass.*

Similar to the transition from the one-price policy to the two-price policy, there is a range of values for θ^p, over which a third mass point grows. The panels in figure 13 show the evolution of the two-price policy and the growth of this new mass point as the cost of the price signal is reduced.

Solution (3PP). *For the profit function $\pi\left(q-y\right)$ and the distribution of states $G\left(y\right)$ specified in section 2.5, the three-price policy $\mathcal{Q} = \left\{q_1, q_2, q_3\right\}$ is optimal for all $0.86\,\overline{\theta}^p \leq \theta^p \leq 0.75\,\overline{\theta}^p$. For $\theta^p \geq 0.87\,\overline{\theta}^p$, no three-price policy price can be found that satisfies condition (1); for $\theta^p \leq 0.74\,\overline{\theta}^p$, no three-price policy can be found that satisfies condition (2); at $\theta^p = 0.74\,\overline{\theta}^p$, a new mass point has reached critical mass.*

The optimal solution can be traced in this manner for lower and lower values of θ^p.

2.5 A Model of Price Setting

I explore the implications for price adjustment of the information structure developed thus far in a standard model of price-setting under monopolistic competition. 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.[18] Appendix B.3 maps a standard monopolistic competition model with Dixit-Stiglitz preferences into the generic setup introduced in section 2.2.

I show that the model can generate pricing regimes that qualitatively match the features of the Dominick's data documented in Stevens (2011). In particular, depending on parameter values, and consistent with empirical evidence, the model can generate both single-price and multiple-price regimes that are updated relatively infrequently. For the case of multiple-price policies, regimes consist of a small number of distinct prices, but are nevertheless characterized by frequent and large within-regime price changes. Hence, the model endogenously generates transitory volatility to and from discrete price levels.

2.5.1 The Objective Function

The profit of an information-constrained firm is proportional to[19]

$$\pi(q-y) = e^{(1-\varepsilon)(q-y)} - \frac{\varepsilon - 1}{\varepsilon\eta} e^{-\varepsilon\eta(q-y)}, \tag{2.83}$$

where q is the log-normalized price charged by the information-constrained firm, y is the optimal full-information log-normalized price, $\varepsilon > 1$ is the elasticity of substitution,

[18] The treatment of price adjustment in a general equilibrium framework in which all firms are information-constrained requires that each firm track not only an exogenous target price, but also the distribution of prices in the economy. I leave the general equilibrium results for future work.

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

and $\eta \equiv \gamma (1 + \nu)$, where $\gamma \geq 1$ captures decreasing returns to scale in production and $\nu \geq 0$ is the inverse of the Frisch elasticity of labor supply.

Equation (2.83) defines the profit function introduced in section 2.2. Note that this profit function is maximized at $q = y$, hence y 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, subject to the costs of acquiring information about the evolution of this target.

2.5.2 The Shocks

The economy is subject to three kinds of shocks: (1) μ_t, permanent monetary shocks, which are the only source of aggregate disturbances in the economy, are generally small, and are summarized in the exogenous evolution of money supply; (2) $\xi_t (i)$, permanent idiosyncratic quality shocks, which affect both the demand for an individual product and the cost of producing it; and (3) $\zeta_t (i)$, i.i.d. idiosyncratic productivity shocks.

The log of money supply is assumed to follow a random walk process,

$$m_t = m_{t-1} + \mu_t, \tag{2.84}$$

$$\mu_t \overset{i.i.d.}{\sim} h_\mu, \tag{2.85}$$

where μ_t is independent over time and from any other disturbances in the economy.

The permanent quality shock also follows a random walk,

$$a_t (i) = a_{t-1} (i) + \xi_t (i) , \tag{2.86}$$

$$\xi_t (i) \overset{i.i.d.}{\sim} h_\xi, \tag{2.87}$$

while $\zeta_t(i)$ is a purely transitory shock,

$$\zeta_t(i) \overset{i.i.d.}{\sim} h_\zeta. \tag{2.88}$$

The law of motion for the normalized pre-review state, $\widetilde{y}_\tau(i)$, following a review τ periods ago, is

$$\widetilde{y}_\tau(i) = \widetilde{y}_{\tau-1}(i) + \mu_\tau + \xi_\tau(i), \tag{2.89}$$

for $\tau > 0$, with $\widetilde{y}_0(i) = 0$. 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 lemma 6.

The law of motion for the normalized target price that enters the firm's period profit function, $y_\tau(i)$, is

$$y_\tau(i) = \widetilde{y}_\tau(i) + \zeta_\tau(i), \tag{2.90}$$

for $\tau \geq 0$. 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 lemma 8.

2.5.3 Parameter Values

The model is parameterized at the weekly frequency. The parameters that specify the firm's objective, shown in table 19, are set to commonly used values used in the literature. I set the weekly discount factor, β, to 0.9994, which implies an annual discount rate of 3%. The inverse of the Frisch elasticity of labor supply, ν, is set to 0, and the decreasing returns to scale parameter, γ, is set to 1.5. Variations in these two parameters change the steepness 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. The elasticity of substitution,

ε, is set to 6. Variation in ε also changes the asymmetry of the profit function: a higher elasticity implies larger losses from setting a price that is too low relative to the optimal full information price. The profit function is shown in figure 15. Losses from charging a price that is below the current target price are significantly larger than those from charging a price that is above the target price. This asymmetry will affect the resulting shape of the optimal policy, as further discussed below.

Table 20 shows the parametrization of the shocks and table 21 shows the parametrization of the costs of information. The volatility of the permanent monetary shock, μ, is relatively standard. The distribution of μ is triangular with a mean equal to 0.0004 and standard deviation of 0.0015. These values imply an annualized inflation rate of 2.1%, and an annualized standard deviation of 1.1%, which are comparable with the volatility of the U.S. inflation rate over the last thirty years. The calibration of the idiosyncratic shocks is chosen jointly with the information costs to separately match the pricing statistics for products with single-price regimes and multiple-price regimes. The parameters used for each type of pricing policy are discussed in the next two sections.

2.5.4 Single-Price Regimes

I begin by determining the optimal policy when the cost of the price signal, θ^p, is high enough (relative to the volatility of the shocks that affect the firm's profits) such that the firm chooses to acquire no information through its price signal. In this case, the optimal pricing policy always involves a single normalized price, $\overline{q}$, and the model endogenously generates the single-price regimes of Woodford (2009).

In the Dominick's data set, for products with single-price regimes, the frequency of regime changes is 2.2%, and the average size of price changes across regimes is 5.8%. In order to match these statistics, the managerial cost of conducting a policy review, κ, is set to 0.76; the monitoring cost for the review decision, θ^r, is set equal to 1.6; and, to ensure the optimality of the single-price policy, the cost of the price signal is $\theta^p \geq 0.1087$.

The permanent quality shock, ξ, is drawn from a symmetric triangular distribution with mean zero and standard deviation $\sigma_\xi = 0.0104$; and the transitory shock, ς, is set to zero.

At the model-implied weekly frequency of reviews of 2.2%, the managerial cost κ implies that the firm spends 4.2% of its profits on conducting policy reviews. The cost of the review signal implies that 4.2% of the firm's profits are spent on monitoring market conditions in order to make the review decision. The cost of the price signal is high enough, relative to the volatility of the shocks, that the firm acquires no information through the price signal. Hence, the total expenditure on information acquisition is 8.4% of profits. Under this policy, the firm's profits, net of information costs, are 84.7% of the full-information profits that it would achieve if information could be costlessly processed. The first part of table 22 compares the empirical statistics for products characterized by single-price regimes with the moments implied by the model.

Figure 16 plots the hazard function $\Lambda(\widetilde{y})$ implied by a single-price pricing policy, as a function of the normalized pre-review state, $\widetilde{y}$. The asymmetry evident in the figure, and highlighted in Woodford (2009), is the result of the asymmetry in the firm's profit function, such that the firm's review policy is much more likely to trigger a review when the target price at the time of receipt of the review signal, $\widetilde{y}$, is relatively high.

Figure 17 plots the distribution of normalized states, $G(y)$, implied by this review policy. Since the firm reviews its policy with higher probability when the target price is relatively high, the resulting distribution of states that survive the review decision has strong negative skew.

2.5.5 Multiple-Price Regimes

In Dominick's data, products with multiple-price regimes are characterized by a higher frequency of regime changes (3.2% versus 2.2% for products with single-price regimes) and a larger shift in the average price across regimes (7.8% versus 5.8% for products with single-price regimes). Hence, I adjust the calibration, by increasing the standard deviation of the idiosyncratic permanent shock, setting $\sigma_\xi = 0.017$, while also slightly increasing the cost of a policy review, setting $\kappa = 0.81$. This parametrization yields a frequency of reviews of 3.2% and an average shift in the distribution of prices across regimes of 7.7%.

The targets for within-regime price volatility are the median number of distinct prices per regime (equal to four in the data), and the size of price changes within regimes (equal to 10.6% in the data).

As in the case of single-price policies, I begin by determining the optimal review policy when the cost of the price signal, θ^p, is high enough such that the optimal pricing policy always involves a single price, $\overline{q}$. Treating the review policy as fixed, and assuming the same hazard function as in the case of this single-price policy, I then reduce the cost of the price signal, θ^p, such that the optimal policy involves multiple prices. Section 2.4.6 traces out the evolution of the optimal support of the pricing policy.

The columns labeled "MPP" in table 22 compare the empirical statistics for products characterized by regimes with multiple rigid prices with the moments implied by the model calibrated to generate four distinct price per regime. For this calibration, the monitoring cost for the pricing decision, θ^p, is set equal to 0.04, which implies that 3.5% of the firm's profits are spent on acquiring information regarding which price to charge in each week. In total, the firm spends 17.7% of its profits on acquiring information. Net of information costs, the firm's profit is 74.5% of the profit that would be obtained in the full-information, flexible price setting, with no costs of processing information. Within regimes, prices change with a frequency of 39.7%, versus 28.6% in the data. The average size of within-regime price changes is 7.5% versus 10.6% in the data.

In order to generate price changes that are larger within regimes than across regimes, I now introduce transitory shocks, ς. To match the within-regime volatility, these shocks are drawn from a triangular distribution with mean zero and standard deviation $\sigma_\varsigma = 0.0965$. However, the higher within-regime volatility increases both the size and frequency of price changes between policy reviews, as shown in the last column of table 22.

Together with the shape of the firm's profit function, the shape of $G(y)$ determines the shape of the firm's pricing policy. In simulations, the price often remains sticky for considerable periods of periods of time. Figure 18 plots the firm's realized price and the full-information optimal target price. The review signal and the price signal jointly ensure that the firm's price tracks the target price relatively well. Since the firm now has added flexibility to respond to shocks inside the regime, the weighted average price per regime is lower compared with the optimal price that would be implied by the single-price policy. Nonetheless, the firm continues to typically charge a price that is above the target price, reflecting the relatively larger losses from "falling behind." Moreover, since the review signal is noisy, the firm can be seen to occasionally review its policy with a lag.

2.6 Conclusion

This paper presents a theory of price-setting in which firms design simple pricing policies that they update infrequently. The key friction in the model 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 cost per unit of information.

The theory generates pricing patterns consistent with the evidence on discrete multiple-price regimes documented in Stevens (2011). In contrast to existing theories of price setting, it generates pricing regimes that are identified by discrete jumps when the policy is reviewed, and are furthermore characterized by within-regime discreteness, due to the coarseness of the pricing policy implemented between reviews. In this model, neither the frequency of regime changes, nor the frequency of price changes are enough, by themselves, to establish how rapidly prices incorporate changes in market conditions. Nevertheless, price statistics can be used to identify the key model parameters that determine the speed of adjustment. Specifically, price statistics pin down the quantity of information acquired by firms, which in turn determines the degree to which prices are tied to market conditions in each period.

Prices in this model change frequently, yet they are always only partially related to concurrent market conditions. Due to the noisy nature of the firm's information, the speed of adjustment in this model depends far less on the frequency with which prices are changed than on the quantity of information acquired. I leave for future work the computation of a general equilibrium version of the model, and the resulting magnitude of the real effects of monetary policy. Nevertheless, non-neutrality in this model will necessarily be higher than that implied by a model with the same frequency of price adjustment, but in which price changes are based on full information regarding market conditions at the time of adjustment.

Table 18: Evolution of the firm's pricing policy as a function of the unit cost of information.

Stage	Information cost, $\theta^p/\overline{\theta}^p$
SPP is optimal	≥ 1.66
2PP is optimal	$[1.65, 0.87]$
3PP is optimal	$[0.865, 0.75]$
$\geq$ 4PP is optimal	≤ 0.74

Table 19: Parametrization of the firm's objective.

Parameter	β	ν	γ	ε
Value	0.9994	0	1.5	6

Table 20: Parametrization of the shocks.

Parameter	$\overline{\mu}$	σ_μ	$\overline{\xi}$	σ_ξ	$\overline{\varsigma}$	σ_ς
SPP[a]	0.0004	0.0015	0	0.0104	0	0
MPP[b]	0.0004	0.0015	0	0.0170	0	0.0965

[a] SPP: Parametrization for single-price policies.

[b] MPP: Parametrization for multiple-price policies.

Table 21: Parametrization of the costs of information.

Parameter[a]	κ	θ^r	θ^p
Single-price policies	0.76	1.6	$\geq$0.11
Multiple-price policies	0.81	1.6	0.04

[a] All costs of information are expressed as 100 x percent of weekly profits.

Table 22: Price statistics for single-price and multiple-price policies.

Statistic[a]	Data SPP[b]	Model SPP	Data MPP[c]	Model MPP	Model MPP-T[d]
Across regimes					
Freq. of reviews	2.2%	2.2%	3.2%	3.2%	3.2%
Mean $\Delta\bar{p}$	5.8%	5.8%	7.8%	7.7%	7.7%
Within regimes					
Number of prices	1	1	4	4	4
Freq. of Δp	-	-	28.6%	39.7%	70.2%
Mean Δp	-	-	10.6%	7.5%	10.8%
Freq. of modal price	-	-	71.4%	64.5%	43.7%
Information expenditure (% of profits)					
Cost of reviews	-	4.2%	-	6.5%	7.8%
Cost of review signal	-	4.2%	-	7.7%	9.3%
Cost of price signal	-	0	-	3.5%	2.3%
Total expenditure	-	8.4%	-	17.7%	19.3%

[a] Data statistics are medians across products.
[b] SPP: Statistics for products with single-price regimes.
[c] MPP: Statistics for products with multiple-rigid-price regimes.
[d] MPP-T: Statistics for multiple-price policies, with transitory shocks.

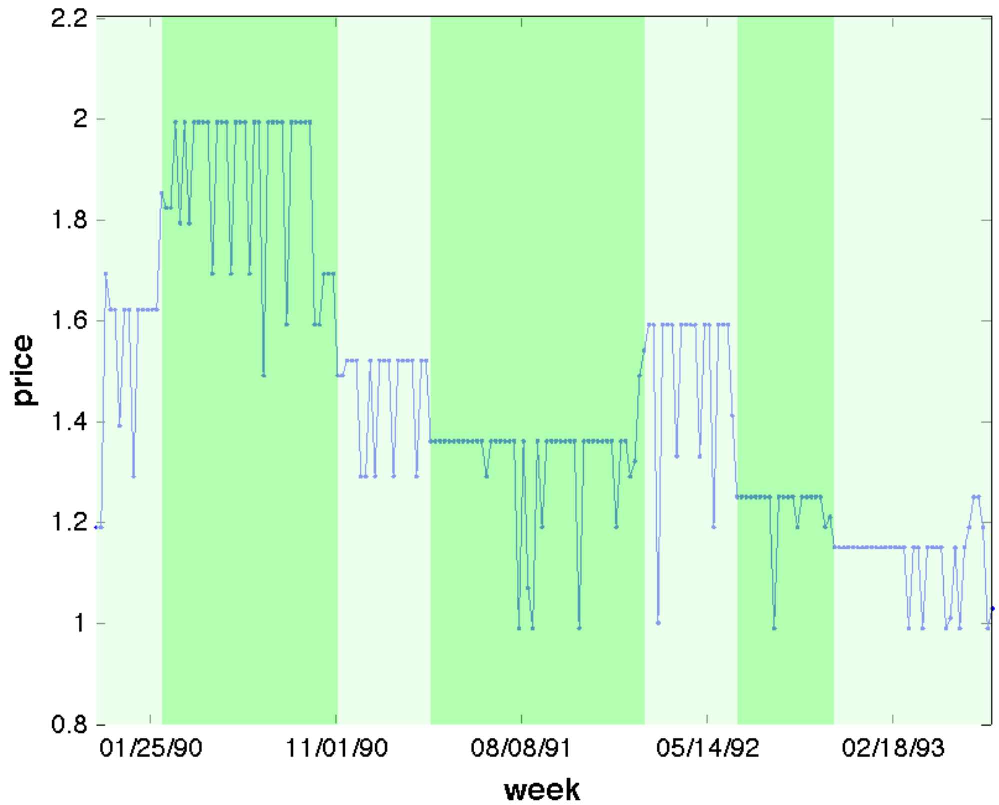


Figure 10: Sample frozen juice price series from Dominick's.

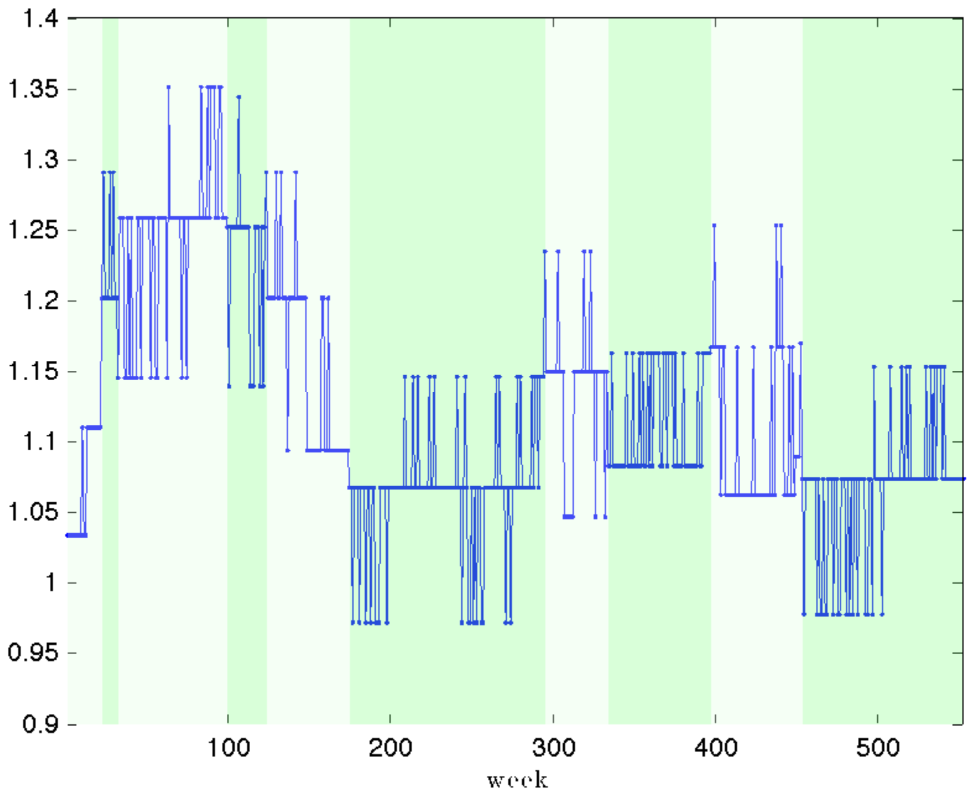


Figure 11: Sample simulated price series.

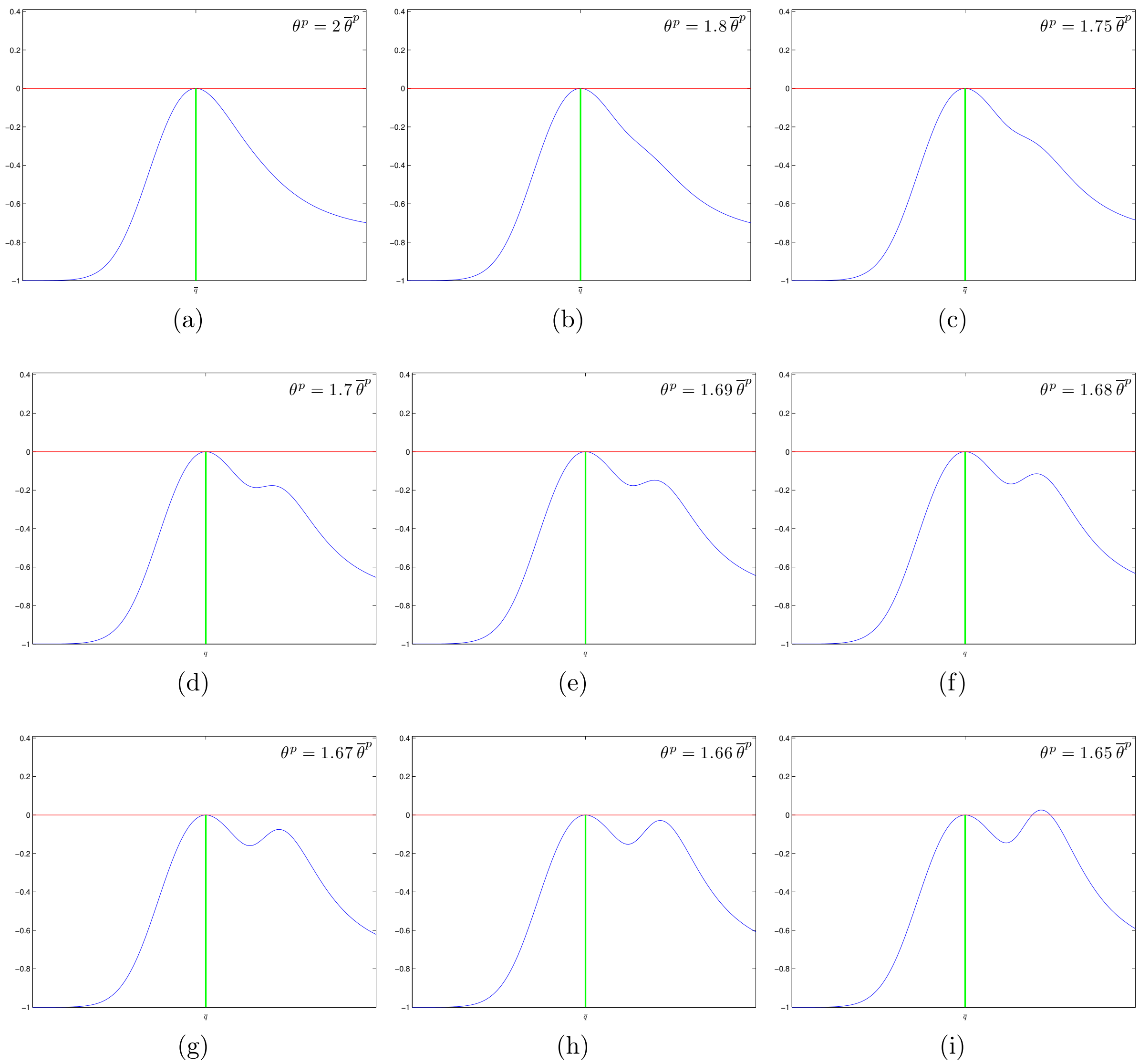


Figure 12: Optimality of the single-price policy and growth of a new mass point as the cost of information is reduced from $\theta^p = 2\,\overline{\theta}^p$ to $\theta^p = 1.65\,\overline{\theta}^p$. At $\theta^p = 1.65\,\overline{\theta}^p$, the single-price policy is no longer optimal. The panels plot the function $Z\left(q;\overline{f}\right) - 1$.

