Pricing Regimes in Disaggregated Data[1]

Luminita Stevens

Ph.D. Candidate, Columbia University

July 2011

ABSTRACT: This paper develops a robust statistical method that identifies patterns in product-level price data that are at odds with current models of price setting and suggest a new theoretical approach based on infrequently updated strategies that consist of a small set of prices charged over the life of the strategy. I develop a non-parametric statistical method - the *break test* - that tests for the existence of breaks in product-level pricing series based on changes in the distribution of prices over time. The method builds on the Kolmogorov-Smirnov statistic, and it is adapted to test for the existence of multiple potential breaks. The test does not require specific a priori knowledge about patterns in the data, and it is shown to be robust to different data generating processes. On average, it outperforms popular price filters in simulations. Using Dominick's grocery store product-level data, I characterize the regimes identified by the break test in terms of their duration and the shape of the distribution of prices realized within each regime. I find that 1) the typical pricing regime lasts seven months, with most regimes lasting between four and 16 months; 2) the typical regime consists of a small number of distinct prices: for the large majority of regimes, five or fewer unique prices account for more than 90% of the regime; 3) 99% of the regimes identified across all product series have a rigid modal price but 76% of product series exhibit additional within-regime rigidity, with multiple prices being revisited over the life of the regime; 4) much of the volatility of prices in terms of both frequency and size of price changes is concentrated in these rigid multi-price series; 5) the frequency with which prices are charged within regimes is steeply declining, from 71% for the modal price to 6% for the third most common price.

[1]INCOMPLETE. I would like to thank Mike Woodford and Jaromir Nosal for valuable suggestions. I am also indebted to Ryan Chahrour, Daniel Dias, Christian Hellwig, Marcelo Moreira, Emi Nakamura, Alexei Onatski, Ernesto Pasten, Ricardo Reis, Jon Steinsson, seminar participants at Columbia University and Toulouse School of Economics, and participants at the 2011 Midwest Macroeconomics Meeting at Vanderbilt University and the 2011 SED Annual Meeting in Ghent for helpful comments. Contact: lld2108@columbia.edu.

1 Introduction

Consumer prices exhibit high volatility even under relatively stable macroeconomic conditions. For instance, Nakamura and Steinsson (2008) report that from 1988 until 2005 - a period of low and stable inflation in the US - the median frequency of price changes in the micro-level data underlying the US CPI was approximately 20% per month, with a median price change of approximately 11% in absolute size. Moreover, this volatility is broad-based: nine of the eleven major product groups in the US CPI have a frequency of price changes that exceeds 15%. It may be the case that this volatility reflects the prevalence of large and frequent idiosyncratic shocks that wash out in the aggregate. However, this view poses a challenge for micro-founded models that seek to explain the dynamics of aggregate inflation and monetary non-neutrality: in models in which all price changes are responses to concurrent market conditions, such high frequency and magnitude of price changes imply essentially immediate and complete adjustment to market conditions, and hence a low degree of non-neutrality. Conversely, if at least some of the observed price changes are only partially tied to concurrent market conditions, then high micro-level volatility can coexist with significant non-neutrality. Hence identifying the motivation behind the observed price changes is a key part of the monetary research agenda.

Recent studies of product-level pricing series[2] have underscored the importance of transitory prices, arguing that a significant portion of the divergence between micro volatility and aggregate sluggishness is driven by large yet transitory price changes occurring to and from a frequently recurring price. Figure 1 below shows the weekly price series for frozen juice from grocery store data, illustrating the importance of transitory price changes: prices change very frequently and by large amounts, yet seem to alternate among a small number of distinct values. More broadly, Nakamura and Steinsson (2008) estimate that a pattern of large price cuts from an apparently rigid high price characterizes products in categories that account for approximately 40% of non-shelter consumer expenditures. The presence of these short-lived price cuts is important quantitatively: eliminating them reduces the overall median frequency of price changes by half. Moreover, they are also fairly widespread across product categories, substantially impacting the frequency

[2] For a comprehensive survey of the empirical literature on good-level pricing, see Klenow and Malin (2010).

of adjustment in six of the eleven major product groups of the CPI. The extent to which these transitory price changes are important for aggregate dynamics, how they should be analyzed empirically, and how they should be modeled theoretically remain open questions in the literature. For example, in models that constrain the firm to return to return to a recurring price level following a short-lived price change, half of the price changes associated with these transient prices are mechanical, and not the result of re-optimization. Non-neutrality in such models is necessarily higher than that implied by models with the same frequency of price changes but in which the firm *chooses* to return to the exact same price level. How much higher depends on the specification of the model, which again depends on the motivation behind implementing temporary versus permanent price changes in the first place.