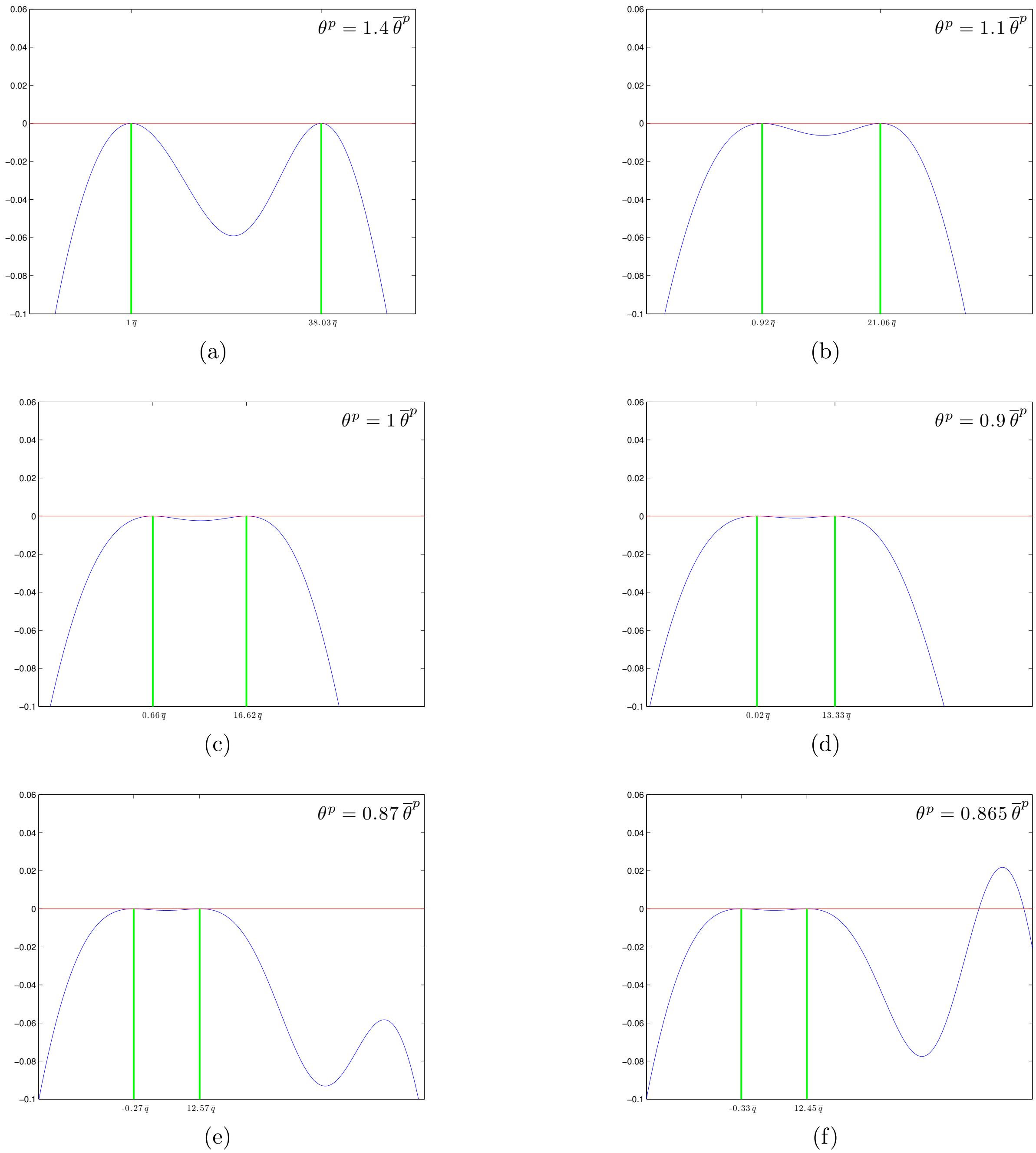
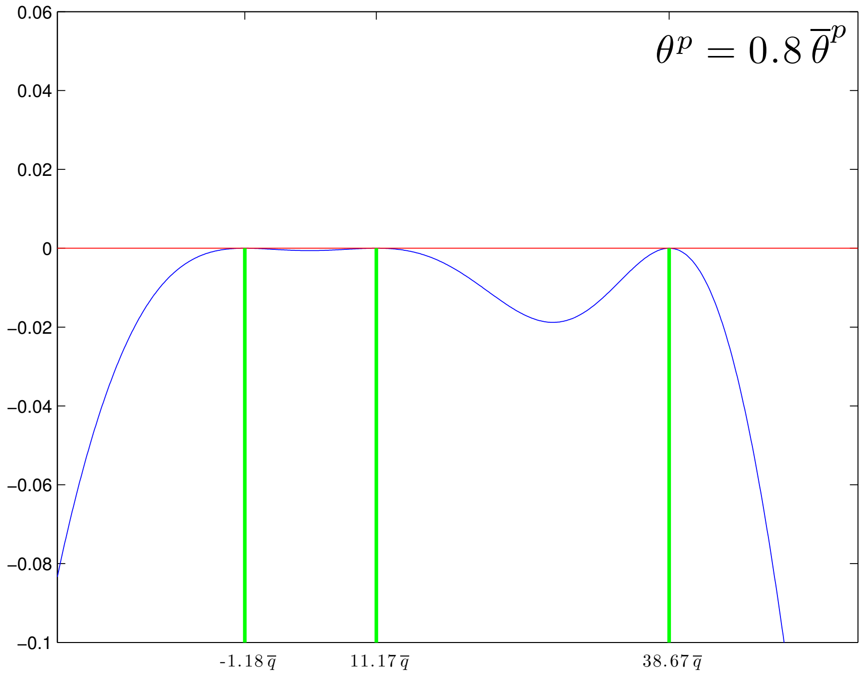
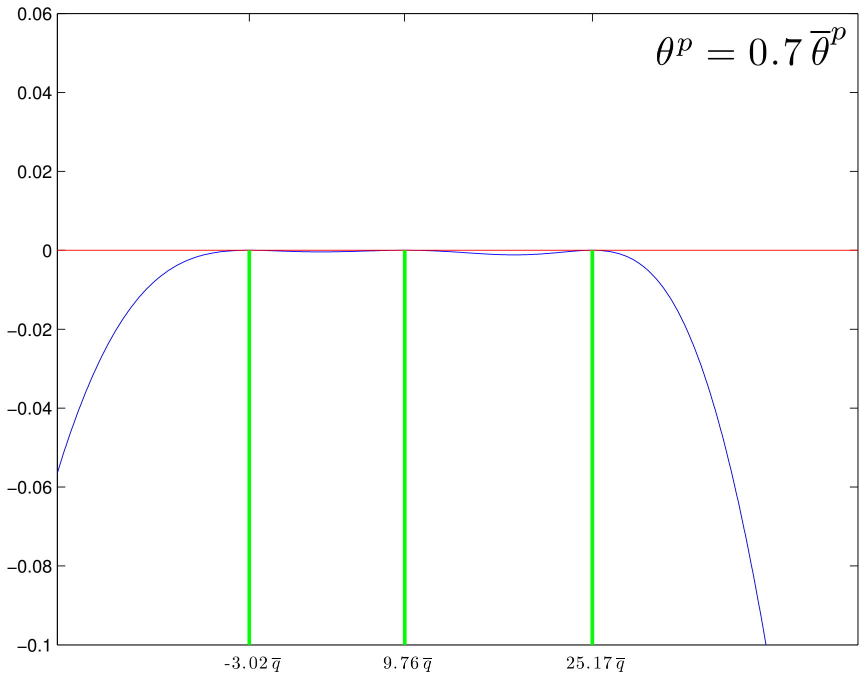


Figure 13: Evolution of two-price policy and growth of a new mass point as the cost of information is reduced from $\theta^p = 1.65\,\overline{\theta}^p$ to $\theta^p = 0.865\,\overline{\theta}^p$. At $\theta^p = 0.865\,\overline{\theta}^p$, no two-price policy is optimal. The panels plot the function $Z\left(q;\overline{f}\right) - 1$.



(a) $\theta^p = 0.8\,\overline{\theta}^p$



(b) $\theta^p = 0.7\,\overline{\theta}^p$

Figure 14: Sample three-price policies.

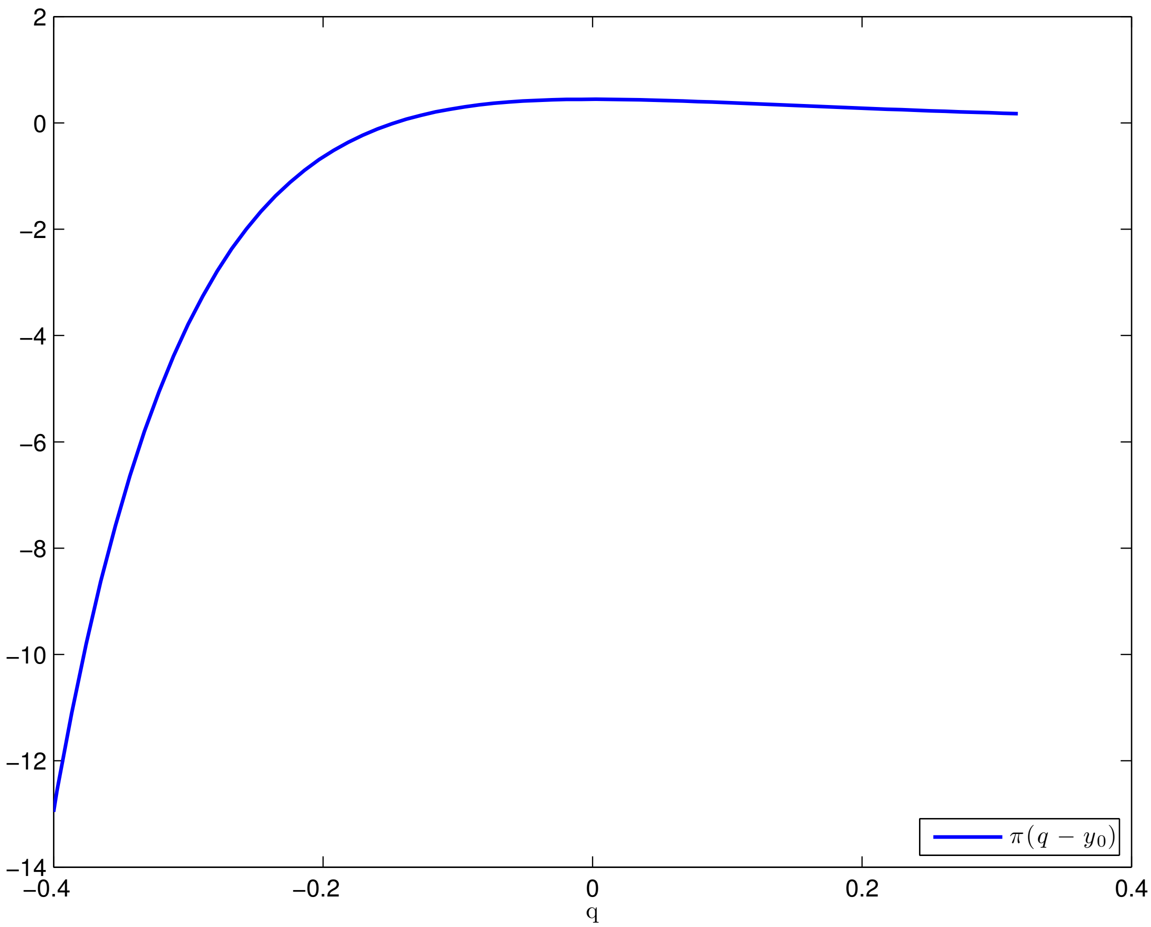


Figure 15: The asymmetric profit function.

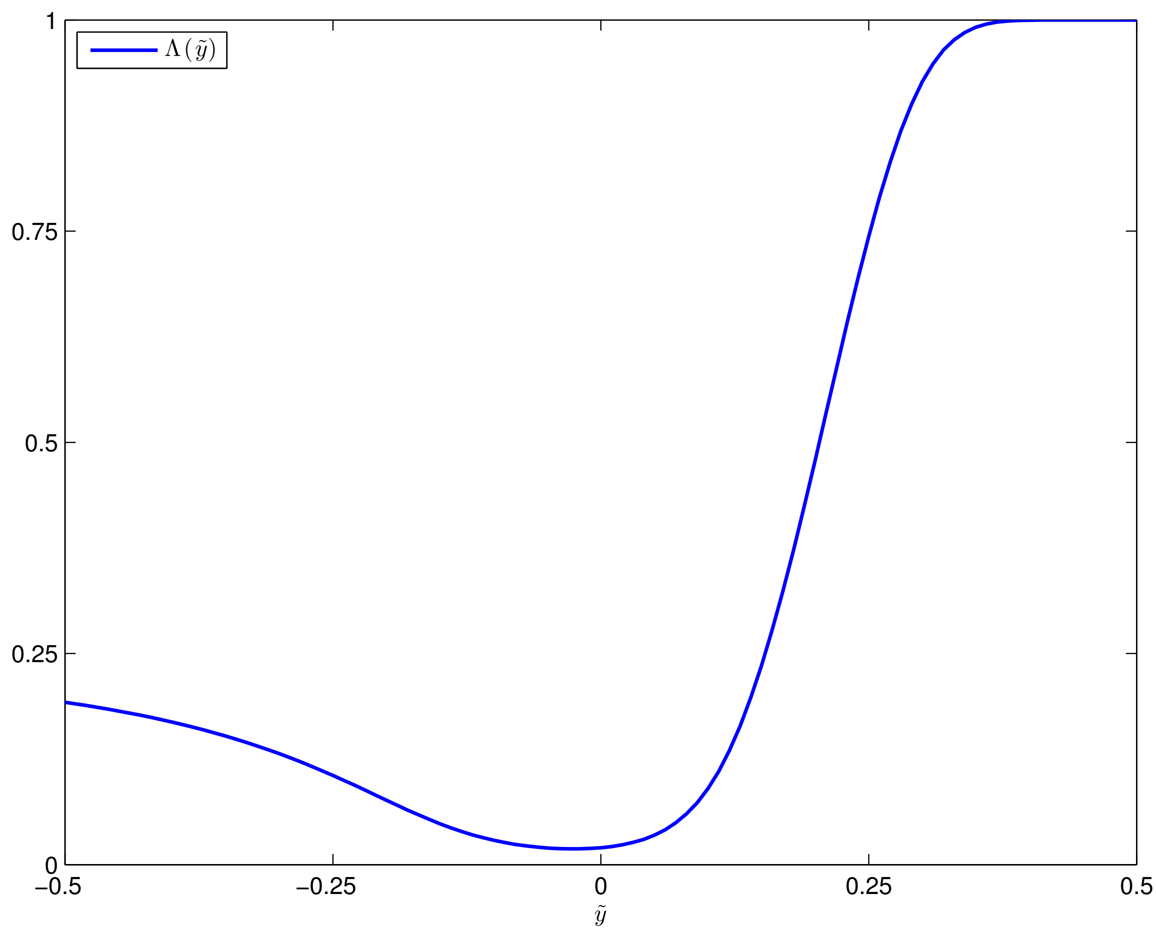


Figure 16: The optimal hazard function for single-price policies.

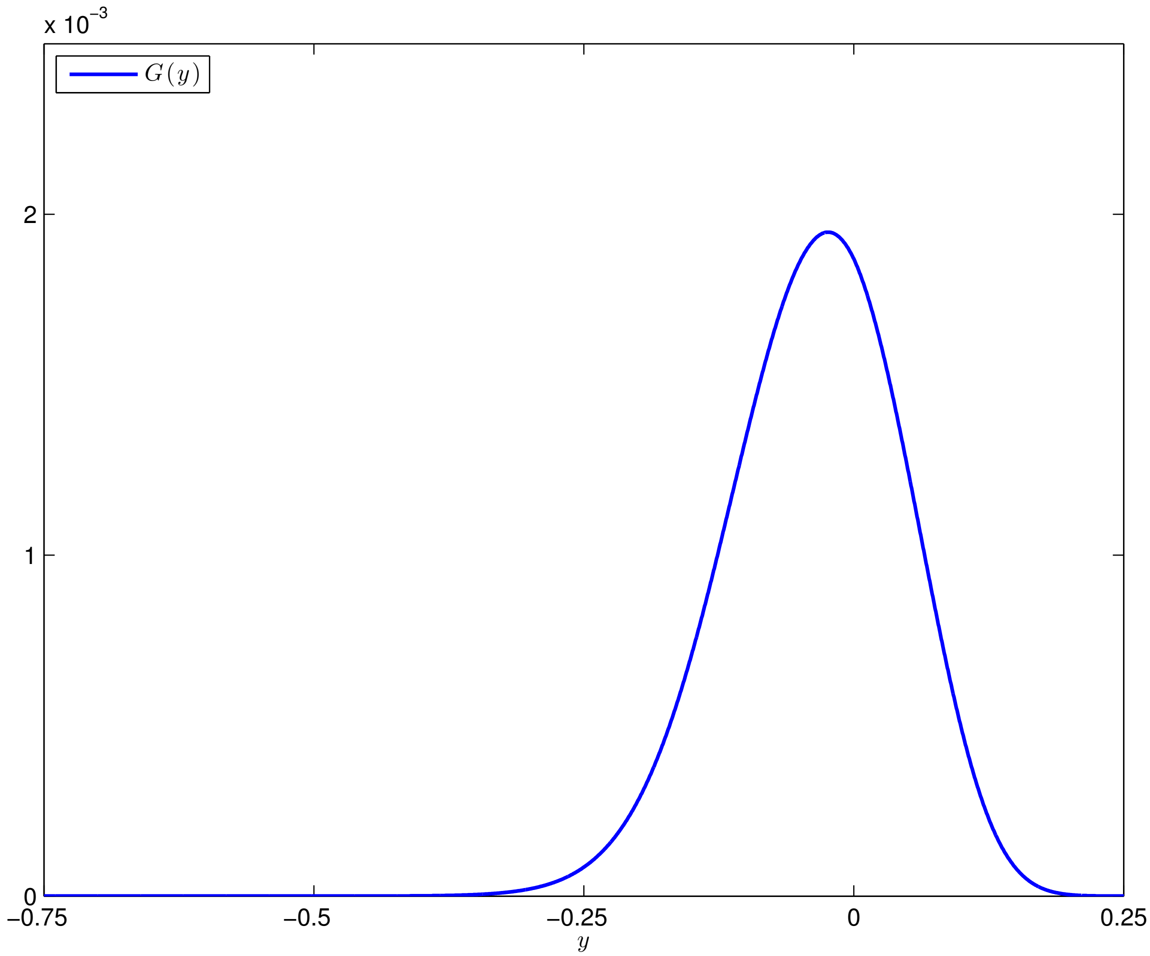


Figure 17: The firm's prior under the hazard function implied by the single-price policy.

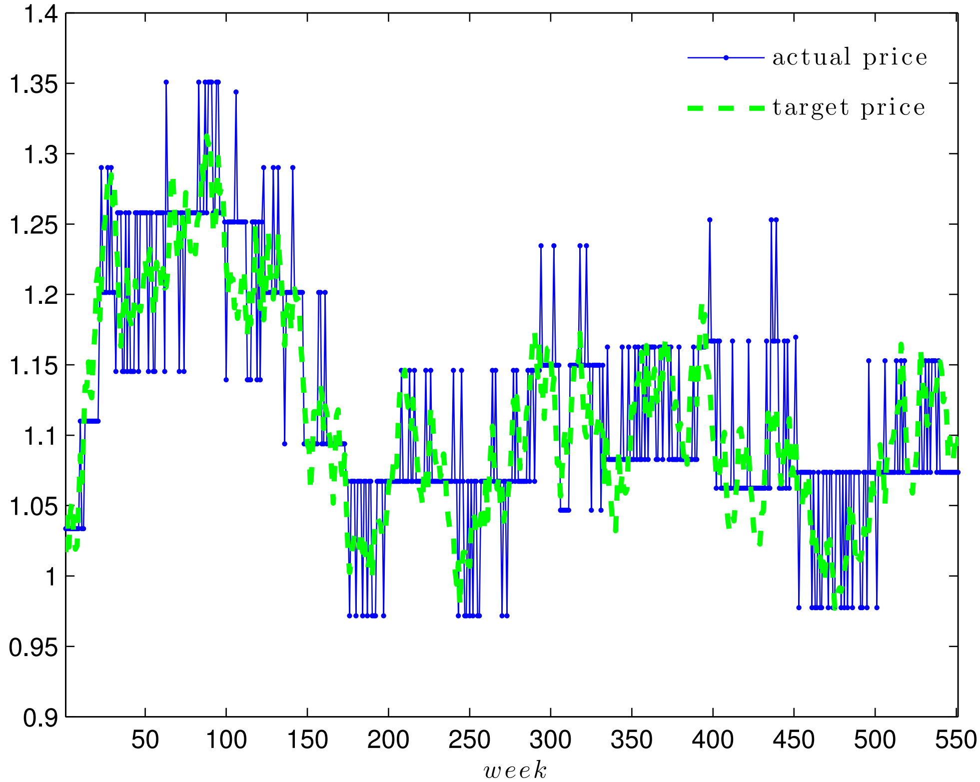


Figure 18: The target price and the actual price.

3 Equilibrium Price Dispersion and the Border Effect

3.1 Introduction

Increasing availability of price data at the product level has led to a resurgence of research into the failure of purchasing power parity and the law of one price across countries. The key feature of the recent literature, compared to the foundational work of Engel and Rogers (1996) and its immediate successors, is that the recent work explicitly considers price dispersion of *individual, identical* products sold in particular cities or stores, rather than the behavior of aggregate price indices. This distinction is important because, as illustrated by Broda and Weinstein (2008), using price indices may overstate relative cross-country dispersion by collapsing large within-country idiosyncratic volatility in relative prices while preserving the variation due to macro-level cross-country differences. Moreover, the use of disaggregated data has enabled researchers to provide a richer set of empirical regularities against which potential theories may be tested. Specifically, recent empirical work[20] has shown that (1) the average (aggregate) real exchange rate closely follows the nominal exchange rate; (2) good-level real exchange rates are widely dispersed within countries, but more so across borders; (3) the volatility of good-level real exchange rates is much greater than the volatility of the nominal exchange rate; and (4) changes in good-level prices are more correlated within than across countries. The new empirical evidence has been accompanied by theoretical models that stress the importance of cross-country segmentation, defined as the complete (Gopinath et al, 2011) or partial (Burstein and Jaimovich, 2009) non-responsiveness of the prices posted by firms in one country to changes in market conditions in another country.

This paper demonstrates that, as hypothesized by Broda and Weinstein (2008), the

[20]We focus on the work of Broda and Weinstein (2008), Gopinath et al (2011), and Burstein and Jaimovich (2009).

existing empirical evidence on price dispersion can be matched by a model in which segmentation across countries is *no stronger* than segmentation across markets within countries. In light of our results, attributing all cross-country dispersion to segmentation at the border may overstate the magnitude of the border effect.

We define segmentation as the degree to which buyers' *access* to one market (namely, their ability to sample prices from firms operating in that market) is lower than their access to another market. In our model, retailers engage in costly sequential search for the best price among producers in the economy. This search friction, combined with a distribution of producer-specific productivity shocks, gives rise to endogenous equilibrium price dispersion, consistent with empirical evidence. Retailers search in a world of two countries, each with two regions. Retailers located in a particular region are more likely to sample prices posted by producers located in their home region than in any of the other three regions. However, conditional on not sampling a price from their own region, retailers are equally likely to sample a price from any of the other three regions. Hence, their access to one of the markets located in the foreign country is no more limited than their access to the "away" market within their own country. In this sense, there is no difference between segmentation across and within countries. Our model also allows for cross-country shocks to relative unit labor costs and aggregate productivity, as well as differences in the distributions of producer-specific productivity.

Our approach is similar to Gopinath et al (2011) and Burstein and Jaimovich (2009) in that we consider a model with a real friction in goods markets, coupled with country heterogeneity in the distribution of firm costs.[21] However, unlike these authors, we employ a model with no additional segmentation across borders.

We show that the basic facts about cross-border price dispersion can be matched using this framework, without any additional impediments to international trade: these

[21] In our model, heterogeneity in costs is supported by nominally fixed wages with monetary shocks. Although we find this an appealing way to link labor costs and exchange rate shocks, the mechanics of the model are not affected by this choice.

facts hold qualitatively whenever there are international differences in the realizations of aggregate shocks or differences in the structural parameters, namely, when the two countries are simply different, and not necessarily isolated from each other. Moreover, we show that price evidence alone is not sufficient to conclusively differentiate between a parameterization with no additional impediments to international trade and one in which we introduce additional segmentation across countries: any additional cross-country segmentation (whereby retailers in one region are more likely to sample from the "away" region in their own country than from one of the two foreign regions) cannot be separately identified from the baseline specification (in which retailers are more likely to sample from their own region than from any of the other three regions).

Although price-data alone cannot identify within-country versus across-country market segmentation, we show that empirical trade quantities can be used to decompose the relative importance of these two types of segmentation. Incorporating a notion of region and country size, we examine the model's implications for the quantity of trade. The calibrated model demonstrates a tension between the US-Canadian price data, which suggest that markets are highly segmented, and the level of US-Canadian trade, which implies a lower degree of border segmentation.

After showing that the model without international barriers can match the basic facts, we demonstrate some of the difficulties in using the standard empirical tools (e.g. Engel and Rogers, 1996, Parsley and Wei, 2001, Gopinath et al, 2010) to measure the degree of market segmentation across borders. We find that typical regressions may indicate that international markets are segmented, even when this is not the case. In this respect, we build on the result of Gorodnichenko and Tesar (2008) that cross-country heterogeneity in price dispersion can lead to positive border coefficients, even without barriers to international trade. We further argue that, according to our model, it is possible to get “false negatives” in these regressions when there are not sufficiently large aggregate shocks to relative productivity; perfectly symmetric economies will in fact generate no evidence of a border effect on prices, regardless of how restricted international trade is. Finally, we show that in the presence of both segmentation and cross-country heterogeneity, standard regressions cannot be used to decompose the measured "border effect" into these two components. Hence in our model, these regressions provide little evidence on the degree of international market segmentation.

The rest of the paper is organized as follows. Section 3.2 reviews the existing empirical evidence. Section 3.3 outlines the search model we employ, extending Reinganum (1979) to a multi-market, two-country setting. Section 3.4 presents our main results. Section 3.5 discusses the identification of regional versus national segmentation, and illustrates the challenges that arise from estimating national segmentation using popular pricing regressions. Section 3.6 concludes.

3.2 Facts on Prices

We evaluate the ability of our theory to match a set of facts that have emerged from the recent empirical literature, concerning both the level and growth rates of relative prices at the good level. For consistency, and reflecting the of availability of data, we focus on papers that study pricing across the US-Canada border.

Let $p_{n,t}(i)$ be the log price, in local currency, of a particular good n, at time t, in a particular location i. Depending on the context, i may index a region, a city, or a specific store. Let $e_t(i,j)$ be the log of the nominal exchange rate between locations i and j. For location pairs within countries, $e_t(i,j)$ is zero. For any two locations, the good-level real exchange rate is defined as

$$d_{n,t}(i,j) \equiv p_{n,t}(i) - p_{n,t}(j) - e_t(i,j). \tag{3.1}$$

Statistics regarding relative prices are most typically pooled across goods, and so from now on we drop the subscript n and assume that all price comparisons are for identical goods.

In its strongest form, the law of one price (LOP) posits that $d_t = 0$ for all t. This is referred to as the *absolute* law of one price and is the focus of Gopinath et al (2011). A weaker hypothesis is the *relative* law of one price (RLOP), which holds that $d_t = a$, where a is a fixed constant for all t. Under the RLOP hypothesis, $\Delta d_t = 0$, and this is the main object of study in Burstein and Jaimovich (2009). Broda and Weinstein (2008) consider violations of both LOP and RLOP.

Fact 1: The aggregate real exchange rate closely follows the nominal exchange rate. This well-known fact has been reconstructed from micro-level data in various forms. Burstein and Jaimovich (2009) show a high correlation between changes in relative unit labor costs and the expenditure weighted average of changes in good-level real exchange rates, Δd_t, across the US-Canada border over the period from 2004 through 2006, when variation in relative labor costs was almost entirely driven by the nominal exchange rate. Using the same data source, Gopinath et al (2011) also show this relationship in levels: the time series constructed by taking the median value of d_t across products for each period t follows the nominal exchange rate almost perfectly from 2004 through the middle of 2007. Similarly, Broda and Weinstein (2008) provide evidence that, controlling for the distance between markets, US-Canada price differences follow the nominal exchange rate over the period 2001 through 2003.

Fact 2: Good-level real exchange rates are more volatile across countries than within. Burstein and Jaimovich (2009) consider the quarterly growth, Δd_t, in *wholesale costs* of a single retailer with stores in multiple locations in Canada and in the US. They find a standard deviation of 6% in the US, 5% in Canada, and 13% across countries from 2004 through 2006. Using weekly *retail price* data from the same retailer, Gopinath et al (2011) find that over the period from 2004 through the middle of 2007, the median standard deviation of d_t for matched goods, measured at the weekly frequency, is approximately 15% within the US, 6% within Canada, but 25% between the two countries.

In contrast, using buyer scanner data aggregated to the city level for the fourth quarter of 2003[22], Broda and Weinstein (2008) find standard deviations of 22% within the US, 19% within Canada, and 27% across the border. We view the two data sets as complementary; the former characterizes the distribution of wholesale prices sampled by a particular retailer across different markets as well as the distribution of prices posted by this retailer; the latter characterizes the distribution of prices sampled by consumers across different markets from all retailers.

Fact 3: Changes in cross-border good-level real exchange rates are significantly more volatile than changes in the aggregate real exchange rate, and hence the nominal exchange rate. In Burstein and Jaimovich (2009), this corresponds to a fact about relative unit labor costs: they find that the standard deviation of Δd_t across borders is at least three times the standard deviation of changes in relative unit labor costs, which is virtually identical to the standard deviation of changes in the exchange rate over that period.

Fact 4: Changes in good level prices (in a common currency) are more correlated within countries than across. Burstein and Jaimovich (2009) find that, within countries, US dollar denominated price changes have a correlation of 75% for the US and 84% for Canada, while the correlation across countries is approximately 7%.

These facts have been taken as evidence that markets are segmented internationally, either completely or partially. In the next section, we lay out a theory which matches these facts without requiring any more segmentation across countries than within countries.

[22]Broda and Weinstein (2008) have scanner data on prices paid by households in different cities in the US and Canada. Their results refer to the price at which a particular good was purchased by a representative household in a given city, rather than the price at which that good was available in a particular store in that city.

3.3 Model

We develop a static model of internal and international price dispersion via search. We follow the search model of Reinganum (1979), in which heterogeneous producer costs and imperfectly elastic customer demand generate a non-degenerate distribution of prices. Our model introduces retailers who search on behalf of consumers. We first present the basic setup, in which producers, retailers, and consumers operate in a single market, or region, and in which the only sources of uncertainty are aggregate and idiosyncratic productivity shocks at the producer level. We then extend the model to a two-region economy and then to a two-country world with two symmetric regions in each country.

3.3.1 The Single-Region Economy

A single consumption good is produced by a continuum of firms with heterogeneous marginal costs. Each producer observes his cost before posting his price. A continuum of retailers purchase the good from the producers subject to a search friction. Retailers know the distribution of prices posted by producers, but they do not know which producer sells at what price. Instead, they pay a fixed cost each time they wish to draw a random price from the distribution of producer prices. The mechanics of retailer search are as follows: (1) producers observe their marginal costs and post their prices, which form a distribution with density $f(\widehat{p})$; (2) retailers pay a fixed cost k to randomly sample a producer price from the population of producers; (3) each retailer chooses either to purchase all of her demand at the sampled price, or to continue searching by paying the search cost and drawing a new price from $f(\widehat{p})$. Search continues until each retailer has settled on a single supplier for their demand. Retailers then costlessly differentiate the good before selling it in a monopolistically competitive consumer market. Consumers buy an index of differentiated retailer goods, and supply labor to producers inelastically.

Consumers maximize

$$\max_{C} u\left(C\right) \quad s.t. \quad PC \leq PY, \tag{3.2}$$

where the utility function is $u\left(C\right) = \log\left(C\right)$, $C\left(\cdot\right)$ is the CES consumption aggregator over retailer goods, P is the retailer price index, and Y denotes real income (the sum of exogenous labor income and profits from firms). For simplicity, consumers cannot borrow or save.

Consumer demand for a particular retail good with price p is given by

$$y(p) = p^{-\eta} P^{\eta} Y, \tag{3.3}$$

where $\eta > 1$ is the constant elasticity of substitution among retailer varieties, P is the retailer price index implied by the CES demand function, and Y is aggregate demand.

Each retailer seeks to maximize profits net of total search costs, which are defined as

$$\pi_j^R = (p - \widehat{p}) y(p) - kn, \tag{3.4}$$

where $\widehat{p}$ is the producer price upon which the retailer settles after completing search for the period, k is the cost of search per producer searched, and n is the number of producers that the retailer samples in the period.

The sequential nature of search implies that a retailer's choice to continue looking for a better price is independent of the number of producers already visited in the period. The value to a retailer of halting the search and purchasing the good from a producer with price $\widehat{p}$ is given by

$$V^{ns}(\widehat{p}; f) = \max_{p} \pi^R(p; \widehat{p}; f). \tag{3.5}$$

Since differentiation of the good at the retail level is costless, $\widehat{p}$ is the retailer's marginal cost, and the retailer maximizes expression (3.5) by charging the monopoly price $p = \mu\widehat{p}$, where $\mu \equiv \eta / (\eta - 1)$.

Conversely, the value to the retailer of continuing to search upon observing $\widehat{p}$ is given by

$$V^s(\widehat{p}; f) = E[V(\widehat{p}; f)] - k, \tag{3.6}$$

where expectations are taken with respect to the producer price density, $f(\widehat{p})$. The overall value function for the retailer is given by

$$V(\widehat{p}; f) = \max\{V^s(\widehat{p}; f), V^{ns}(\widehat{p}; f)\}. \tag{3.7}$$

The retailer will continue to search so long as she benefits in expectation from continued search,

$$V^s(\widehat{p}; f) \geq V^{ns}(\widehat{p}; f). \tag{3.8}$$

As in Reinganum (1979), the optimal search strategy is a stopping rule described by a unique reservation price $\widehat{p}_r$ that sets expression (3.8) to equality: all retailers sampling a price less than or equal to $\widehat{p}_r$ stop search and purchase all their demand at the sampled price, while all retailers sampling a price above $\widehat{p}_r$ continue to search for a better offer. Since retailers pass on the demand from consumers to producers, the demand of a retailer settling on a producer posting price $\widehat{p}$ is

$$D(\widehat{p}) = \mu^{-\eta} \widehat{p}^{-\eta} P^{\eta} Y. \tag{3.9}$$

Hence, the demand faced by a producer with price $\widehat{p}$ is given by

$$x(\widehat{p}) = \begin{cases} D(\widehat{p}) & \texttt{if} \quad \widehat{p} \leq \widehat{p}_r \\ 0 & \texttt{if} \quad \widehat{p} > \widehat{p}_r. \end{cases} \tag{3.10}$$

Finally, each producer i faces a production function that is linear in labor,

$$x_i = A_i h_i, \tag{3.11}$$

where h_i is hours and A_i is productivity, which is the product of aggregate productivity, ϵ, and an independent idiosyncratic productivity, ζ_i, distributed independently across all producers:

$$\log(A_i) = \epsilon + \zeta_i. \tag{3.12}$$

Unit labor costs, w, are given exogenously, and consumers always satisfy labor demand at the going wage. Thus, the marginal cost of producer i is $mc_i = w/A_i$. The implied distribution of marginal costs is denoted by $g(mc)$.

An equilibrium in the producer-retailer market is a retailer reservation price $\widehat{p}_r$ and a distribution of producer prices $f(\widehat{p})$ such that (1) given $f(\widehat{p})$, retailers choose the optimal stopping rule governed by $\widehat{p}_r$ and (2) given $\widehat{p}_r$, producers maximizing profits generate $f(\widehat{p})$. The optimal price set by producers and the resulting cumulative distribution of producer prices are given by

$$\widehat{p} = \min\{\mu mc, \widehat{p}_r\} \tag{3.13}$$

and

$$F(\widehat{p}) = \begin{cases} G(\frac{\widehat{p}}{\mu}) & \texttt{if} \quad \widehat{p} \leq \widehat{p}_r \\ 1 & \texttt{if} \quad \widehat{p} > \widehat{p}_r, \end{cases} \tag{3.14}$$

respectively.

The retailers' stopping rule implies that there will be no search in equilibrium since all producers will post prices that are weakly below the retailers' reservation price. However, because demand is elastic, producers with heterogeneous costs do not find it optimal to generate a single price equilibrium at $\widehat{p}_r$. Hence, consistent with empirical evidence, the model generates equilibrium price dispersion within a single market. The degree of dispersion in prices is determined by the cross-sectional dispersion of producer costs. Markups are constant for all producers with marginal costs less than $\widehat{p}_r/\mu$, and are smaller for all those producers with costs larger than this threshold.

3.3.2 The Two-Region Economy

We now divide the economy into two equal-sized regions, a and b, separated by a regional border. The wage rate, aggregate productivity, and the distribution of idiosyncratic shocks are the same in both regions.[23] The only difference across the two regions is that retailers from a particular region may be more likely to sample prices from producers located within their own region than from those located in the other region. Specifically, let $f_a(\widehat{p})$ and $f_b(\widehat{p})$ denote the distributions of producer prices in each region. During her search, a retailer in region a has probability α of drawing a price from the distribution of producer prices posted in her own region, $f_a(\widehat{p})$, and a probability $1-\alpha$ of drawing a price from the distribution of producer prices posted in the neighboring region, $f_b(\widehat{p})$. The degree of regional sampling bias for retailers from region a implies a regional segmentation parameter

$$\lambda \equiv \frac{\alpha}{1-\alpha}. \tag{3.15}$$

For simplicity, we take the segmentation parameter λ as exogenous. It captures all the frictions and barriers to trade, either bilateral or unilateral, that may make transacting across regions less likely. It may be motivated by informational advantages that ease access to the chain of production in one's own market, or by external barriers that make transacting with firms located outside one's own network more difficult. The exogeneity assumption can be relaxed, as long as λ remains independent of relative prices in the

[23] The assumption that markets within the same country are symmetric is supported by the evidence in Gopinath et al (2011) that, within countries, price differentials are centered around zero, and in Burstein and Jaimovich (2009) that average changes in relative prices within countries are zero.

two regions. In this environment, the price distribution faced by retailers in regions a is given by

$$f_a^{ret}(\widehat{p}) \equiv \alpha f_a(\widehat{p}) + (1-\alpha) f_b(\widehat{p}). \tag{3.16}$$

Otherwise, each retailer's problem is the same as in the single-region setting.

Retailers in market b face a similar tendency to over-sample prices from their own region, although their search bias may be different from that of region a retailers. Hence, it is important to note that there is no notion of segmentation at the border, but only of segmentation of one retail market with respect to producers in another market. For instance, it may be the case that producers in one region export easily to the other region, while at the same time, retailers in this region encounter higher frictions in importing from the other region.

The producers' problem remains unchanged. Since we do not incorporate any region-specific shocks and since producers across both regions draw their marginal costs independently from the same distribution, $g(mc)$, their desired prices are given in expression (3.13) and retailers have the same reservation price across the two regions, $\widehat{p}_{r,a} = \widehat{p}_{r,b}$. Hence, producers in both regions post prices drawn from the same distribution, $f_a(\widehat{p}) = f_b(\widehat{p})$.

Since there are no differences in the distributions of prices posted by producers, retailers in the two regions also sample from identical distributions, $f_a^{ret}(\widehat{p}) = f_b^{ret}(\widehat{p})$, *regardless of the value of* α. As before, retailers add a constant markup to the sampled prices, hence the distributions of prices paid by the consumers of the two regions are also identical: $f_a^{cons}(p) = f_b^{cons}(p)$. As a consequence, the distributions of prices at the producer, retailer, and consumer level are identical across the two regions, and cannot be used to infer the degree of regional segmentation. This result foreshadows the challenge of separately identifying regional versus national segmentation in the two-country model. As shown below, pricing statistics generate an estimate of the overall segmentation between countries. However, since in the absence of regional shocks, regional segmentation cannot be identified using price data alone, any estimate of overall segmentation based only on price data will confound regional and national barriers.

3.3.3 The Two-Country Model

We now extend the model to a two-country setup, in which there are two regions in each country: a and b in the *Home* country, and c and d in the *Foreign* country. Each region is of unit mass. The realizations of aggregate shocks differ across the two countries and the distributions of idiosyncratic shocks are also country-specific.

Similar to the two-region case, a retailer in region a of the *Home* country samples prices from the following distribution

$$f_a^{ret}(\widehat{p}) \equiv \alpha_1 f_a(\widehat{p}) + \alpha_2 f_b(\widehat{p}) + \alpha_3 f_c(\widehat{p}) + \alpha_4 f_d(\widehat{p}), \tag{3.17}$$

where $\sum_{i=1}^{4} \alpha_i = 1$, and where $f_r(\widehat{p})$ is the density of prices posted by the producers of region r, $r \in \{a, b, c, d\}$. We assume that retailers in region a have a regional search bias, in that they are more likely to sample from their own region: $\alpha_1 > \alpha_i$, $i \in \{2, 3, 4\}$.

By assumption, the *Home* regions, a and b, are structurally identical, as are the *Foreign* regions, c and d. Thus, the prices posted by producers in each country are identically distributed,

$$f_a(\widehat{p}) = f_b(\widehat{p})$$
$$f_c(\widehat{p}) = f_d(\widehat{p}).$$

Therefore, equation (3.17) becomes

$$f_a^{ret}(\widehat{p}) = \alpha f_a(\widehat{p}) + (1 - \alpha) f_c(\widehat{p}), \tag{3.18}$$

where we now redefine α, setting it equal to $\alpha_1 + \alpha_2$.

Retailers in region a have an *apparent* national sampling bias if they are more likely to sample from their own country, namely, if $\alpha > 1/2$. If, in addition to having a regional sampling bias, retailers in region a are equally likely to sample prices from any of the other three regions, as in Figure 19, specifically, if $\alpha_2 = \alpha_3 = \alpha_4 = \frac{1-\alpha_1}{3}$, then the regional sampling bias embedded in α_1 implies a national bias in α without any additional segmentation of markets at the national border:

$$\alpha = \frac{2\alpha_1 + 1}{3} > \frac{1}{2}. \tag{3.19}$$

This case is illustrated in figure 19.

Retailers in region b of the *Home* country face a similar tendency to over-sample prices from their own region. Although the regional and national sampling biases could, in principle, be different for the two *Home* regions, either because of relative size differences or because of differences in segmentation, we assume, for now, that the two regions are of equal size and symmetrically segmented. Hence, retailers in region b also sample prices from the distribution given by equation (3.18).

Retailers in region c of the *Foreign* country samples prices from the distribution specified by

$$f_c^{ret}(\widehat{p}) \equiv \gamma_1 f_a(\widehat{p}) + \gamma_2 f_b(\widehat{p}) + \gamma_3 f_c(\widehat{p}) + \gamma_4 f_d(\widehat{p}), \tag{3.20}$$

where $\sum_{i=1}^{4} \gamma_i = 1$, and where the regional sampling bias is captured by $\gamma_3 > \gamma_i$, $i \in \{1, 2, 4\}$. The equivalent of the sampling distribution (3.18) for retailers in region c is given by

$$f_c^{ret}(\widehat{p}) \equiv (1 - \gamma)\, f_a(\widehat{p}) + \gamma f_c(\widehat{p}), \tag{3.21}$$

where the *Foreign* national bias, γ, is a function of *Foreign* regional bias, γ_3, and may be different from the *Home* national bias, α:

$$\gamma = \frac{2\gamma_3 + 1}{3} > \frac{1}{2}. \tag{3.22}$$

By symmetry, retailers from region d also sample prices from (3.21). The full derivations for retailers in regions b, c and d are shown in the appendix.