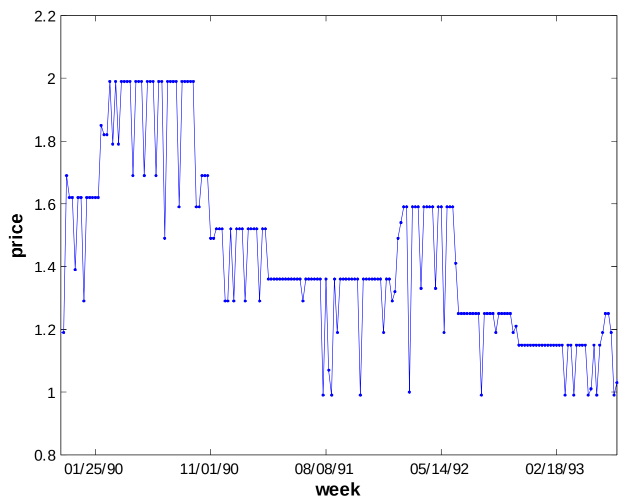

Figure 1. Frozen juice price series from Dominick's data set.

An ongoing challenge for empirical work on product-level price-setting is the identification of transitory versus permanent price changes. The definition of a transitory price change and how it is identified in the data are crucial for informing models of pricing: important statistics such as the frequency of regular versus transitory price changes vary significantly depending on assumptions and implementation. As an alternative, I propose a "break test" - a method that sidesteps the need for a priori definitions of transitory or persistent price changes 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. In addition to its advantage as a purely statistical approach, described more below, this approach is also consistent with a potentially appealing theoretical interpretation of product-level volatility, which views all prices within a regime as part of an integrated pricing strategy chosen by the firm. A theoretical model of multi-price strategies can then investigate the determinants of these strategies and analyze how the strategies changes over time.

The *break test* is based on the Kolmogorov-Smirnov test, which determines if two samples are likely to have been drawn from the same distribution. An advantage of this approach is its generality, and hence robustness across different types of data structures. It embeds the familiar filtering methods, while allowing for more complex price-setting patterns to be uncovered. Simulation exercises show that the test correctly rejects the null of no change in the distribution 91% of the time across different data generating processes. Conversely, the method yields false positives less than 2% of the time. Upon rejection of the null, the test finds the true location of the break exactly 94% of the time; in the remaining cases, it is off by two periods, on average. Figure 2 shows the resulting regimes for the frozen juice price series plotted above: the test successfully reveals the underlying structure in the series by identifying breaks at the expected locations.

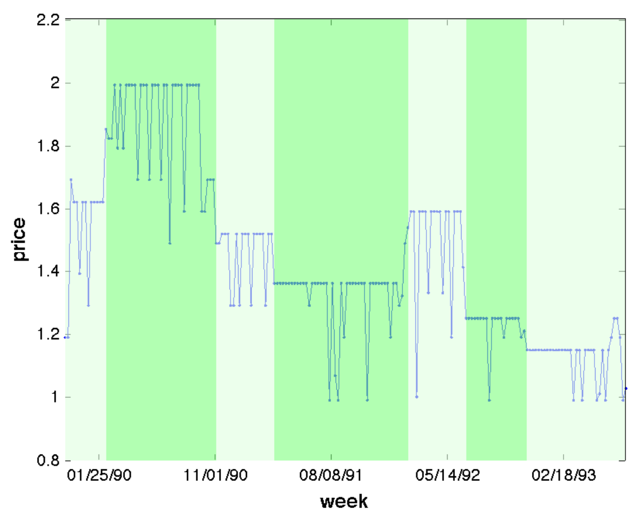

Figure 2. Frozen juice price series from Dominick's. Shading marks pricing regimes identified by the break test.

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 (2009), and the rolling mode filter proposed by Kehoe and Midrigan (2010). Simulations suggest that the break test is preferable: while each filter does particularly well on specific data generating processes, the break test does well on average: 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 weekly product-level pricing data from Dominick's chain of grocery stores. 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. Hence, 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).