Due to cross-country heterogeneity, producers in each country generate different producer price distributions, $f_a(\widehat{p}) \neq f_c(\widehat{p})$, and retailers in each country may have different reservation prices, $\widehat{p}_{r,a} \neq \widehat{p}_{r,c}$. Without loss of generality, let the *Home* country be relatively less expensive, with $\widehat{p}_{r,a} < \widehat{p}_{r,c}$. Differences in reservation prices arise across countries due to (1) aggregate productivity differences, (2) nominal wage differences, or (3) differences in the distribution of idiosyncratic productivity shocks. In equilibrium, all producers in both countries post prices that are weakly lower than the high reservation price, $\widehat{p}_{r,c}$. High-cost producers in either country who post prices between the two reservation prices only sell to retailers in the *Foreign* regions c and d. Conversely, producers charging prices weakly below $\widehat{p}_{r,a}$ sell to retailers in both countries at a single monopoly price that takes into account the different price levels and relative demand in the two countries.

In order to construct the demand functions faced by producers in the two-country setup, we first consider the demand from retailers located in the *Foreign* regions c and d. These retailers search only once, since they have the highest reservation price. From the unit mass of retailers in region c, a fraction $(1-\gamma)$ purchase from *Home* producers and a fraction γ purchase from *Foreign* producers. The unit mass of retailers in region d behave in the same way, such that the total mass of *Foreign* retailers buying from *Home* producers is $2\,(1-\gamma)$. This demand is evenly split between region-a and region-b producers, since these two regions are symmetric from the perspective of all *Foreign* retailers. The total mass of *Foreign* retailers that are matched with *Foreign* producers is

2γ, and it is also, due to symmetry, evenly split between region-c and region-d producers.

Next, we consider the search process of the retailers located in the *Home* regions a and b. These retailers have the low reservation price. Unlike in the two-region economy in which there was no equilibrium search, they may search repeatedly, until they come across a price that is below their reservation price. The measure of region-a retailers who initially search in the *Home* country and settle on a *Home* producer is $\alpha F_a(\widehat{p}_{r,a})$, and the measure of region-a retailers who initially search in the *Foreign* country and settle is $(1-\alpha)\,F_c(\widehat{p}_{r,a})$. Hence, the total mass of region-a retailers left to search again is

$$L_a \equiv 1 - \alpha F_a(\widehat{p}_{r,a}) - (1-\alpha)\,F_c(\widehat{p}_{r,a}). \tag{3.23}$$

Of these, a fraction L_a will again be left to search after the second round of search, and $(1 - L_a)$ will find an acceptable price, either at home or abroad. Hence, after each round n of search, the remaining mass of region-a retailers who continue to search is L_a^n. The mass of region-a retailers matched to *Home* producers after round n of search is

$$M_{a,n}^H = \alpha F_a(\widehat{p}_{r,a}) L_a^{n-1} \tag{3.24}$$

The cumulative mass of region-a retailers matched to *Home* producers after round n, $n \to \infty$, which is given by $\mu_a^H \equiv \lim_{n\to\infty} \sum_{i=1}^{n} M_{a,i}^H$, becomes

$$\mu_a^H = \frac{\alpha F_a(\widehat{p}_{r,a})}{\alpha F_a(\widehat{p}_{r,a}) + (1-\alpha)\,F_c(\widehat{p}_{r,a})} > \alpha. \tag{3.25}$$

The remaining $1 - \mu_a^H$ of region-a retailers are matched to *Foreign* producers. Hence, because of repeated search, more region-a retailers are matched with the (relatively cheaper) *Home* producers than would be matched if there were no opportunity to search again after the first round. Conversely, *Foreign* retailers do not benefit as much from the cheaper *Home* prices, since they only search once.

Due to regional symmetry, the cumulative mass of region-b retailers matched to *Home* producers after round n, $n \to \infty$, is also equal to μ_a^H. Therefore, $2\mu_a^H$ retailers from the *Home* country are eventually matched to *Home* producers (and are evenly split between region-a and region-b producers); and $2\left(1-\mu_a^H\right)$ retailers from the *Home* country are eventually matched to *Foreign* producers (and are also evenly split between the producers from regions c and d).

In summary, the demand function faced by a producer in either region a or region b is given by

$$x_a(\widehat{p}) = x_b\left(\widehat{p}\right) = \begin{cases} \left[\mu_a^H + (1-\gamma)\right] D(\widehat{p}) & \texttt{if} \quad \widehat{p} \leq \widehat{p}_{r,a} \\ (1-\gamma)\, D\left(\widehat{p}\right) & \texttt{if} \quad \widehat{p}_{r,a} < \widehat{p} \leq \widehat{p}_{r,c} \\ 0 & \texttt{if} \quad \widehat{p} > \widehat{p}_{r,c}, \end{cases} \tag{3.26}$$

and the demand function faced by a producer in either region c or region d is given by

$$x_c(\widehat{p}) = x_d\left(\widehat{p}\right) = \begin{cases} \left[\left(1-\mu_a^H\right) + \gamma\right] D(\widehat{p}) & \texttt{if} \quad \widehat{p} \leq \widehat{p}_{r,a} \\ \gamma D\left(\widehat{p}\right) & \texttt{if} \quad \widehat{p}_{r,a} < \widehat{p} \leq \widehat{p}_{r,c} \\ 0 & \texttt{if} \quad \widehat{p} > \widehat{p}_{r,c}, \end{cases} \tag{3.27}$$

where the quantity demanded $D(\widehat{p})$ is given by (3.9). In equilibrium, the distribution of prices may contain mass points at one or both reservation prices, $\widehat{p}_{r,a}$ and $\widehat{p}_{r,c}$.

A similar derivation yields demand functions if $\widehat{p}_{r,a} > \widehat{p}_{r,c}$, in which case the *Foreign* country retailers are the ones who may search repeatedly in equilibrium. Conversely, if $\widehat{p}_{r,a} = \widehat{p}_{r,c}$, both the *Home* and the *Foreign* retailers settle on the first producer sampled, and the model collapses to the two-region case discussed in the previous section.

3.3.4 Exchange-Rate Determination

Although the model can be solved using an exogenous process for labor costs, we instead make assumptions that permit a simple model of the link between exchange rates and real labor costs. In particular, we assume that wages are fixed nominally in the local currency and that money demand follows a standard velocity equation, with fixed velocity normalized to one. Under these assumptions, $P_H Y_H = M_H^s$, $P_F Y_F = eM_F^s$, and

$$e = \frac{M_H^s}{M_F^s}\frac{P_F Y_F}{P_H Y_H},$$

where $P_i Y_i$ gives the common-currency value of total output in each country. Shocks to relative money supply, which follows a persistent AR(1) process, generate differences in real unit labor costs between countries, and therefore additional cross-country differences in average costs of production beyond those generated by productivity shocks alone.

3.3.5 Model Intuition

We illustrate the properties of the model in three different settings in which the distribution of prices posted by producers differs across countries. Despite this difference, the distribution of prices sampled by retailers, and in turn the distribution of consumer prices, differs only in the last case.

First, we consider the case in which the two countries differ only in the distributions of producer-specific productivity shocks. The countries are symmetric and there is no bias in search, but $\sigma_{\zeta,H} > \sigma_{\zeta,F}$. Figure 20 shows the pricing function, $\widehat{p}(mc_i)$, and the cumulative distribution of producer prices in both the *Home* and *Foreign* countries. Both countries have the same pricing function, $\widehat{p}(mc_i)$, and the same reservation price. However, since the wider distribution of prices in the *Home* country puts more mass on cost-price combinations below the reservation level, the average price posted by producers in the *Home* country is lower than the average price of the *Foreign* producers. Nevertheless, in the absence of any search bias, prices sampled by retailers in each country are drawn from the same distribution.

Next, figure 21 illustrates the case in which the *Home* country experiences low relative unit labor costs ($w_H < w_F$). Once again, prices posted by producers differ across the two countries. The *Home* and *Foreign* reservation prices are now slightly different because the fixed search cost, k, represents a larger proportion of the *Home* retailers' profits than it does for the *Foreign* retailers, so that *Foreign* retailers are more willing to search for a better price. The difference in labor costs across countries means that prices posted in the *Home* country are lower on average. Nevertheless, since there is no bias in search, the price distributions sampled by retailers in each country are identical.

Finally, figure 22 shows the case where the two countries have the same structural parameters, but experience different unit labor costs ($w_H < w_F$) *and* face regional bias in search. In this case, the reservation price in the *Foreign* country is substantially higher than in the *Home* country because producers with high costs attempt to capitalize on "trapped" *Foreign* retailers, rather than set a lower price that appeals to retailers in both regions. Under some parametrizations, this may even be true in the low cost *Home* country to the extent that some retailers from the *Foreign* country sample first in the *Home* country, and find it worthwhile to pay the higher reservation price rather than search again. In this case, the prices sampled by retailers - and by extension, by consumers - differ across the two countries.

Figures 20 through 22 demonstrate an important point: in order to observe pricing to market (namely, firms with identical marginal costs charging different prices), markets must be at least partially segmented, and also experience some asymmetry, in terms of either their average productivities or the distributions of idiosyncratic productivities. Without both segmentation and differences between markets, firms with the same cost will charge same price regardless of their location. This basic intuition underlies our later results regarding the identification of market segmentation using reduced form regressions.

Figure 23 breaks down the profit-maximizing policy for the *Foreign* producers in the case of partial segmentation, showing profits as function of marginal costs, given different pricing policies. As shown in panel (a), when marginal cost is less than $\widehat{p}_{r,H}/\mu$ - the threshold set by the low reservation price - producers charge the monopoly price (indicated by the blue line) to retailers in *both* countries. At higher marginal costs, once the desired monopoly price exceeds the *Home* reservation price, producers charge the *Home* reservation price (green line, panel (b)), thereby maintaining market share in the *Home* country. Under this policy, the producer more than makes up in volume from *Home* retailers what he loses in pricing from *Foreign* retailers. Marginal costs eventually reach a high enough critical point, c^*, where the producer no longer finds it worthwhile to keep selling to *Home* retailers and forgo the profits of charging *Foreign* retailers a higher price; instead, he starts charging the *Foreign* monopoly price (purple line, in panel (c)). Finally, with high enough marginal costs, the producer simply charges the *Foreign* reservation price (red line, panel (d)), or drops out of the market. In this setting, retailer markups remain constant, while producer markups are heterogeneous, with a mass point at the maximum producer markup, μ.

3.4 Results

In this section, we discuss the baseline calibration of the model as well as the implied importance of pricing-to-market in generating the observed cross-border price dispersion.

3.4.1 Parameter Values

Table 23 shows the calibrated parameters values required to match the four pricing facts presented above. On the retailer side of the model, we fix the elasticity of substitution between retail goods at $\eta = 5$, a value standard in the literature, which yields retail markups of 20%. We set the retailers' search cost parameter at $k = 0.006$, which, given η, implies an unconditional producer markup of 15%.

We calibrate an AR(1) process for the relative money supply, $\frac{M_H^s}{M_F^s}$, in order to match the variance and high persistence of the US-Canada nominal exchange rate. We set the persistence, $\rho_M = 0.95$ and the volatility of innovations, $\sigma_{\varepsilon_m} = 0.029$. These values yields an exchange rate with the same persistence, $\rho_e = 0.95$, and a variance (in growth rates), $\sigma_{\Delta e} = 0.03$.

On the producers' side, we assume that average relative productivity is distributed according to $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and that producer-specific productivity shocks are distributed within each period according to $\zeta_i \sim \mathcal{N}(0, \sigma_{\zeta,c}^2)$, with $c \in \{\text{us, can}\}$. We calibrate the cost shock parameters to match the main moments of the cost data for the US and Canada considered by Burstein and Jaimovich (2009). These data correspond most closely our interpretation of the model as an interaction between retailers and wholesalers. We choose the shock variances $(\sigma_{\zeta,us}^2, \sigma_{\zeta,can}^2, \sigma_\epsilon^2)$ to match the variability of good-level real exchange rates: $\sigma_{\Delta d_{t,us}} = 0.06$, $\sigma_{\Delta d_{t,can}} = 0.05$, and $\sigma_{\Delta d_{t,bord}} = 0.13$. Generating higher variability of good-level real exchange rates across the border requires a fairly large variance of the shock to relative productivities, hence we set $\sigma_\epsilon = 0.059$. We generate higher dispersion within the US versus Canada by assuming a higher variance of US idiosyncratic shocks, setting $\sigma_{\zeta,us} = 0.093$ and $\sigma_{\zeta,can} = 0.052$.

The regional bias parameters are selected in order to match the correlation of aggregate real exchange rates with the nominal exchange rate (fact 1). For the baseline calibration, we assume symmetry, so that $\alpha_1 = \gamma_3$. Since Burstein and Jaimovich (2009) do not provide an numerical correlation, we target a correlation coefficient of 0.70. This requires a very high regional bias ($\alpha_1 = \gamma_3 = 0.998$), so that retailers in all regions are extremely unlikely to search outside their own region.

3.4.2 Implications for Price Dispersion

The first and second rows of table 24 show that the baseline calibration of the model can almost perfectly match the targeted moments. In particular, the average relative price for the search good is highly correlated with unit labor costs (fact 1), changes in real relative prices are far more volatile across countries than within (fact 2), and relative prices across countries are approximately four times more volatile than the nominal exchange rate (fact 3). The shock to relative aggregate productivities is essential for matching fact 3 because it increases the dispersion of international relative prices, beyond the levels that would be created with a relative unit labor cost shock alone.

Table 25 shows that the model has reasonable implications for other moments. In particular, price-change correlations are approximately 80% within the US and Canada, and are very close to zero across countries (fact 4). The shock to relative aggregate productivities is also important for matching this final fact, since it decreases the correlation of international price changes, relative to the correlation of within-country price changes. Finally, the average relative price also closely follows the unit labor costs ratio, which is natural given the assumption of fixed nominal wages.

The model can match evidence of price dispersion using a regional bias parameter that is close to one. In fact, assuming complete segmentation yields pricing statistics that are almost identical to our baseline results, as shown in row four of table 24 and in row three of table 25. This result is consistent with the evidence presented in Gopinath et al (2011), who conclude that the US and Canadian markets are virtually fully segmented. However, we generate this result using regional segmentation alone. We return to the issue of identifying regional versus national segmentation in section 3.5.

3.4.3 The Importance of Pricing-to-Market

We next investigate to what extent our results are driven by pricing-to-market, as opposed to retailers simply sampling from producers with different marginal costs. Since producers in our model only set one price in each period, we define pricing-to-market as the tendency of producers with *equal marginal costs* to set different prices depending on the market in which they are located. We parameterize the model so that search costs are high enough that, in equilibrium, retailers always purchase from the first producer they search, and producers always charge the monopoly price (rather than one of the reservation prices). This parameterization shuts down the pricing-to-market created by the presence of different reservation prices across countries.

The third row of table 24 shows that, under this calibration, the degree of additional price dispersion created by the border is significantly reduced, though it is still evident. Yet, qualitatively, the pricing facts cited above remain unchanged. In our model, the basic pricing facts can be matched without any pricing-to-market. Further study is required in order to determine if a similar model, with detailed locations of production, can generate the pricing-to-market evidence of Burstein and Jaimovich (2009) for goods sold in both countries, but produced in a common location.

3.5 Identification

In the previous section, we demonstrated that the model of regional bias without any additional international segmentation can match the recent empirical evidence regarding price dispersion within and across countries. We now introduce the possibility that retailers have a greater chance of sampling prices from a region within their own country compared with one of the two foreign regions. This is the case of *additional* cross-country segmentation. We first show that, although the model's pricing implications remain intact when we introduce cross-country segmentation, using price data alone is insufficient to separately identify the degrees of regional versus national segmentation. We then illustrate how introducing quantity data, in particular trade relative to internal demand, can yield the desired identification. Finally, we show how while matching statistics on price dispersion implies a very high degree of segmentation, trade data suggests that markets are much more integrated across borders, at least viewed through the lens of our model.

3.5.1 Prices and Segmentation

Without loss of generality, we continue to assume that the *Home* country is relatively less expensive, with reservation prices satisfying $\widehat{p}_{r,a} < \widehat{p}_{r,c}$. For ease of exposition, we focus on the behavior of retailers in region c of the *Foreign* country, who do not search in equilibrium. As noted above, the region-c retailers sample prices from the distribution given by (3.20), reproduced here for convenience:

$$f_c^{ret}(\widehat{p}) \equiv \gamma_1 f_a(\widehat{p}) + \gamma_2 f_b(\widehat{p}) + \gamma_3 f_c(\widehat{p}) + \gamma_4 f_d(\widehat{p}). \tag{3.28}$$

The *regional* search bias can be re-expressed as $\lambda \equiv \gamma_3/\gamma_4$. Similarly, we can express the degree of *additional cross-country* segmentation induced strictly by crossing the national border with the parameter $B \equiv \gamma_4/\gamma_1$. Given symmetry across the *Home* regions,

$$f_c^{ret}(\widehat{p}) = \gamma_3 \left[\frac{1}{\lambda B} f_a(\widehat{p}) + \frac{1}{\lambda B} f_b(\widehat{p}) + f_c(\widehat{p}) + \frac{1}{\lambda} f_d(\widehat{p}) \right]. \tag{3.29}$$

We find that a broad range of combinations of the segmentation parameters λ and B are consistent with the empirical evidence on price dispersion. In fact, the two parameters cannot be independently identified using price data alone. Using the symmetry of regions within each country, $f_c^{ret}(\widehat{p}) = \gamma f_a(\widehat{p}) + (1-\gamma) f_c(\widehat{p})$, as before. Therefore, we can express the national sampling bias, γ, as a function of the two segmentation parameters, λ and

B:

$$\gamma = \frac{\lambda B + B}{\lambda B + B + 2}. \tag{3.30}$$

If aggregate output is held fixed, then reservation prices and demand, and thus the producers' pricing functions, are determined only by γ, and not by the particular breakdown between regional versus national segmentation. However, since aggregate output depends only on γ as well, there exists a continuum of different combinations of regional and national segmentation parameters that are observationally equivalent.

3.5.2 Quantities and Segmentation

In this section we show that the degree of regional versus national segmentation can be determined using additional empirical evidence on the patterns of trade within and across countries. In order to generate empirically-based measures of the different degrees of segmentation, we begin by relaxing the assumption that all regions are of equal size. Let s_r denote the size of region r, $r \in \{a, b, c, d\}$, namely the mass of producers and retailers operating in that region. Incorporating relative size differences, the relative regional bias of region-c retailers is modified to take the form

$$\frac{\gamma_3}{\gamma_4} = \lambda \frac{s_c}{s_d}, \tag{3.31}$$

so that $\lambda > 0$ now measures the degree of regional segmentation for region-c retailers that cannot be attributed to size differences between the two regions. A retailer in region c is more likely to sample a price from her own region than from region d if her own region is relatively larger, or if there is larger regional search bias, λ. For $\lambda \to \infty$, $\gamma \to 1$ and retailers in region c never search in any other region; hence the producers in region c are entirely isolated. For $\lambda = 1$, the two regions are perfectly integrated from the perspective of retailers in region c. Conversely, for $\lambda < 1$, retailers in region c

are more likely to search region d than they are to search their own region. Although this situation could arise for particular product categories, it is unlikely to apply to the representative good. Hence, we focus on the case $\lambda \geq 1$.

Similarly, the total national bias of region-c retailers is modified to take the form

$$\frac{\gamma_3}{\gamma_1} = \lambda B \frac{s_c}{s_a}, \tag{3.32}$$

with $B \geq 1$ measuring the degree of cross-country segmentation for region-c retailers that cannot be attributed to relative market size differences or to the spillover from regional segmentation.

Next, we determine the fraction of purchases by retailers in a particular region from producers in all regions, as a function of the retailers' sampling distribution. Let $Q_F \equiv \int f_c\left(\widehat{p}\right) \widehat{p} D\left(\widehat{p}\right) d\widehat{p}$ be the expected (or average) value of purchases made by a retailer, conditional on having sampled a price from the distribution of *Foreign* producer prices. Then, the quantities purchased by a region-c retailer from producers located in each of the two *Foreign* regions are given by $Q_{c,c} = \gamma_3 Q_F$ and $Q_{c,d} = \gamma_4 Q_F$. Similarly, $Q_H \equiv \int f_a\left(\widehat{p}\right) \widehat{p} D\left(\widehat{p}\right) d\widehat{p}$ is the expected value of purchases made by a retailer, conditional on having sampled a price from the distribution of *Home* producer prices. The quantities purchased a region-c retailer from producers located in each of the two *Home* regions are given by $Q_{c,a} = \gamma_1 Q_H$ and $Q_{c,b} = \gamma_2 Q_H$. From the perspective of retailers in region c, the fraction purchases made within region c, and across regions domestically, respectively, are given by

$$\tau_{c,c} = \frac{\gamma_3 Q_F}{(1-\gamma) Q_H + \gamma Q_F} \tag{3.33}$$

$$\tau_{c,d} = \frac{\gamma_4 Q_F}{(1-\gamma) Q_H + \gamma Q_F}. \tag{3.34}$$

Together with empirical estimates of the size of each regional market and an estimate for γ, these trade fractions identify regional versus national segmentation. First, note that the relative regional bias of region-c retailers is equal to the relative trade fractions,

$$\frac{\gamma_3}{\gamma_4} = \frac{\tau_{c,c}}{\tau_{c,d}}. \tag{3.35}$$

The relationships determined in equations (3.35) and (3.31) yield an estimate of the degree of regional segmentation of retailers in region c from producers in region d,

$$\lambda = \frac{\tau_{c,c}}{\tau_{c,d}} \frac{s_d}{s_c}. \tag{3.36}$$

Regional segmentation is estimated to be higher the higher is the fraction of demand that is satisfied internally and that cannot be attributed to relative size differences in the production sector. Regional sampling bias, which incorporates both regional segmentation and relative size differences, is given by[24]

$$\gamma_3 = \frac{\lambda s_c}{\lambda s_c + s_d} \gamma. \tag{3.37}$$

Using these values, the degree of excess segmentation at the border is identified using

$$\frac{\gamma}{1-\gamma} = B \frac{\lambda s_c + s_d}{s_a + s_b}. \tag{3.38}$$

The above analysis establishes that adding trade quantities to the set of moments targeted in the calibration is, in principle, sufficient to answer the question of whether

[24] The remaining sampling parameters are given by

$$\gamma_1 = \frac{s_a}{s_a + s_b} (1-\gamma)$$
$$\gamma_2 = \frac{s_b}{s_a + s_b} (1-\gamma)$$
$$\gamma_4 = \frac{s_d}{\lambda s_c + s_d} \gamma.$$

markets are segmented primarily within countries, or primarily across countries. Yet, it also raises a new challenge for theories calibrated to match pricing data alone: pricing data suggest that markets are so isolated as to effectively preclude any substantial degree of international trade. In our model, for example, the calibrated value of γ implies trade levels between the US and Canada that are well below one percent of GDP, which is clearly counter-factual. Simultaneously matching pricing evidence of market segmentation (whether regional or at the border) and the reality of substantial international trade appears a difficult task, one which we leave for future research.

3.5.3 Border Effect Regressions

This section reviews the empirical specifications that are commonly used in estimating the border effect. The original regressions of Engel and Rogers (1996) are intended to measure failures of relative LOP in city-level price indexes. They regress the time series volatility of relative real prices on distance and a border dummy variable:

$$std(d_t(i,j)) = \beta_0 + \beta_1 dist(i,j) + \beta_2^{ER} D(i,j) + X\gamma' + \varepsilon(i,j), \tag{3.39}$$

where $dist(i,j)$ is the log-distance between locations i and j, $D(i,j)$ is a dummy equal to one if the locations are in different countries, and X is a vector of variables controlling for demand characteristics in each city. Engel and Rogers (1996) find that $\hat{\beta}_2^{ER}$ is positive and significant, and refer to the magnitude as the border effect. Other papers (Parslet and Wei, 2001, Engel and Rogers, 2001) consider multi-country versions of this regression, as this allows for additional controls in this regression, notably nominal exchange rate volatility.

Arguing that using price indexes can create significant bias in the Engel and Rogers regression, Broda and Weinstein (2008) estimate the cross-sectional regression:

$$d_t(i,j)^2 = \beta_0 + \beta_1 dist(i,j) + \beta_2^{BW} D(i,j) + \varepsilon_t(i,j) \tag{3.40}$$

on product-level price data. This regression is different from the Engel and Rogers regression in a few key respects. First, since the authors have good-level prices (as opposed to indexes) the constant $\hat{\beta}_0$ can be used to test absolute LOP. Second, the regression is run cross-sectionally (t is fixed) implying that the coefficient estimates will vary over time. We study this property of the regression below.

Finally, Gopinath et al (2011) use a regression discontinuity approach to test for discrete jumps in the *price level* at the border. Their regression takes the following form:

$$p_t(i) = \beta_0 + \beta_1 dist(i,b) + \beta_2^G I(i \in Home) + \beta_3 dist(i,b) I(i \in Home) + X\gamma' + \varepsilon(i,j). \tag{3.41}$$

Here, $dist(i,b)$ represents the log-distance from location i to the countries' common border. The value of $dist(i,b)$ is positive whenever i is in the "home" country, and negative if i is a location abroad. X is a vector of variables controlling for demand characteristics in each city.

In regressions (3.39)-(3.41), rejections of the null $\beta_2 = 0$ are typically taken as evidence of segmentation. We simulate data from the model with and without any segmentation, and show that estimates of β_2 from the simulated data cannot correctly identify whether (or the degree to which) markets are segmented. We view our model as a model of residual price dispersion once distance and other variables are controlled for. Accordingly, we fix the distance and the controls coefficients to zero.

To better understand the properties of the regression measures of the border effect, we study the empirical distributions of the coefficients using data simulated from the model. The first row of table 26 gives some basic statistics for the ER regression under the baseline calibration. The ER regressions are performed on 1,000 time series of 100 periods each, with 15 cities per country. Column one shows that the mean of $\hat{\beta}_2^{ER}$ is greater than zero. Furthermore, the first row of column two shows that, according a standard t-test, we always reject the null $\beta_2^{ER} = 0$. Under the baseline calibration, we would conclude (under the standard interpretation) that markets are more segmented across borders one-hundred percent of time, even though the model takes no stand on this fact.

The second and third rows of the table demonstrate results using the cross-sectional regressions of Broda and Weinstein (2008) and Gopinath et al (2011). These regressions are performed on a single time-series simulation of 1,000 periods, with 50 locations per country. Again, on average, these regressions would typically lead to the conclusion of greater segmentation across the border. Importantly, however, the regressions each do not reject the null of $\hat{\beta} = 0$ around ten percent of the time. This suggestions that disagreements about the size of border effect, e.g. between the authors of these two papers, could well be explained by the time period in which the data for the regressions is collected.

Figures 25 and 26 show estimated coefficients $\hat{\beta}_{BW}$ and $\hat{\beta}_G$ over a simulated 50-quarter time span, along with the nominal exchange rate. The coefficients are highly time-varying, and strongly correlated with the nominal exchange rate, in much the same way as the real exchange rate is correlated with the nominal exchange rate. Furthermore, this high correlation holds as long as there is some degree of segmentation (namely, $\gamma > 0.5$). According to our model, these cross-sectional regressions are informative about the state of exchange rates far more than they are about the extent of segmentation within or between countries.

Finally, table 27 shows regression results for simulations of a world with no segmentation ($\gamma = 0.5$), but with cross-country differences in the distributions of producers' idiosyncratic shocks. In this case, the regressions correctly fail to reject a zero border effect. This result stands in contrast to Gorodnichenko and Tesar (2008), who argue that such asymmetries can lead to falsely rejecting the zero effect null. There is no contradiction, however. In our model retailers are sampling prices with equal probability from producers in both countries. Since retailers in both countries have the same reservation price, they never re-sample a second price within a period, and the prices paid by all retailers are identically distributed. It is important to note, however, that we would get an entirely different result if we considered the price posted by producers, rather than prices paid by retailers. These two distributions are not equal, as shown in figure 20, and would therefore falsely imply segmentation.

3.6 Conclusion

We have demonstrated that a simple model of customer search can replicate the most prominent facts about international real exchange rates at the good-level. It can do so without relying on any additional friction in international trade. While the extent of the border effect remains in some dispute, we have shown that the standard interpretation of the border effect as a measure of market segmentation may be quite misleading. In our model, price data alone are not sufficient to answer the question of whether international borders create market segmentation beyond that which already occurs within countries. Incorporating data on trade shares resolves this identification problem, but raises a tension between the high dispersion of observed prices across the border and the relatively large quantity of international trade that occurs between countries. Other models incorporating endogenous price dispersion are likely to generate similar difficulties.

One possible objection to the calibration of the model is the high volatility of average productivity across countries. This shock is crucial for matching both the volatility of international prices relative to the exchange rate, and the low correlation of international price changes. A more general version of the model could include many sectors and, therefore, sector-specific productivity shocks. Both of the roles played by the aggregate shock could then be played by sector specific shocks which are more correlated within countries than across.[25] This would permit matching the targeted moments without resorting to large aggregate shocks.

In our view, model-based structural estimation represents a promising avenue forward for exploring the extent of barriers to international trade. Identifying the degree of market segmentation, however, is likely to require data on variables other than prices, most notably quantities. Although the model we have presented here incorporates a reduced-form wage friction, we leave for future work the introduction of nominal price stickiness. Although it cannot account for the data on its own, it is possible that in a dynamic setting, price stickiness interacts in an important way with our search friction, supporting the persistence of price dispersion.

[25] Burstein and Jaimovich (2009) create a similar effect using Bertrand competition, assuming that firms always face the same latent competitor within countries, but only sometimes face the same competitor across countries.

Table 23: Parameters values for the baseline model calibration.

Parameter	η	κ	$\alpha_1 = \gamma_3$	ρ_{ms}	σ_{ms}	$\sigma_{\zeta,us}$	$\sigma_{\zeta,can}$	σ_{ϵ}
Value	5.000	0.006	0.998	0.950	0.029	0.093	0.052	0.059

Table 24: Targeted model moments. Moments are computed directly from policy functions and a discretized approximation to the shock processes.

	$\bar{\mu}_{us}$	$\sigma_{\Delta d_{t,us}}$	$\sigma_{\Delta d_{t,can}}$	$\sigma_{\Delta d_{t,bord}}$	$\rho(\bar{d}_{t,bord}, log(e_t))$	$\frac{\sigma_{\Delta d_{t,bord}}}{\sigma_{\Delta \log(e)}}$
Target	0.150	0.060	0.050	0.130	0.700	4.333
Baseline Calibration	0.150	0.061	0.050	0.131	0.696	4.377
No P.T.M.	0.223	0.184	0.103	0.191	0.699	6.377
Pefect Segmentation	0.150	0.061	0.050	0.136	0.719	4.531

Table 25: Other model moments not targeted.

	$\bar{\mu}_{can}$	$\sigma_{\Delta log(e)}$	$\rho(\bar{d}_{t,bord}, rw_t)$	$\rho_{\Delta p_t,us}$	$\rho_{\Delta p_t,can}$	$\rho_{\Delta p_t,bord}$
Baseline Calibration	0.195	0.030	0.696	0.789	0.849	0.010
No P.T.M.	0.223	0.030	0.699	0.300	0.576	-0.000
Pefect Segmentation	0.195	0.030	0.719	0.801	0.854	-0.026

Table 26: Simulation of regression results for baseline calibration - 1000 exemplars of each regression.

	Mean($\hat{\beta}_2$)	**% Reject**
Engel-Rogers (1996)	0.066	1.000
Broda-Weinstein (2008)	0.018	0.794
Gopinath et al. (2009)	0.073	0.861

Table 27: Simulation of regression results under no segmentation, country asymmetry - 1000 exemplars of each regression.

	Mean($\hat{\beta}_2$)	**% Reject**
Engel-Rogers (1996)	-0.000	0.006
Broda-Weinstein (2008)	-0.000	0.012
Gopinath et al. (2009)	0.000	0.041

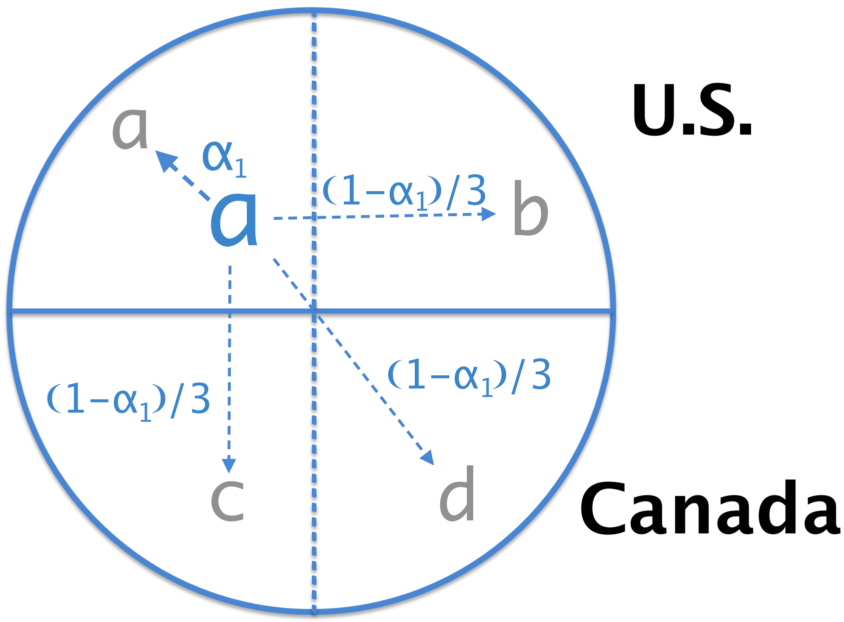
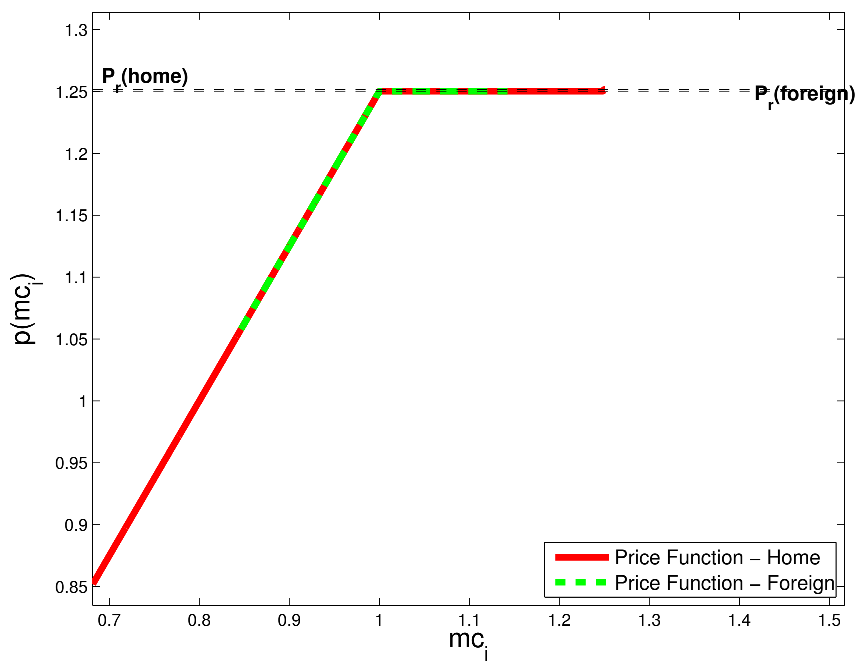
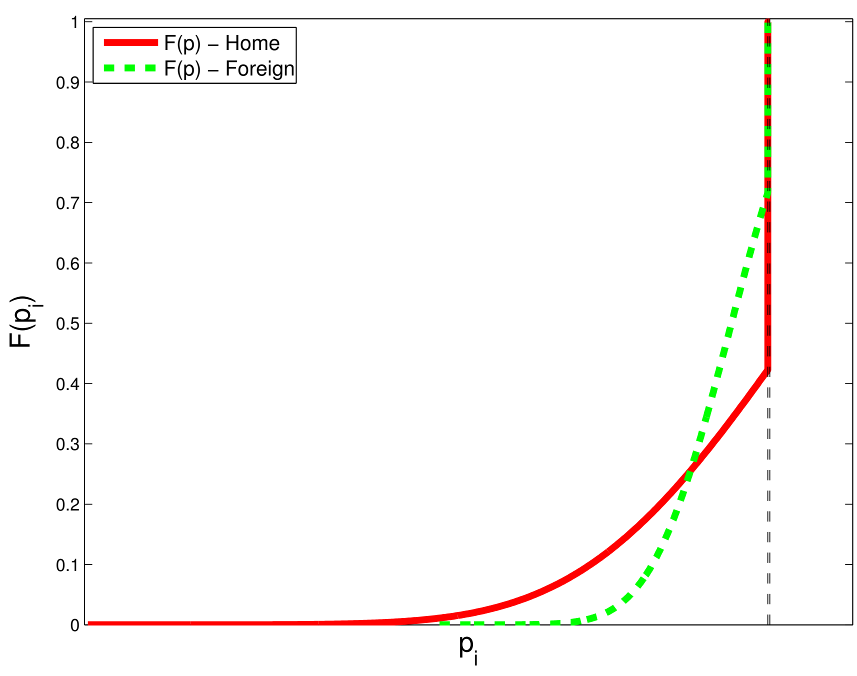


Figure 19: The two-country setup.

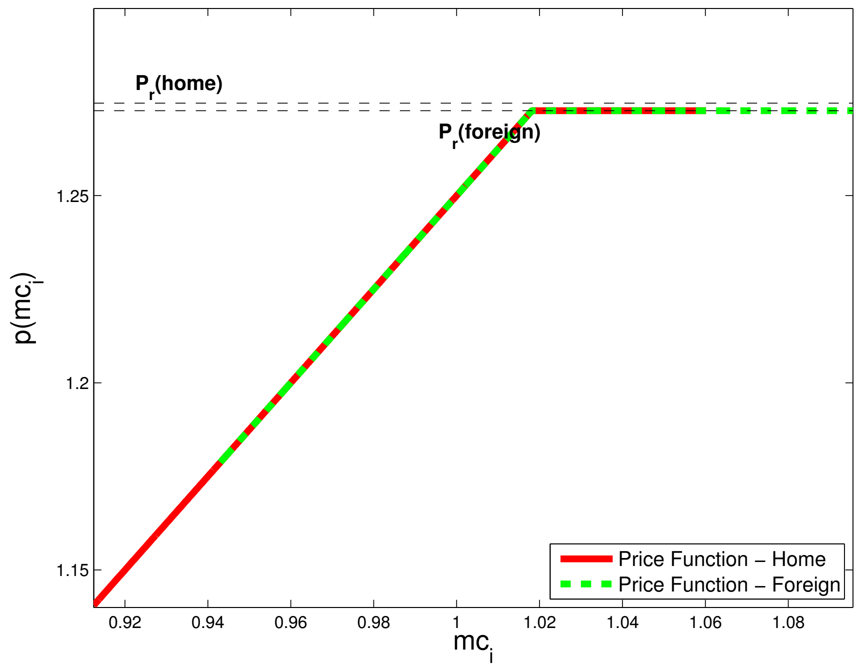


(a)

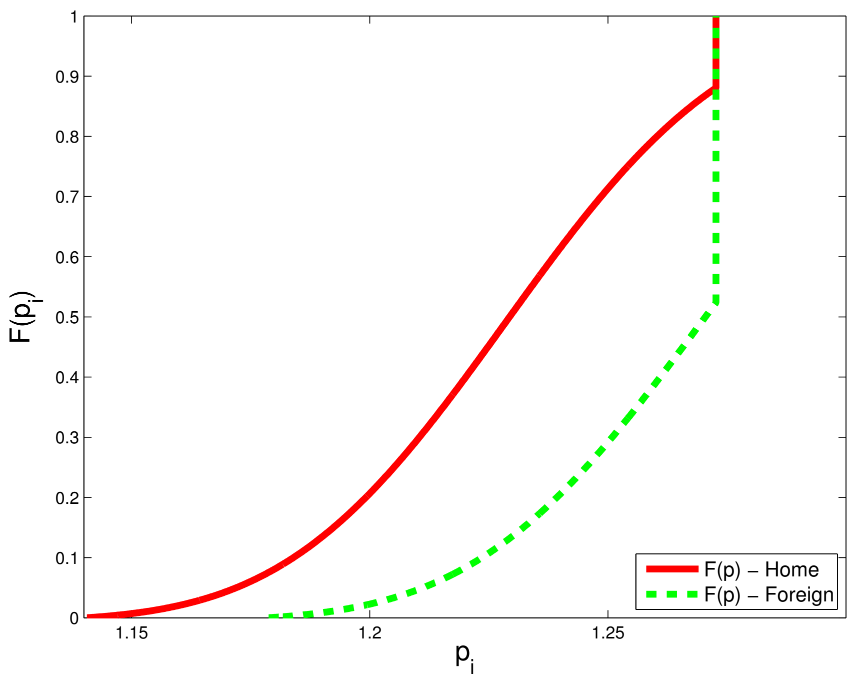


(b)

Figure 20: Pricing functions and price distributions for symmetric, unsegmented economies with different cost dispersions ($\sigma_{\zeta,H} > \sigma_{\zeta,F}$).

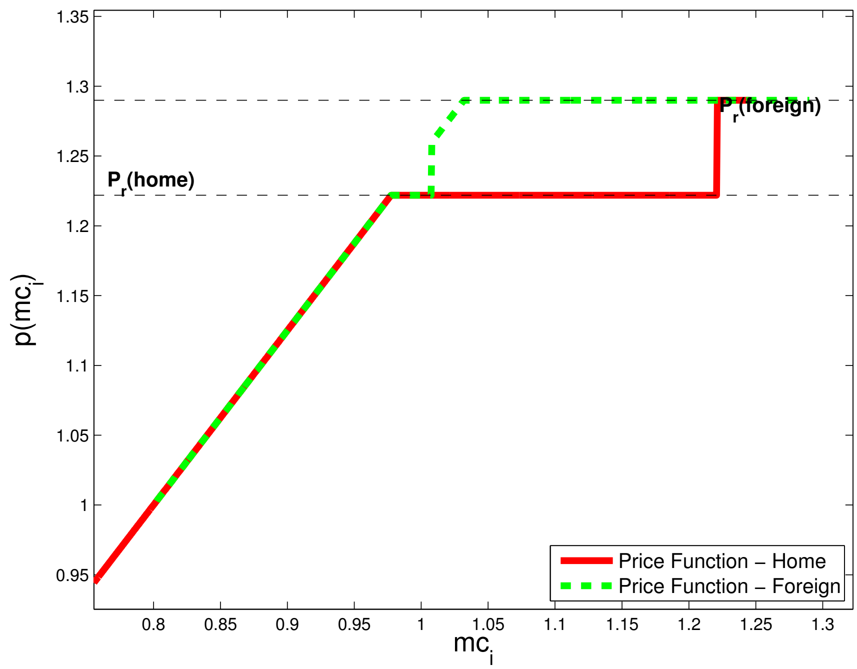


(a)

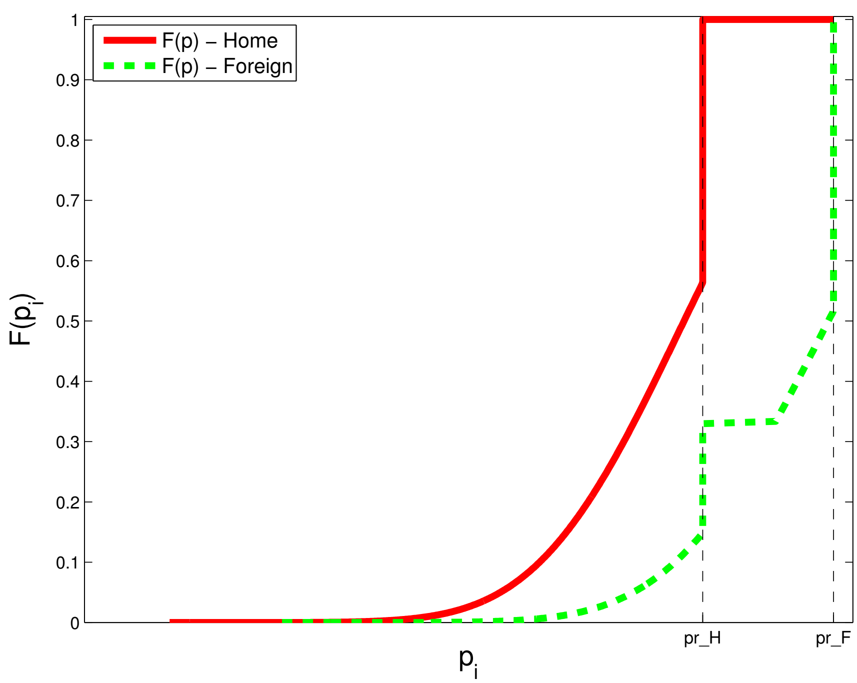


(b)

Figure 21: Pricing functions and price distributions for symmetric, unsegmented economies with different unit labor costs ($w_H < w_F$).