The break test identifies infrequent regime changes in Dominick's data. While individual prices change with a median weekly frequency of 22%, implying a price duration of four weeks, the median duration of regimes is 31 weeks. However, regime lengths exhibit considerable heterogeneity, both across products and over time, and there is no evidence of synchronization in the timing of regime changes across products. Hence, a regime-based model of pricing would require staggered updating of regimes.

The identified regime changes can be viewed as estimates of transition points to new pricing plans. Each pricing plan is identified with a distribution of prices that consists of a small number of distinct prices relative to both the duration of the plan and the frequency of price changes inside the plan. While the frequency of regime changes is very low, inside regimes that consist of more than two distinct prices, prices change with a frequency of 32%. Nevertheless, across regimes with more than two unique price quotes, the median number of unique quotes per regime is 5 and the interquartile range is $[3, 8]$. For approximately 90% of all regimes, five or fewer unique price quotes account for more than 90% of the prices inside the regime.

Consistent with the existing empirical literature, I find significant rigidity at the regime's mode: the median frequency of the most frequently charged price per regime is 74% (excluding regimes that consist of a single price). I also find that the highest price is the mode of the regime in 67% of regimes (excluding ties), hence a model in which the high price dominates would capture the majority of empirical regimes. However, rigidity extends beyond the modal price, as I uncover substantial evidence against both the single sticky price model and the *one-to-flex* hypothesis that deviations from this mode are flexible. Only 5.2% of product series consist entirely of single-price regimes. More surprisingly, only 4.3% are purely one-to-flex series, in which prices deviate flexibly for one period from a rigid mode. The remaining 90.5% of products exhibit within-regime rigidity among non-modal prices. In 14.1% of these series, transitory price changes last more than one period, hence they can be included in the *one-to-flex* category. The remaining 76.4% of products consist of regimes in which multiple rigid prices are revisited over the life of the regime. These findings are not consistent with the recent theories of pricing such as those proposed by Kehoe and Midrigan (2010) or Guimaraes and Sheedy (2010), which assume that within regimes prices deviate flexibly from a rigid mode. In these models, revisiting non-modal prices would require that the shocks themselves are drawn from discrete distributions. It is also not the case that all prices other than the modal price are equally likely to be observed: inside regimes that contain more than two prices, the frequency of the non-modal prices declines from 12% (for the second most commonly observed price) to 3% (for the fifth most commonly observed price). Generating this pattern within the theories referenced above would require additional assumptions on the distributions of shocks.

Within-regime volatility is especially high for the rigid multi-price products: the within-regime frequency of price changes is 31.3% versus 13.0% for one-to-flex products. Moreover, the within-regime absolute size of price changes is 12.7% for rigid multi-price products and 7.2% for one-to-flex products. In a standard model, such large and frequent price changes would imply large and frequent shocks, and would additionally require that firms continuously monitor and respond to these shocks. In this sense, a price discrimination incentive a la Guimaraes and Sheedy (2010), which breaks the tight link between shock volatility and price volatility, seems an appealing ingredient in a model of multi-price strategies.

An important question relates to the timing of price changes inside regimes: is it the case that price changes occur randomly over the life of the regime, or are they clustered, for instance towards the end of the regime? As a starting point, I investigate the extent to which regimes can be interpreted as purely i.i.d. multi-price plans: for each regime identified in the data, I take the prices and their frequencies as given and perform random i.i.d. draws, and then I concatenate all resulting regimes into artificial series. When comparing the breaks and resulting statistics from the artificial series with those obtained from the actual data, I find that actual regimes are quite similar to the regimes obtained from i.i.d. simulations. Synchronization between the actual data and the artificial data is even stronger than that obtained when comparing two artificial series, suggesting that most of the within-regime volatility is interior, not at the edges of the regimes. Conversely, suppose that the data were mostly characterized by single-price regimes followed by periods of volatility marking the transition to a new sticky price. Generating artificial series in this way yields regimes that are significantly different from those found in the data, because the test breaks out the tranquil, single-price periods from the periods of volatility.

Section 2 details the statistical method employed to identify multiple potential change points in a given series, building on the basic two-sample Kolmogorov-Smirnov test. Section 3 presents the properties of pricing regimes identified by the break test in Dominick's data. Section 4 compares the performance of the break test to that of three different filtering methods in simulated data, underscoring the robustness of the break test across different data generating processes. Section 5 concludes.