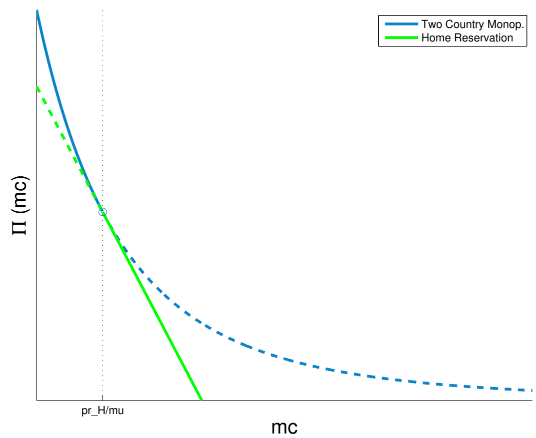


(a)

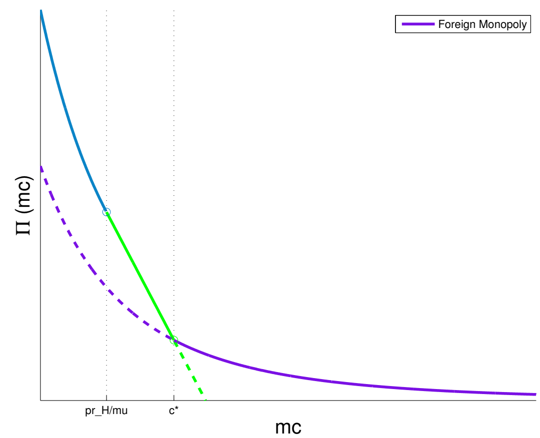


(b)

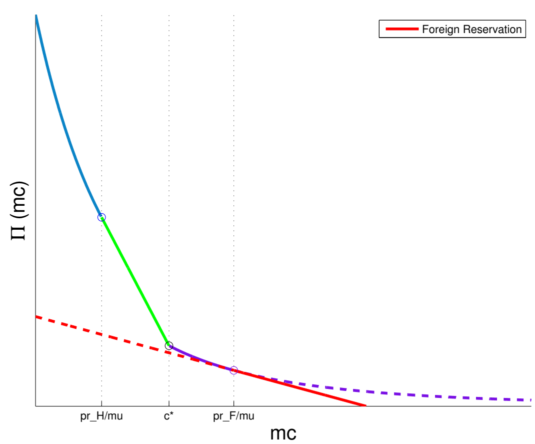
Figure 22: Pricing functions and pricing distributions for symmetric, *segmented* economies with different unit labor costs ($w_H < w_F$).



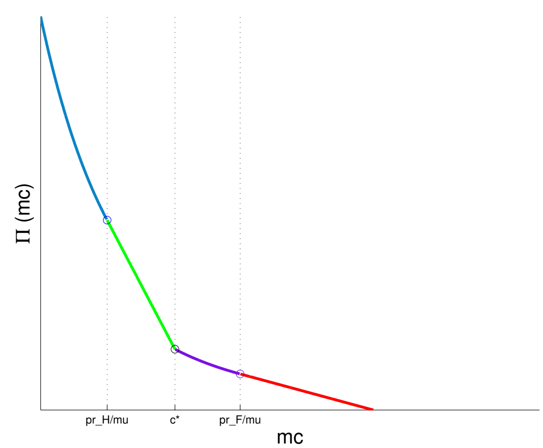
(a) When marginal cost is below p_r^H/μ, Foreign firms charge the two-country monopoly price.



(b) When marginal cost is between p_r^H/μ and c^*, Foreign firms charge the home reservation price.



(c) When marginal cost is between c^* and the p_r^F/μ, Foreign firms charge the foreign monopoly price.



(d) When marginal cost is greater than p_r^F/μ, Foreign firms charge the foreign reservation price.

Figure 23: Profit as a function of marginal cost under different pricing policies.

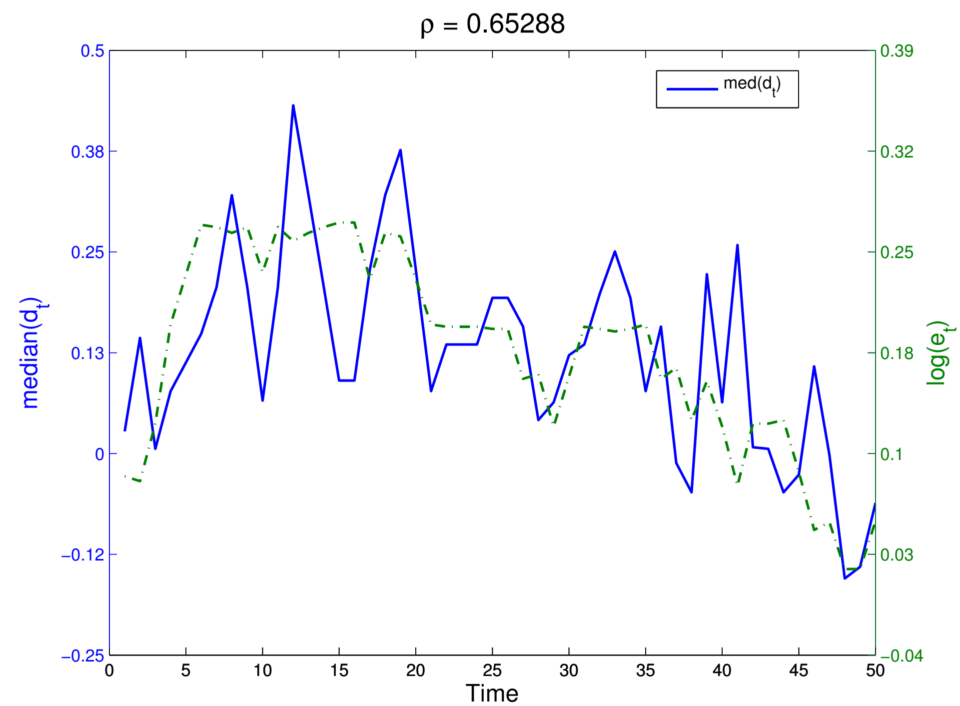


Figure 24: Median real and nominal exchange rate for a single time series realization under the baseline parameterization.

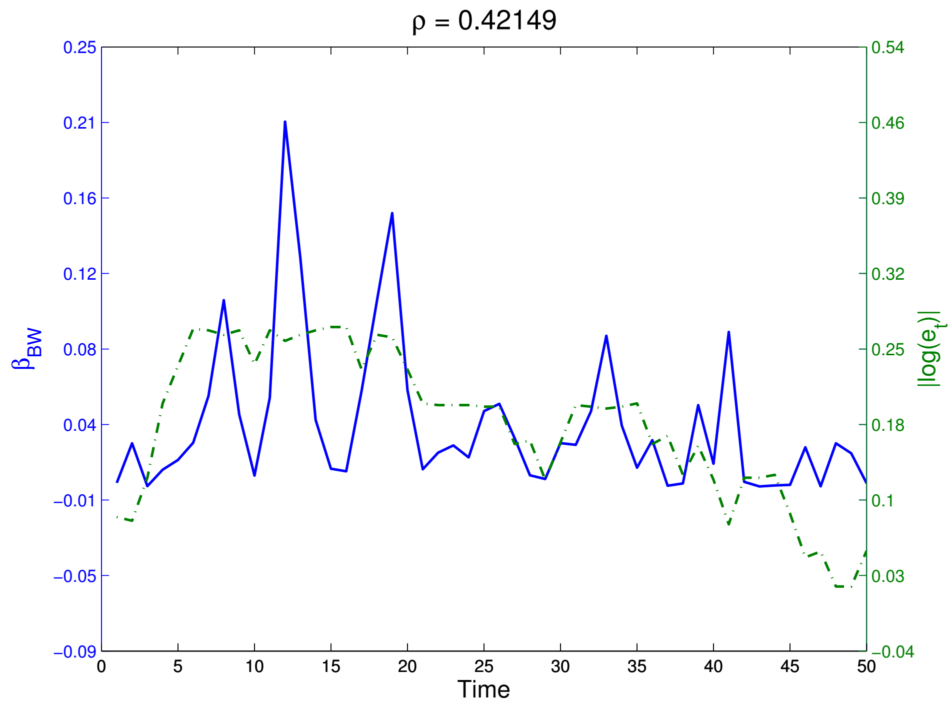


Figure 25: Time series of the nominal exchange rate versus the border coefficient in the regression of Broda and Weinstein (2008)

.

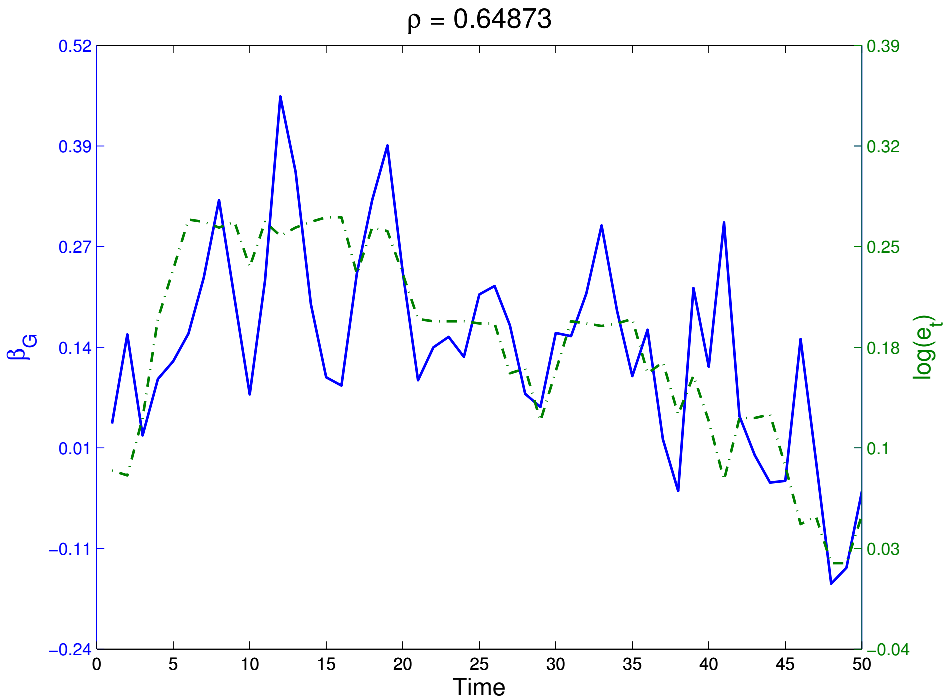


Figure 26: Time series of the nominal exchange rate versus the border coefficient in the regression of Gopinath et al (2011). The estimated coefficient and the exchange rate are highly correlated.

References

[1] Abe, Naohito and Akiyuki Tonogi (2010), "Micro and Macro Price Dynamics in Daily Data," Journal of Monetary Economics, 57(6): 716-728.

[2] Alessandria, George (2009), "Consumer Search, Price Dispersion, and International Relative Price Fluctuations," International Economic Review, 50(3): 803-829.

[3] Arimoto, Suguru (1973), "An Algorithm for Calculating the Capacity of an Arbitrary Discrete Memoryless Channel," IEEE Trans. Inf. Theory, IT-18: 14-20.

[4] Bai, Jushan and Pierre Perron (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," Econometrica, 66(1): 47-78.

[5] Berger, Toby (1971), "Rate Distortion Theory," Prentice-Hall Inc., Englewood Cliffs, NJ.

[6] Bils, Mark and Peter J. Klenow (2004), "Some Evidence on the Importance of Sticky Prices," Journal of Political Economy, 112: 947-985.

[7] Blahut, Richard E. (1972), "Computation of Channel Capacity and Rate Distortion Functions," IEEE Trans. Inf. Theory, IT-18: 460-473.

[8] Broda, Christian and David E. Weinstein (2008), " Understanding International Price Differences Using Barcode Data," working paper 14017, National Bureau of Economic Research.

[9] Brodsky, B. and B. Darkhovsky (1993), Nonparametric Methods in Change-Point Problems. Kluver Academic Publishers, the Netherlands.

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

[11] Burstein, Ariel and Christian Hellwig (2007), "Prices and Market Shares in a Menu Cost Model," working paper, UCLA.

[12] Burstein, Ariel and Nir Jaimovich (2009), " Understanding Movements in Aggregate Product-Level Real-Exchange Rates," unpublished paper, UCLA.

[13] Caballero, Ricardo J., and Eduardo Engel (2007), "Price Stickiness in Ss Models: New Interpretations of Old Results," Journal of Monetary Economics, 54(S1): 100-121.

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

[15] Campbell, Jeffrey R. and Benjamin Eden (2005), "Rigid Prices: Evidence from US Scanner Data," unpublished paper, Federal Reserve Bank of Chicago.

[16] Caplin, A., and D. Spulber (1987), "Menu Costs and the Neutrality of Money," Quarterly Journal of Economics, 102(4): 703-725.

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

[18] Chahrour, Ryan A. (2011), "Sales and Price Spikes in Retail Scanner Data," Economic Letters, 110: 143-146.

[19] Chevalier, Judith. A., Anil K. Kashyap and Peter E. Rossi (2003), "Why Don't Prices Rise during Periods of Peak Demand," American Economic Review, 93(1): 15-37.

[20] Conover, William J. (1972), "A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions," Journal of the American Statistical Association, 67(339): 591-596.

[21] Conover, William J. (1980), Practical Nonparametric Statistics. New York: Wiley.

[22] Cover, Thomas M. and Joy A. Thomas (2006), Elements of Information Theory, second ed. Wiley-Interscience, New York.

[23] Csiszar, I. (1974), "On the Computation of Rate Distortion Functions," IEEE Trans. Inf. Theor, IT 20: 122-124.

[24] Deshayes, Jean and Dominique Picard (1986), "Off-Line Statistical Analysis of Change-Point Models using Nonparametric and Likelihood Methods," Detection of Abrupt Changes in Signals and Dynamical Systems, New York: Springer-Verlag: 103-168.

[25] Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo (2011), "Reference Prices, Costs, and Nominal Rigidities," American Economic Review, 101(1): 234-262.

[26] Engel, Charles and John H. Rogers (1996), "How Wide is the Border?" American Economic Review, 86(5): 1112-1125.

[27] Engel, Charles and John H. Rogers (2001), "Deviations From Purchasing Power Parity: Causes and Welfare Costs," Journal of International Economics, 55(1): 29-57.

[28] Fix, Stephen, L. (1978), "Rate Distortion Functions for Squared Error Distortion Measures," Proceedings of the Sixteenth Annual Allerton Conference on Communication, Control, and Computing: 704-711.

[29] Golosov, Michael, and Robert E. Lucas, Jr. (2007), "Menu Costs and Phillips Curves," Journal of Political Economy, 115(2): 171-199.

[30] Gopinath, Gita, Pierre-Olivier Gourinchas, Chang-Tai Hsieh, and Nicholas Li (2009), "Estimating the Border Effect: Some New Evidence," working paper 14938, National Bureau of Economic Research

[31] Gopinath, Gita, Pierre-Olivier Gourinchas, Chang-Tai Hsieh, and Nicholas Li (2011), "International Prices, Costs, and Markup Differences," American Economic Review, 101(6): 2450-2486.

[32] Gorodnichenko, Yuriy and Linda Tesar (2008), "Border Effect or Country Effect? Seattle May Not be So Far From Vancouver After All," American Economic Journal: Macroeconomics, 1(1): 219-241.

[33] Guimaraes, Bernardo and Kevin D. Sheedy (2011), "Sales and Monetary Policy," American Economic Review, 101(2): 844-876.

[34] Hoch, Stephen, Xavier Dreze and Mary E Purk (1994), "EDLP, Hi-Lo, and Margin Arithmetic," Journal of Marketing, 58: 16-27.

[35] Kehoe, Patrick and Virgiliu Midrigan (2010), "Prices Are Sticky After All," NBER working paper 16364.

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

[37] Klenow, Peter J. and Benjamin A. Malin (2010), "Microeconomic Evidence on Price-Setting," in B.M. Friedman and M. Woodford, eds., Handbook of Monetary Economics, Amstedam: Elsevier Press.

[38] Klenow, Peter J. and Jonathan L. Willis (2007), "Sticky Information and Sticky Prices," Journal of Monetary Economics, 54: 79-99.

[39] Levy, D., H. Chen, G. Mueller, S. Dutta and M. Bergen (2010), "Holiday Price Rigidity and Cost of Price Adjustment," Economica, 77: 172-198.

[40] Levy, D., D. Lee, H. Chen, R. J. Kauffman and M. Bergen (2011), "Price Points and Price Rigidity," Review of Economics and Statistics, 93(4): 1417-1431.

[41] Mackowiak, Bartosz and Mirko Wiederholt (2009a), "Optimal Sticky Prices under Rational Inattention," American Economic Review, 99(3): 769-803.

[42] Mackowiak, Bartosz and Mirko Wiederholt (2009b), "Business Cycle Dynamics Under Rational Inattention," discussion paper, European Central Bank and Northwestern University.

[43] Mankiw, N. Gregory, and Ricardo Reis (2002), "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," Quarterly Journal of Economics, 117: 1295-1328.

[44] Matejka, Filip (2011), "Rationally Inattentive Seller: Sales and Discrete Pricing," unpublished paper, CERGE-EI.

[45] Matejka, Filip, and Christopher A. Sims (2010), "Discrete Actions in Information-Constrained Tracking Problems," discussion paper, Princeton University.

[46] Midrigan, Virgiliu (2011), "Menu Costs, Multi-Product Firms, and Aggregate Fluctuations," Econometrica, 79: 1139-1180.

[47] Nakamura, Emi and Jon Steinsson (2008), "Five Facts about Prices: A Reevaluation of Menu Cost Models," Quarterly Journal of Economics, 123(4): 1415-1464.

[48] Parsley, David C. and Shang-Jin Wei (2001), "Explaining the Border Effect: the Role of Exchange Rate Variability, Shipping Costs, and Geography," Journal of International Economics, 55(1): 87-105.

[49] Reinganum, Jennifer F. (1979), "A Simple Model of Equilibrium Price Dispersion," The Journal of Political Economy, 87 (4): 851-858.

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

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

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

[53] Shannon, Claude E. (1959), "Coding Theorems for a Discrete Source with a Fidelity Criterion," IRE Nat. Conv. Rec., Pt 4: 142-163.

[54] Sheshinski, Eytan, and Yoram Weiss (1977), "Inflation and Costs of Price Adjustment," Review of Economic Studies, 54: 287-303.

[55] Sims, Christopher A. (1998), "Stickiness," Carnegie-Rochester Conference Series on Public Policy, 49(1): 317-356.

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

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

[58] Sims, Christopher A. (2010), "Rational Inattention and Monetary Economics," in B.M. Friedman and M. Woodford, eds., Handbook of Monetary Economics, vol 3A, Amstedam: Elsevier Press.

[59] Taylor, John B. (1980), "Aggregate Dynamics and Staggered Contracts," Journal of Political Economy, 88: 1-24.

[60] Wood, L. Constance and Michele M. Altavela (1978), "Large-Sample Results for Kolmogorov-Smirnov Statistics for Discrete Distributions," Biometrika, 65(1): 235-239.

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

[62] Woodford, Michael (2008), "Inattention as a Source of Randomized Discrete Adjustment," unpublished paper, Columbia University.

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

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

A Addendum to Chapter 1

A.1 Implementation of Filters

The V-shaped Sales Filter

I implement the v-shaped sales filter proposed by Nakamura and Steinsson (2008). The algorithm requires choosing four parameters: J, K, L, F. The parameter J is the period of time within which a price cut must return to the regular price in order to be considered a transitory sale. For asymmetric v-shaped sales, in which 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. In the case of asymmetric sales, the parameter F determines whether to associate the sale with the existing regular price or with the new one.

1. $r_0 = p_0$
2. If $p_t = r_{t-1}$, then $r_t = r_{t-1}$
3. Else, if $p_t > r_{t-1}$, then $r_t = p_t$
4. Else, if $r_{t-1} \in \{p_{t+1}, ..., p_{t+J}\}$, and the price never rises above r_{t-1} before returning to r_{t-1}, then $r_t = r_{t-1}$
5. Else, if the set $\{p_{t+1}, ..., p_{t+L}\}$ has K or more distinct prices, then $r_t = p_t$
6. Else, define $p_{\max} = \max\{p_{t+1}, ..., p_{t+L}\}$ and $t_{\max}$ = first occurrence of $p_{\max}$. If $p_{\max} \in \{p_{t\max+1}, ..., p_{t\max+L}\}$, then
   (a) if $F = 1$, then $r_t = p_{\max}$
   (b) elseif $F = 0$, then $r_t = r_{t-1}$
7. Else, $r_t = p_t$.

The Reference Price Filter

The reference price filter proposed by Eichenbaum, Jaimovich and Rebelo (2011) requires one parameter, W, the width of the fixed window.

1. Divide each price series into non-overlapping intervals of length W.

2. For each interval, compute the modal price, p^R.

The Rolling Mode Filter

The rolling mode filter proposed by Kehoe and Midrigan (2010) requires two parameters: W, the width of the rolling window, and C, the minimum required frequency for the modal price to count as a regular price.

1. For each rolling window of width W

   (a) compute p_w^M, the modal price

   (b) compute f_w, the fraction of observations with $p_t = p_w^M$

2. For each t, set p_t^R, the regular price, equal to the modal price for that t, if $f_w \geq C$

B Addendum to Chapter 2

B.1 Proofs

Proof of lemma 1. See Woodford (2008). □

Proof of corollary 1. Follows from lemma 1. □

Proof of lemma 2. Prices are distributed independently of states conditional on signals. As a result, by the data-processing inequality theorem (Cover and Thomas, 2006), the relative entropy between prices and states is weakly less than the relative entropy between signals and states. If prices are a random function of signals, then the inequality is strict. □

Proof of corollary 2. Follows from lemma 2. □

Proof of lemma 3. Recall that the objective is given by

$$\Pi_\tau\left(\varpi_\tau\right) \equiv \sum_{q\in\mathcal{Q}} f_\tau\left(q|\varpi_\tau\right)\left\{\pi(q-y_\tau) - \theta^p\left[\log f_\tau\left(q|\varpi_\tau\right) - \log \overline{f}\left(q\right)\right]\right\}. \tag{B.1}$$

Forming the Lagrangian with multipliers μ and $\eta\left(q\right)$ on the constraints specified in equations (2.46) and (2.47),

$$\begin{aligned}\mathcal{L}(f) \quad = \quad & \sum_{q\in\mathcal{Q}} f_\tau\left(q|\varpi_\tau\right)\pi(q-y_\tau) - \theta^p \sum_{q\in\mathcal{Q}} f_\tau\left(q|\varpi_\tau\right)\left[\log f_\tau\left(q|\varpi_\tau\right) - \log \overline{f}\left(q\right)\right] \\ & -\mu \sum_{q\in\mathcal{Q}} f_\tau\left(q|\varpi_\tau\right) - \sum_{q\in\mathcal{Q}} \eta\left(q\right) f_\tau\left(q|\varpi_\tau\right).\end{aligned} \tag{B.2}$$

For $f_\tau(q|\varpi) > 0$, such that $\eta(q) = 0$, differentiating $\mathcal{L}(f)$ with respect to $f_\tau(q|\varpi)$, for a fixed $\overline{f}(q)$, yields

$$\frac{1}{\theta^p}\pi(q - y_\tau) - \left[\log f_\tau(q|\varpi_\tau) - \log \overline{f}(q)\right] - 1 - \frac{\mu}{\theta^p} = 0 \quad \Leftrightarrow \tag{B.3}$$

$$\log f_\tau(q|\varpi_\tau) - \log \overline{f}(q) = \frac{1}{\theta^p}\pi(q - y_\tau) - \left(1 + \frac{\mu}{\theta^p}\right). \tag{B.4}$$

Letting $\phi \equiv \exp\left\{1 + \frac{\mu}{\theta^p}\right\}$,

$$\log f_\tau(q|\varpi_\tau) - \log \overline{f}(q) = \log\left[\exp\left\{\frac{1}{\theta^p}\pi(q - y_\tau)\right\}\right] - \log\phi \quad \Leftrightarrow \tag{B.5}$$

$$f_\tau(q|\varpi_\tau) = \frac{1}{\phi}\overline{f}(q)\exp\left\{\frac{1}{\theta^p}\pi(q - y_\tau)\right\}. \tag{B.6}$$

Summing over q, we obtain

$$\phi = \sum_{\widehat{q}\in\mathcal{Q}} \overline{f}(\widehat{q})\exp\left\{\frac{1}{\theta^p}\pi(\widehat{q} - y_\tau)\right\}, \tag{B.7}$$

which yields

$$f_\tau(q|\varpi_\tau) = \overline{f}(q)\frac{\exp\left\{\frac{1}{\theta^p}\pi(q - y_\tau)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}} \overline{f}(\widehat{q})\exp\left\{\frac{1}{\theta^p}\pi(\widehat{q} - y_\tau)\right\}}. \tag{B.8}$$

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

The conditional distribution, $f_\tau(q|\varpi_\tau)$, 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 $f(q|y_\tau)$, for all $\tau \geq 0$, and for each normalized target price y_τ in each ϖ_τ. □

Proof of lemma 4 . From lemma 3, the firm's per-period profit net of the cost of the price signal is an invariant function, $\Pi(y)$, for all $y \in \mathcal{Y}$. 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. Hence, equation (2.50) can alternatively be written as

$$E_\tau\left\{\Pi(y_\tau)+\sum_{\tau'=\tau+1}^{\infty}\beta^{\tau'-\tau}\Gamma_{\tau,\tau'}(\widetilde{\varpi}_{\tau'-1})\left[\begin{array}{c}(1-\Lambda_{\tau'}(\widetilde{\varpi}_{\tau'}))\,\Pi(y_{\tau'})\\ +\Lambda_{\tau'}(\widetilde{\varpi}_{\tau'})\left[\overline{V}-\kappa\right]\\ -\theta^r I^r\left(\Lambda_{\tau'}(\widetilde{\varpi}_{\tau'}),\overline{\Lambda}\right)\end{array}\right]\right\} \qquad \text{(B.9)}$$

Recall that $\widetilde{y}_\tau$ is a random walk and $y_\tau = \widetilde{y}_\tau + \nu_\tau$, where ν_τ is i.i.d. Therefore, the probability distributions for realizations of $\widetilde{y}_{\tau'}$ and $y_{\tau'}$ conditional on $\widetilde{\varpi}_\tau$ depend only on the value of $\widetilde{y}_\tau$ for any $\tau' \geq \tau$. The maximum attainable value of the objective specified in equation (B.9) must therefore only depend on the value of $\widetilde{y}_\tau$, $V_\tau(\widetilde{\varpi}_\tau) = V(\widetilde{y}_\tau)$, for some invariant function $V(\widetilde{y})$.

The problem of maximizing the objective specified in equation (B.9) has the recursive form

$$V(\widetilde{y}_\tau)=\max_{\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})} E_\tau\left\{\Pi(y_\tau)+\beta\left[\begin{array}{c}(1-\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}))\,V(\widetilde{y}_{\tau+1})\\ +\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})\left[\overline{V}-\kappa\right]\\ -\theta^r I^r\left(\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}),\overline{\Lambda}\right)\end{array}\right]\right\}, \qquad \text{(B.10)}$$

where $E_\tau\{\cdot\}$ integrates over all possible innovations to the state, $\widetilde{\varpi}_{\tau+1}$, that follow $\widetilde{\varpi}_\tau$ under the current review policy. The expression (B.10) defines the problem of finding the optimal hazard as a function of the difference between the value of updating to a new policy, $\overline{V}-\kappa$, and the value of continuing with the existing policy, $V(\widetilde{y})$, net of the information expenditure required to receive the signal from this hazard function.

For each state $\widetilde{\varpi}_{\tau+1}$, the hazard function $\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})$ is chosen to maximize

$$(1-\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}))\,V(\widetilde{y}_{\tau+1})+\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})\left[\overline{V}-\kappa\right]-\theta^r I^r\left(\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}),\overline{\Lambda}\right). \qquad \text{(B.11)}$$

This problem, and hence its solution, 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 (B.11) with respect to $\Lambda(\widetilde{y}_{\tau+1})$ yields

$$\overline{V} - \kappa - V(\widetilde{y}_{\tau+1}) - \theta^r \frac{\partial I^r\left(\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1}), \overline{\Lambda}\right)}{\partial \Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})} = 0, \tag{B.12}$$

where

$$\frac{\partial I^r\left(\Lambda, \overline{\Lambda}\right)}{\partial \Lambda} = \log \frac{\Lambda}{1-\Lambda} - \log \frac{\overline{\Lambda}}{1-\overline{\Lambda}}. \tag{B.13}$$

Hence

$$\frac{\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})}{1-\Lambda_{\tau+1}(\widetilde{\varpi}_{\tau+1})} = \frac{\overline{\Lambda}}{1-\overline{\Lambda}} \exp\left\{\frac{1}{\theta^r}\left[\overline{V} - \kappa - V(\widetilde{y}_{\tau+1})\right]\right\} \tag{B.14}$$

for all $\widetilde{\varpi}_{\tau+1}$ and all $\tau > 0$, which can be written directly as

$$\frac{\Lambda(\widetilde{y})}{1-\Lambda(\widetilde{y})} = \frac{\overline{\Lambda}}{1-\overline{\Lambda}} \exp\left\{\frac{1}{\theta^r}\left[\overline{V} - \kappa - V(\widetilde{y})\right]\right\}, \tag{B.15}$$

for all $\widetilde{y} \in \widetilde{\mathcal{Y}}$.

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

$$V(\widetilde{y}) = E\left\{\Pi(y) + \beta\left[(1-\Lambda(\widetilde{y}'))V(\widetilde{y}') + \Lambda(\widetilde{y}')\left[\overline{V} - \kappa\right] - \theta^r I^r\left(\Lambda(\widetilde{y}'), \overline{\Lambda}\right)\right]\right\}, \tag{B.16}$$

where $E\{\cdot\}$ 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

$$\overline{V} = V(0). \tag{B.17}$$

□

Proof of lemma 5 . Given the definition of $I^r\left(\Lambda\left(\widetilde{y}_\tau\right), \overline{\Lambda}\right)$ in equation (2.12), the minimization of the cost of the review policy given in equation (2.56) is equivalent to the maximization of

$$E\left\{\sum_{\tau=1}^{\infty}\beta^\tau\Gamma_\tau\left(\widetilde{y}^{\tau-1}\right)\left[\Lambda\left(\widetilde{y}_\tau\right)\log\overline{\Lambda}+\left(1-\Lambda\left(\widetilde{y}_\tau\right)\log\left(1-\overline{\Lambda}\right)\right)\right]\right\}. \tag{B.18}$$

The first order condition is, for $\overline{\Lambda}\in(0,1)$,

$$E\left\{\sum_{\tau=1}^{\infty}\beta^\tau\Gamma_\tau\left(\widetilde{y}^{\tau-1}\right)\left[\frac{\Lambda\left(\widetilde{y}_\tau\right)}{\overline{\Lambda}}-\frac{1-\Lambda\left(\widetilde{y}_\tau\right)}{\left(1-\overline{\Lambda}\right)}\right]\right\}=0 \;\Leftrightarrow \tag{B.19}$$

$$E\left\{\sum_{\tau=1}^{\infty}\beta^\tau\Gamma_\tau\left(\widetilde{y}^{\tau-1}\right)\left[\Lambda\left(\widetilde{y}_\tau\right)-\overline{\Lambda}\right]\right\}=0. \tag{B.20}$$

Rearranging yields

$$\overline{\Lambda}=\frac{E\left\{\sum_{\tau=1}^{\infty}\beta^\tau\Gamma_\tau\left(\widetilde{y}^{\tau-1}\right)\Lambda\left(\widetilde{y}_\tau\right)\right\}}{E\left\{\sum_{\tau=1}^{\infty}\beta^\tau\Gamma_\tau\left(\widetilde{y}^{\tau-1}\right)\right\}}. \tag{B.21}$$

□

Proof of lemma 6. Follows from lemma 5 and the law of motion for the normalized pre-review state, $\widetilde{y}$. □

Proof of lemma 7. Forming the Lagrangian with multiplier μ, the first order condition for each q charged with positive probability yields

$$E\left\{\sum_{\tau=0}^{\infty}\beta^\tau\Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)\frac{f\left(q|y\right)}{\overline{f}\left(q\right)}\right\}=\mu. \tag{B.22}$$

Summing over q yields an expression for the multiplier,

$$\mu = E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)\right\}. \tag{B.23}$$

Hence,

$$\overline{f}\left(q\right) = \frac{E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)f\left(q|y\right)\right\}}{E\left\{\sum_{\tau=0}^{\infty}\beta^{\tau}\Gamma_{\tau+1}\left(\widetilde{y}^{\tau}\right)\right\}}. \tag{B.24}$$

□

Proof of lemma 8. Follows from lemma 7 and the laws of motion for the normalized states $\widetilde{y}$ and y. □

Proof of lemma 9. The proof follows from the strict concavity of (2.69) in $f\left(q|y\right)$ and $\overline{f}\left(q\right)$. See also Csiszar (1974) in the information theory literature. □

Proof of lemma 10. Forming the Lagrangian with multipliers μ and $\eta\left(q\right)$ on the constraints specified in equations (2.71) and (2.72),

$$\mathcal{L}(\overline{f}) = \int G\left(y\right)\log\left[\sum_{q\in\mathcal{Q}}\overline{f}\left(q\right)\exp\left\{\frac{1}{\theta^{p}}\pi\left(q-y\right)\right\}\right]dy + \mu\sum_{q\in\mathcal{Q}}\overline{f}\left(q\right) + \sum_{q}\eta\left(q\right)\overline{f}\left(q\right). \tag{B.25}$$

Differentiating with respect to $\overline{f}\left(q\right)$ yields

$$\int G\left(y\right)\frac{\exp\left\{\frac{1}{\theta^{p}}\pi\left(q-y\right)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}}\overline{f}\left(\widehat{q}\right)\exp\left\{\frac{1}{\theta^{p}}\pi\left(\widehat{q}-y\right)\right\}}dy + \mu + \eta\left(q\right) = 0. \tag{B.26}$$

For $\overline{f}(q) > 0$, such that $\eta(q) = 0$, multiplying by $\overline{f}(q)$, and then summing over $q \in \mathcal{Q}$ yields

$$\int G(y) \frac{\overline{f}(q) \exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}} dy + \mu\overline{f}(q) = 0, \tag{B.27}$$

$$\int G(y)\, dy + \mu = 0, \tag{B.28}$$

yielding the Lagrange multiplier, $\mu = -1$.

Hence,

$$\int G(y) \frac{\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\sum_{\widehat{q}\in\mathcal{Q}} \overline{f}(\widehat{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widehat{q}-y)\right\}} dy \leq 1, \tag{B.29}$$

with equality for each q such that $\overline{f}(q) > 0$. □

Proof of lemma 11. The following proof follows the method of proof indicated by Fix (1978).

Let $Z(q)$ be a complex analytic function on the entire complex plane.

1. If $Z(q)$ is equal to some constant M for all $q \in \mathbb{C}$, then the roots of the function $Z(q) - M$ are the entire complex plane.

2. If $Z(q)$ is non-constant, then the roots of the function $Z(q) - M$, for any constant M, are a set of isolated points.

Hence, the real roots of $Z(q) - M$ are either the entire real line or a set of isolated points.

In order to apply this result, one needs to establish that $Z\left(q; \overline{f}\right)$, defined in equation (2.73) for $q \in \mathbb{R}$, and for fixed $\overline{f}$, $\mathcal{Q}$, and $G(y)$, can be extended to a complex function on the entire complex plane.

The function $\pi(q-y)$, which is specified in equation (2.83), is a sum of two exponentials. Hence, the function $Z\left(q; \overline{f}\right)$ is a composition of exponentials. The exponential is analytic on $\mathbb{R}$ (and $\mathbb{C}$). Hence, $Z\left(q;\overline{f}\right)$ is analytic. Any real analytic function on some open set on the real line can be extended to a complex analytic function on some open set of the complex plane. In this particular case, $Z\left(q;\overline{f}\right)$, for $q \in \mathbb{C}$, is complex analytic on the entire complex plane. It follows from above that the real roots of $Z\left(q;\overline{f}\right) - 1$ are either the entire real line or a set of isolated points.

Since lemma 10 establishes that prices in the support $\mathcal{Q}$ are roots of the function $Z\left(q;\overline{f}\right) - 1$, it then follows that the support of the distribution of prices is either the entire real line or a discrete set of prices. □

Proof of lemma 12. Differentiating $\mathcal{F}$ with respect to q yields

$$\int \frac{\partial \pi(q-y)}{\partial q} f(q|y) G\left(y\right) dy = 0. \tag{B.30}$$

The optimal support is the subset of the points q satisfying condition (B.30) that also satisfy the second order condition, which, differentiating (B.30), is solved to yield

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

Using Bayes rule, the two conditions are rearranged as in equations (2.75) and (2.76). □

B.2 Computational Method

The algorithm for finding the optimal policy iterates between finding the optimal review policy for a given pricing policy, and finding the optimal pricing policy for a given review policy.

The review policy requires from the pricing policy the expected per-period profit net of the cost of the price signal, $\Pi(y)$, defined in equation (2.49), while the pricing policy requires from the review policy the distribution over all possible states, $G(y)$, defined in equation (2.67), as a function of the hazard for reviews, $\Lambda(\widetilde{y})$.

I begin by determining the optimal review policy when the optimal pricing policy involves a single price, $\overline{q}$, such that $\Pi(y) = \pi(\overline{q} - y)$ for all y.

I then find the optimal pricing policy for different values of the unit cost of the price signal, θ^p, fixing the distribution of the possible post-review states, $G(y)$, at the distribution implied by this review policy.

Algorithm for the Optimal Review Policy for a given Pricing Policy

Solving for the firm's optimal review policy, $\overline{\Lambda}$, Λ, requires the discount factor, β, the distributions of the shocks, $h_{\widetilde{\nu}}$ and h_{ν}, the function Π, implied by the pricing policy, the cost of holding a policy review, κ, and the unit cost of the review signal, θ^r.

The steps that solve for the optimal review policy are as follows:

1. Initialize the frequency of reviews, $\overline{\Lambda}_{(0)}$.

2. Iterate

   (a) Initialize the hazard function, $\Lambda_{(0)}$.

   (b) Given $\Lambda_{(k)}$, iterate on the fixed point equation (2.53) to solve for the value function $V_{(k)}$.

   (c) Given $V_{(k)}$, solve equation (2.52) for the hazard function, $\Lambda_{(k+1)}$.

(d) Given $\Lambda_{(k+1)}$, solve for the frequency of reviews, $\overline{\Lambda}_{(k+1)}$, using equation (2.57).

(e) If $\left|\overline{\Lambda}_{(k+1)} - \overline{\Lambda}_{(k)}\right|$ exceeds a prescribed tolerance level, return to step 2.b. Otherwise, continue to step 3.

3. Compute the implied distribution of post-review states, $G(y)$, using equation (2.59).

The method is the based on that used by Woodford (2009), hence I omit further details.

Algorithm for the Optimal Pricing Policy for a given Review Policy

Solving for the firm's optimal pricing policy, $\mathcal{Q}$, $\overline{f}$, f, requires the firm's period objective function, $\pi(q-y)$, the distribution over all possible states, $G(y)$, and the cost of the price signal, θ^p.

The steps that solve for the optimal pricing policy for an arbitrary information cost are as follows:

1. Discretize the state, $y \sim G(y)$.

2. Solve for the full information solution.

3. Solve for the single-price policy.

4. Determine the boundaries of the support $\mathcal{Q}$.

5. Initialize the cardinality of the support $\mathcal{Q}$.

6. Iterate, for a given cardinality:

   (a) Initialize the support, $Q_{(0)}$.

   (b) Given $Q_{(k)}$, determine the optimal distributions $\overline{f}_{(k)}$ and $f_{(k)}$.

(c) Given $f_{(k)}$, determine the optimal support $Q_{(k+1)}$.

(d) If $\left\|Q_{(k+1)} - Q_{(k)}\right\|$ or $\left\|\overline{f}_{(k+1)} - \overline{f}_{(k)}\right\|$ exceed prescribed tolerance levels, return to STEP 6.b. Otherwise, continue to STEP 7.

7. Check the cardinality of the solution:

(a) If $\overline{f}(q) < tol_f$ or $1 - Z\left(q; \overline{f}\right) > tol_Z$ for any $q \in \mathcal{Q}$, remove a point from the support and return to STEP 6.b.

(b) Else, if $Z\left(q; \overline{f}\right) - 1 > tol_Z$ for some $q \notin \mathcal{Q}$, add a point to the support and return to STEP 6.b.

(c) Else, if $\left|Z\left(q; \overline{f}\right) - 1\right| \leq tol_Z$ for all $q \in \mathcal{Q}$ and $Z\left(q; \overline{f}\right) - 1 \leq tol_Z$ for any $q \notin \mathcal{Q}$, END.

Details

STEP 1: Discretize the state.

Let the discretization of $y \sim G(y)$ be denoted by the n_y nodes $\{y_i\}$ and weights $\{G_i\}$, $i = 1, ..., n_y$.

STEP 2: Solve for the full information solution.

The full information solution is given by the set of prices $\mathcal{Q}^{FI} = \left\{q_i^{FI}\right\}$, $i = 1, .., n_y$,

$$q_i^{FI} = \arg\max_q \pi\left(q - y_i\right). \tag{B.32}$$

STEP 3: Solve for the single-price policy.

The single-price policy, if optimal, is given by the price

$$\overline{q} = \arg\max_q \sum_{i=1}^{n_y} G_i \pi\left(q - y_i\right). \tag{B.33}$$

The threshold information cost, $\overline{\theta}^p$, below which the cost of information is sufficiently low to require a multiple-price policy, solves

$$\sum_{i=1}^{n_y} G_i \left[\frac{\partial^2}{\partial q^2} \pi \left(q - y_i \right) + \frac{1}{\theta^p} \left(\frac{\partial}{\partial q} \pi \left(q - y_i \right) \right)^2 \right] = 0, \tag{B.34}$$

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

STEP 4: Determine the boundaries of the support $\mathcal{Q}$.

(a) Initialize the boundaries of $\mathcal{Q}$ at

$$a_{(0)} = \min \left\{ q^{FI} \right\}, \tag{B.35}$$

$$b_{(0)} = \max \left\{ q^{FI} \right\}. \tag{B.36}$$

(b) Iterate:

i. Construct a grid $Q_{(j)}$ of n_y equidistant points on the interval $[a_{(j)}, b_{(j)}]$, where n_y is the size of the set $\mathcal{Q}^{FI}$. Let $w_{(j)}$ denote the distance between points in $Q_{(j)}$.

ii. Given $Q_{(j)}$, find the optimal distributions $\overline{f}_{(j)}$ and $f_{(j)}$ using the Blahut-Arimoto algorithm (detailed in a later section of this appendix).

iii. Given $\overline{f}_{(j)}$, compute the function $Z\left(q; \overline{f}\right)$ for each $q \in Q_{(j)}$,

$$Z\left(q; \overline{f}_{(j)}\right) = \sum_{i=1}^{n_y} \frac{G_i \exp\left\{ \frac{1}{\theta^p} \pi(q - y_i) \right\}}{\sum_{l=1}^{n_y} \overline{f}_{(j)} \left(q_l \right) \exp\left\{ \frac{1}{\theta^p} \pi(q_l - y_i) \right\}}; \tag{B.37}$$

- find the *first* point in $Q_{(j)}$ for which $Z\left(q; \overline{f}_{(j)}\right) - 1 \geq tol_Z$; denote this point by q_{first}; set the new lower bound

$$a_{(j+1)} = q_{first} - w_{(j)}; \tag{B.38}$$

- find the *last* point in $Q_{(j)}$ for which $Z\left(q;\overline{f}_{(j)}\right)-1\geq tol_Z$; denote this point by q_{last}; set the new upper bound

$$b_{(j+1)}=q_{last}+w_{(j)};\tag{B.39}$$

iv. If $a_{(j+1)}=a_{(j)}$ and $b_{(j+1)}=b_{(j)}$, end this STEP; otherwise, return to i.

STEP 5. Initialize the cardinality of the support $\mathcal{Q}$.

Set the initial cardinality of $\mathcal{Q}$, $n=n_y$.

STEP 6. Iterate for a given cardinality.

(a) Initialize the support, $Q_{(0)}=Q_J$, where Q_J is the last grid in STEP 4.

(b) Given $Q_{(k)}$, determine the optimal distributions $\overline{f}_{(k)}$ and $f_{(k)}$ using the Blahut-Arimoto algorithm (detailed in a later section of this appendix).

(c) Given $f_{(k)}$, the optimal support, $Q_{(k+1)}$ is given by all points q that satisfy

$$\sum_{i=1}^{n_y} G_i f_{(k)}\left(q|y_i\right)\frac{\partial}{\partial q}\pi\left(q-y_i\right)=0.\tag{B.40}$$

(d) If $\left\|Q_{(k+1)}-Q_{(k)}\right\|$ or $\left\|\overline{f}_{(k+1)}-\overline{f}_{(k)}\right\|$ exceed prescribed tolerance levels, return to (b). Otherwise, continue to STEP 7.

At the end of STEP 6, we have a solution that satisfies the necessary optimality conditions for $f\left(q|y\right)$, $\overline{f}\left(q\right)$ and q, for a given cardinality n of the support. Let this solution be denoted by f_n, $\overline{f}_n$, $\mathcal{Q}_n$, where the subscript indicates the current cardinality.

STEP 7. Check the cardinality of the solution.

(a) Compute $Z\left(q;\overline{f}_n\right)$ for each $q \in \mathcal{Q}_n$. Let this vector be denoted by Z_n^{in}. If $1 - Z_n^{in} \geq tol_Z$, set $n = n-1$, remove from the support the point $q = \arg\min_q Z_n^{in}$, and return to STEP 6.b. Else, continue to STEP 7.b.

(b) Construct a grid, Q_n^{out}, and compute $Z\left(q;\overline{f}_n\right)$ for each $q \in Q_n^{out}$. Let this vector be denoted by Z_n^{out}. If $Z_n^{out} - 1 \geq tol_Z$, set $n = n+1$, add to the support the point $q = \arg\max_q Z_n^{out}$, and return to STEP 6.b. Else, END

The grid Q_n^{out} contains $M+1$ densely-spaced grid points over the entire range of $\mathcal{Q}^{FI}$, where M is a very large number. The density of this grid is chosen such that for a point $q \in \mathcal{Q}_n$ and either of its nearest neighbors, $q_{next} \in \mathcal{Q}^{DENSE}$, $Z_n\left(q;\overline{f}\right) - Z_n\left(q_{next};\overline{f}\right) \approx tol_{ZDIFF}$.

Discussion

STEP 4: Determining the boundaries of the support.

Since the optimal signal is a compression of the full information solution, the support of the price signal is weakly contained in the support of the full information solution, $\mathcal{Q}^{FI}$. Hence, the initialization of the support in STEP 4.1 ensures that we find a globally optimal solution, given the discretization of the state, y. Since the boundaries of the solution are constrained by the discretization of the state, the resulting solution needs to be checked for sensitivity to n_y and to the minimum and maximum values of y_i.

STEP 4 improves the efficiency of the algorithm in later stages. Once it has converged, it ensures that we search for the solution on the narrowest possible interval of size n_y. Maintaining the cardinality fixed at n_y as we shrink the boundaries implies that each iteration increases the density of the grid. As a result, STEP 4.2 also yields good approximations to the first and last points of support in the solution.

Figure B.1 illustrates how this step shrinks the boundaries of the support by plotting the function $Z\left(q;\overline{f}\right)$ for the first iteration in the top panel, where the support of the price distribution is initialized inside the interval bounded by $a_{(0)} = -1.1$ and $b_{(0)} = 0.4$, and for the last iteration in the bottom panel, where the possible range of q has been reduced to $a_{(J)} = -0.001$ and $b_{(J)} = 0.178$.

STEP 7. Checking a given cardinality.

The challenging part of this step is verifying that $Z\left(q;\overline{f}\right) \leq 1$ for any $q \notin \mathcal{Q}$. In principle, since $Z\left(q;\overline{f}\right)$ can be easily evaluated for any q, one could construct an arbitrarily wide and dense grid Q_n^{out} over which to evaluate the function. In practice, I find that it is sufficient to evaluate the function over a union of a wide sparse grid and a narrow, dense grid. For the sparse grid, I choose $\mathcal{Q}^{FI}$, which serves to confirm that although in STEP 4 the boundaries are determined using an approximation to the final distributions, f and $\overline{f}$, that step did not eliminate optimal points. The density of the narrow grid is constrained by the numerical error in the computation of f and $\overline{f}$.

It is convenient to err on the side of giving the algorithm too high of a cardinality. If a candidate cardinality is too high, then procedure that computes the optimal distribution in STEP 6 adjusts by returning zero mass at the excess points. This does not adversely the computation of the mass at any of the other points, nor does it affect any of the other steps in the algorithm. Doing so ensuring that when STEP 7 is reached, the cardinality is mostly adjusted down, based on the computation of Z_n^{in} at the current points in the support.

The initialization of the support in STEP 6.a is the most conservative approach, which ensures that we rarely rely on Z_n^{out} to adjust the cardinality of the support. An alternative approach, which relies more on Z_n^{out} but also substantially reduces the running time of the algorithm, is to initialize the support at the points for which $Z\left(q;\overline{f}\right) \geq 1$ at the end of STEP 4.

Finally, if the algorithm repeatedly iterates between cardinalities n and $n+1$, then we are at a point where the solution cannot be accurately estimated. In this case, the tolerance levels need to be reduced to obtain convergence.

The Blahut-Arimoto Algorithm

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

$$\overline{f}(q) = \overline{f}(q) \int \frac{\exp\left\{\frac{1}{\theta^p}\pi(q-y)\right\}}{\int \overline{f}(\widetilde{q}) \exp\left\{\frac{1}{\theta^p}\pi(\widetilde{q}-y)\right\} d\widetilde{q}} G(y)\, dy, \tag{B.41}$$

which is obtained by integrating equation (2.48) over y. For a given $\overline{f}(q)$, the conditional distribution is then given by equation (2.48). For a proof of convergence, see for example Csiszar (1974).

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

1. If not preset elsewhere, set $f_j^{(0)} = 1/n$, $j = 1, .., n$.

2. Compute the $n_y \times n$ matrix d whose $(ij)^{th}$ entry is given by

$$d_{ij} = \exp\left\{\frac{1}{\theta^p}\left[\pi(q_j - y_i)\right]\right\}. \tag{B.42}$$

3. Compute

$$D_i = \sum_{k=1}^{n} f_j^{(k)} d_{ij}, \quad i = 1, .., n_y; \tag{B.43}$$

$$Z_j = \sum_{i=1}^{n_y} G_i \frac{d_{ij}}{D_i}, \quad j = 1, .., n; \tag{B.44}$$

$$f_j^{(k+1)} = f_j^{(k)} Z_j, \quad j = 1, .., n. \tag{B.45}$$

4. Compute

$$TU = -\sum_{j=1}^{n} f_j^{(k+1)} \ln Z_j; \tag{B.46}$$

$$TL = -\max_j \ln Z_j. \tag{B.47}$$

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

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

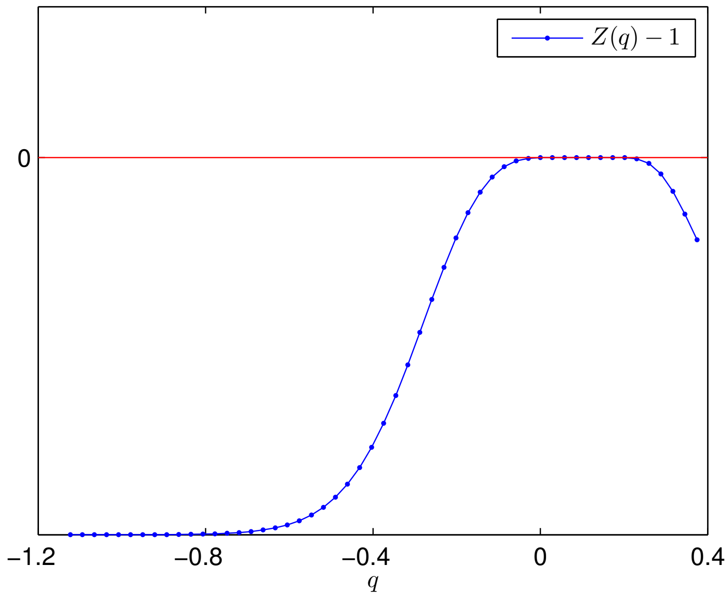
$$f_{jk} = f_k \frac{d_{jk}}{D_j}; \tag{B.48}$$

$$f_k = \sum_{j=1}^{n_y} f_{jk} G_j; \tag{B.49}$$

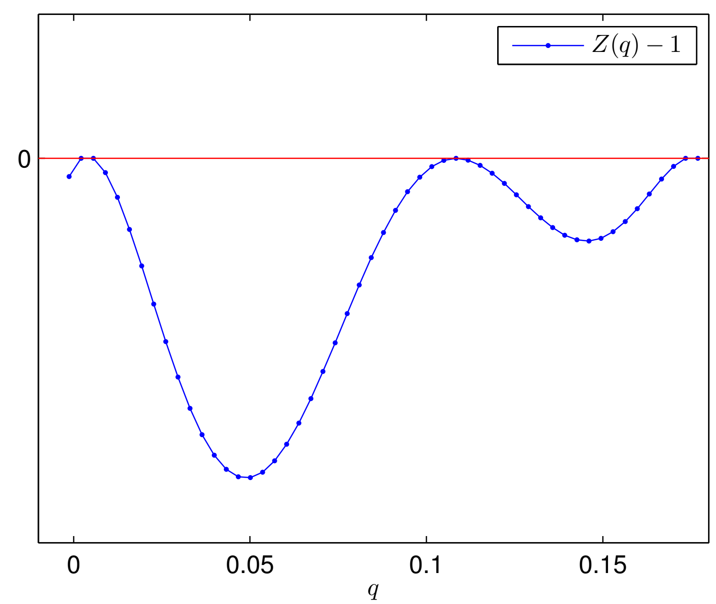
$$\Pi = \sum_{j=1}^{n_y} \sum_{k=1}^{n} \pi(q_k - y_j) f_{jk} G_j; \tag{B.50}$$

$$I = \frac{1}{\theta^p} \Pi - \sum_{j=1}^{n_y} G_j \log D_j. \tag{B.51}$$

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



(a) initial support at the beginning of STEP 4



(b) final support at the end of STEP 4

Figure B.1: Determining the boundaries of the optimal support in STEP 4 of the algorithm, as a function of $Z\left(q;\overline{f}\right) - 1$.

B.3 Model of Price Setting

I explore the implications for price adjustment of the information structure developed thus far in a standard model of price-setting under monopolistic competition. 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 Agents

The economy consists of three types of agents: an infinitely-lived representative household, a continuum of infinitely-lived 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 to all firms i in the economy. It chooses sequences of consumption, hours, and bond holdings to maximize a discounted utility stream defined by:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{1-\sigma} C_t^{1-\sigma} - \frac{1}{1+\nu} \int_0^1 H_t(i)^{1+\nu} di \right], \tag{B.52}$$

where $\beta \in (0,1)$ is the discount factor, C_t is the consumption basket, $\sigma > 1$ is the constant relative risk aversion parameter, $H_t(i)$ is the total amount of labor (in hours) supplied by the representative household to sector i, and $\nu \geq 0$ is the inverse of the Frisch elasticity of labor supply.

Maximization of the objective defined in equation (B.52) is subject to a standard budget constraint in each period t,

$$\int_0^1 W_t(i) H_t(i)\, di + \int_0^1 \Pi_t(i)\, di + B_t + M_{t-1} + T_t \geq P_t C_t + E_t\left[R_{t,t+1} B_{t+1}\right] + M_t, \quad \text{(B.53)}$$

where $W_t(i)$ is the nominal hourly wage of sector i, $\Pi_t(i)$ is the dividend received from sector i, B_t is the portfolio of nominal bond holdings in the period, M_{t-1} is the household's money balance entering period t, T_t is the net monetary transfer received from the government, P_t is the aggregate price index for the consumption basket C_t, and $R_{t,t+1}$ is the stochastic discount factor used to discount income streams between time t and time $t+1$.

The representative household also faces a no-Ponzi-scheme condition, and a cash-in-advance constraint on consumption purchases,

$$P_t C_t \leq M_{t-1} + T_t. \quad \text{(B.54)}$$

Finally, the consumption basket, C_t, is given by a Dixit-Stiglitz aggregator over a continuum of differentiated products $i \in [0,1]$, with elasticity of substitution $\varepsilon > 1$ and good-specific preference shocks, $A_t(i)$, whose law of motion is specified in the next subsection:

$$C_t \equiv \left[\int_0^1 \left[A_t(i) C_t(i)\right]^{\frac{\varepsilon-1}{\varepsilon}} di\right]^{\frac{\varepsilon}{\varepsilon-1}}. \quad \text{(B.55)}$$

Inter-temporal consumer optimization yields the following standard first order conditions for the optimal supply of labor and for the stochastic discount factor:

$$\frac{W_t(i)}{P_t} = \frac{H_t(i)^{\nu}}{C_t^{-\sigma}} \qquad \text{and} \qquad R_{t,T} = \beta^{T-t}\left(\frac{C_t}{C_T}\right)^{\sigma} \frac{P_t}{P_T}. \quad \text{(B.56)}$$

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

i,

$$C_t(i) = A_t(i)^{\varepsilon-1}\left(\frac{P_t(i)}{P_t}\right)^{-\varepsilon} C_t, \tag{B.57}$$

where P_t is the aggregate price index defined by

$$P_t \equiv \left[\int_0^1 \left(\frac{P_t(i)}{A_t(i)}\right)^{1-\varepsilon} di\right]^{\frac{1}{1-\varepsilon}}. \tag{B.58}$$

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

$$Y_t(i) = H_t(i)^{\frac{1}{\gamma}}/Z_t(i), \tag{B.59}$$

where $\gamma \geq 1$ captures decreasing returns to scale in production, $H_t(i)$ is the differentiated labor input, and, for later convenience, $Z_t(i)$ denotes the *inverse* of firm-specific productivity. The evolution of $Z_t(i)$ is described in the next subsection.

The firm's nominal profit each period, excluding the resources spent to acquire information about market conditions, is

$$\Pi_t(i) = P_t(i)Y_t(i) - W_t(i)H_t(i). \tag{B.60}$$

In every period, the firm sets its price and commits to fulfill all demand at that price.

In the absence of information costs, the firm would seek to maximize its discounted profit stream,

$$E_0 \sum_{t=0}^{\infty} R_{0,t}\Pi_t(i). \tag{B.61}$$

Government For simplicity, the government pursues an exogenous policy. The net monetary transfer in each period is equal to the change in money supply, which is assumed to evolve exogenously, as described in the next subsection

$$T_t = M_t - M_{t-1}. \tag{B.62}$$

The Shocks

The economy is subject to three kinds of shocks: (1) μ_t, permanent monetary shocks, which are the only source of aggregate disturbances in the economy, are generally small, and are summarized in the exogenous evolution of money supply; (2) $\xi_t(i)$, permanent idiosyncratic quality shocks, which affect both the demand for an individual product and the cost of producing it; and (3) $\zeta_t(i)$, i.i.d. idiosyncratic productivity shocks.

The log of money supply[26] is assumed to follow a random walk process,

$$m_t = m_{t-1} + \mu_t, \tag{B.63}$$

$$\mu_t \overset{i.i.d.}{\sim} h_\mu, \tag{B.64}$$

where μ_t is independent over time and from any other disturbances in the economy.

The inverse of firm-specific productivity contains independently distributed permanent and transitory components, where the permanent component is the same as the household preference shock. In logs, this term evolves according to

$$z_t(i) = a_t(i) + \zeta_t(i), \tag{B.65}$$

$$a_t(i) = a_{t-1}(i) + \xi_t(i), \tag{B.66}$$

$$\xi_t(i) \overset{i.i.d.}{\sim} h_\xi, \tag{B.67}$$

$$\zeta_t(i) \overset{i.i.d.}{\sim} h_\zeta. \tag{B.68}$$

The permanent shock $a_t(i)$ is a quality shock that increases both the utility from

[26] I use lower-case letters to denote logs of different variables introduced in subsection B.3.

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 permanent nominal shock m_t and by the permanent idiosyncratic shock, $a_t(i)$. This assumption enables a reduction in the state space of the problem, thus increasing tractability. The same assumption is made by Midrigan (2010) and Woodford (2009). The permanent quality shock will generate large and persistent movements in both individual prices and relative prices over time, consistent with the data.

The shock $\zeta_t(i)$ is a purely transitory productivity shock that helps to generate large price changes, as observed in the data.

Partial Equilibrium

In the flexible-price equilibrium with *no* information costs and no other costs to nominal price adjustment, the firm chooses its price in each period to maximize its per-period profit in units of marginal utility. The full-information optimal log-price,[27] denoted by $x_t(i)$, is a linear combination of all the shocks in the economy:

$$x_t(i) = m_t + a_t(i) + \phi\zeta_t(i), \quad \phi \equiv \frac{\eta}{\varepsilon\eta - \varepsilon + 1} < 1. \tag{B.69}$$

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. When substituting the full-information equilibrium outcomes, the profit of an information-constrained firm is proportional to $\pi(p_t(i) - x_t(i))$[28], where $p_t(i)$ is the log-price charged by the information-constrained firm[29], $x_t(i)$ is the optimal full-information log-price de-

[27] The optimal log-price is rescaled by a constant that is omitted.

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

[29] The log-price charged by the rationally inattentive firm and the optimal log-price are rescaled by the same (omitted) constant.

termined in equation (B.69), and

$$\pi(p-x) = e^{(1-\varepsilon)(p-x)} - \frac{\varepsilon - 1}{\varepsilon\eta} e^{-\varepsilon\eta(p-x)}. \tag{B.70}$$

Equation (B.70) defines the profit function introduced in section 2.2. Note that the profit function defined in equation (B.70) 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 shocks are mapped into the notation used in section 2.2 by defining

$$\widetilde{\upsilon}_t(i) \equiv \mu_t + \xi_t(i), \tag{B.71}$$

$$\upsilon_t(i) \equiv \phi\zeta_t(i). \tag{B.72}$$

To map the current model into the notation of section 2.4, which employs the normalized variables q, y, and $\widetilde{y}$, and the profit function $\pi(q-y)$, the full information price in each period t can then be written as a function of the permanent state at the time of the last review, τ periods ago, and the accumulated shocks since then,

$$x_t(i) = m_{t-\tau} + a_{t-\tau}(i) + y_\tau(i), \tag{B.73}$$

$$y_\tau(i) \equiv \widetilde{y}_\tau(i) + \phi\zeta(i), \tag{B.74}$$

$$\widetilde{y}_\tau(i) \equiv \sum_{j=1}^{\tau}\mu_j + \sum_{j=1}^{\tau}\xi_j(i), \tag{B.75}$$

with $\widetilde{y}_0(i) = 0$.

Conditional on no review, the information-constrained price is

$$p_t(i) = m_{t-\tau} + a_{t-\tau}(i) + q_\tau(i). \tag{B.76}$$

The per-period profit $\pi(p_t(i) - x_t(i))$ is replaced by $\pi(q_\tau(i) - y_\tau(i))$, a function of the normalized price and the normalized state, both of which are indexed by τ, the number of periods since the last policy review, with $\pi(q - y)$ defined by equation (B.70).

C Addendum to Chapter 3

C.1 Details of Retailer Search

We present the problem of retailer search in the two-country model. Each region is of unit mass. As shown in the main text, retailers in region a of the *Home* country sample prices from the following distribution:

$$f_a^{ret}(\widehat{p}) \equiv \alpha_1 f_a(\widehat{p}) + \alpha_2 f_b(\widehat{p}) + \alpha_3 f_c(\widehat{p}) + \alpha_4 f_d(\widehat{p}),$$

with $\sum_{i=1}^{4} \alpha_i = 1$. They have a regional bias captured by $\alpha_1 > \alpha_i$, $i \in \{2, 3, 4\}$. Since the prices posted by producers in each country are identically distributed,

$$f_a^{ret}(\widehat{p}) \equiv \alpha f_a(\widehat{p}) + (1 - \alpha) f_c(\widehat{p}), \tag{C.1}$$

where $\alpha \equiv \alpha_1 + \alpha_2$. The regional bias translates into an apparent national bias, with $\alpha > 1 - \alpha$, since we assume for simplicity that $\alpha_2 = \alpha_3 = \alpha_4$.

Retailers in region b of the *Home* country sample prices from the following distribution:

$$f_b^{ret}(\widehat{p}) \equiv \beta_1 f_a(\widehat{p}) + \beta_2 f_b(\widehat{p}) + \beta_3 f_c(\widehat{p}) + \beta_4 f_d(\widehat{p}),$$

with $\sum_{i=1}^{4} \beta_i = 1$. They have also a regional bias, $\beta_2 > \beta_i$, $i \in \{1, 3, 4\}$. As in the case of region-a retailers, using the within-country symmetry of producers, we obtain

$$f_b^{ret}(\widehat{p}) \equiv \beta f_a(\widehat{p}) + (1 - \beta) f_c(\widehat{p}),$$

where $\beta \equiv \beta_1 + \beta_2$. Although in principle the bias of retailers in region b may differ from that of retailers in region a, since the two regions are otherwise identical, we impose symmetry in segmentation as well, and assume that $\beta_2 = \alpha_1$, and $\beta_1 = \beta_3 = \beta_4 = \alpha_2$.

Hence, retailers in region b sample prices from the same distribution, $f_b^{ret}(\widehat{p}) = f_a^{ret}(\widehat{p})$.

Retailers in region c of the *Foreign* country sample prices from the following distribution

$$f_c^{ret}(\widehat{p}) \equiv \gamma_1 f_a(\widehat{p}) + \gamma_2 f_b(\widehat{p}) + \gamma_3 f_c(\widehat{p}) + \gamma_4 f_d(\widehat{p}),$$

with $\sum_{i=1}^{4} \gamma_i = 1$. They have a regional bias captured by $\gamma_3 > \gamma_i$, $i \in \{1, 2, 4\}$. In turn, this translates into an apparent national bias since $\gamma_1 = \gamma_2 = \gamma_4$. Using the within-country symmetry of producers, we obtain

$$f_c^{ret}(\widehat{p}) \equiv (1 - \gamma)\, f_a(\widehat{p}) + \gamma f_c(\widehat{p}), \tag{C.2}$$

where $\gamma \equiv \gamma_3 + \gamma_4$.

Retailers in region d of the *Foreign* country sample prices from

$$f_d^{ret}(\widehat{p}) \equiv \delta_1 f_a(\widehat{p}) + \delta_2 f_b(\widehat{p}) + \delta_3 f_c(\widehat{p}) + \delta_4 f_d(\widehat{p}),$$

with $\sum_{i=1}^{4} \delta_i = 1$. They have a regional bias captured by $\delta_4 > \delta_i, i \in \{1, 2, 3\}$. As in the case of the *Home* country, we assume symmetry between regions c and d, such that $\delta_4 = \gamma_3$ and $f_d^{ret}(\widehat{p}) = f_c^{ret}(\widehat{p})$.

Hence, all retailers in the *Home* country sample prices from the distribution given by equation (C.1), and all retailers in the *Foreign* country sample prices from the distribution given by equation (C.2). Regional bias may be another source of heterogeneity across the two countries, hence we allow for the possibility that $\alpha_1 \neq \gamma_3$, which in turn implies that $\alpha \neq \gamma$.