2 The Break Test

I propose to test for the existence of one or more break points at unknown locations in individual pricing series. The *break test* builds on the Kolmogorov-Smirnov two-sample statistic, which tests whether or not two samples are likely to have been drawn from the same distribution. Hence, transitions to new pricing regimes are identified when the distribution of prices is estimated to have changed. The break test is an iterative procedure that identifies breaks in a series 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.

In simulations, the break test correctly rejects the null of no break for 91% of the simulated breaks; of all the breaks identified by the test, on average less than 2% are based on an incorrect rejection of the null. Upon rejecting the null, the estimated change points coincide exactly with the true change points 94% of the time; otherwise they are off by two periods, on average. The strength of this non-parametric test is that it is sensitive to any salient differences in the empirical distributions of the two samples. Furthermore, it can be applied to discrete or continuous random variables, it can be adapted to serially correlated data, and – importantly for our application to pricing data – it performs well on small samples.[3]

The first part of this section describes the break test, defining 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 approximate product-level pricing patterns. The last subsection evaluates the overall performance of the test: its ability to correctly reject the null and to correctly identify the timing of a change

[3]Brodsky and Darkhovsky (1993) provide a review of the Kolmogorov-Smirnov test and other non-parametric change point methods. To my knowledge, the existing literature on estimating change points using Kolmogorov-Smirnov focuses on the identification of a single change point. Moreover, applications to potentially discrete distributions are restricted to the one-sample goodness-of-fit Kolmogorov test. Hence both the standard method and the critical value for the test have to be adjusted when testing for an unknown number of break points at unknown locations.

given a rejection (the joint performance of the statistics S_n and τ_n).

2.1 Test Statistics

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:

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

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

$$\widehat{F}_1^t(p) \equiv \frac{1}{t}\sum_{s=1}^{t} \mathbf{1}_{\{p_s^n \leq p\}} \quad \text{and} \quad \widehat{G}_{t+1}^n(p) \equiv \frac{1}{n-t}\sum_{s=t+1}^{n} \mathbf{1}_{\{p_s^n \leq p\}} \tag{2}$$

Figure 3 plots the CDFs of two consecutive sub-samples of the pricing 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.

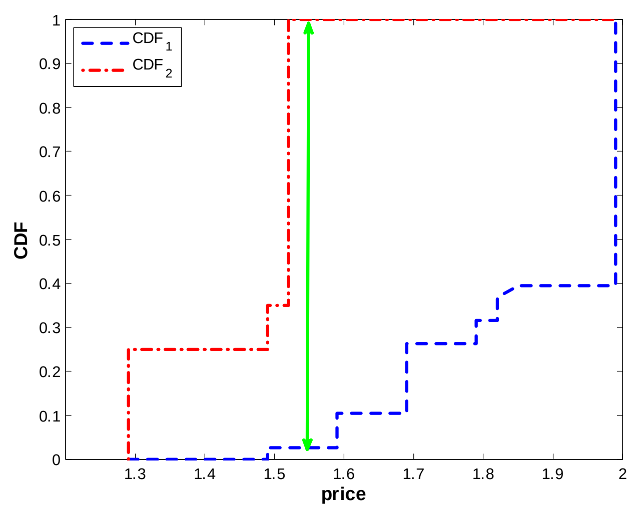

Figure 3. Kolmogorov-Smirnov distance (maximum marked by vertical line).

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):

$$S_n \equiv \sqrt{n} \max_{t \in T_n} \left[\frac{t}{n} \left(\frac{n-t}{n} \right) D_n(t) \right] \tag{3}$$

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 hence 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 sub-section), the next step is to estimate the location of the change: the change point estimate τ_n is given by:

$$\tau_n \equiv \arg\max_{t \in T_n} \sqrt{\frac{t\,(n-t)}{n}} D_n(t) \tag{4}$$

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 change point 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. This sequential approach to estimating change points is similar to that proposed by Bai and Perron (1998) for sequentially estimating multiple breaks in a linear regression model.

A potential concern with this multi-break adaptation of the Kolmogorov-Smirnov test is that the algorithm is testing for the existence of a *single* change point at each step. In the extreme, consider a series in which there are only two regimes that alternate over the course of the series. Because of this symmetry, the break test may fail to find a change point at the very first stage. However, this concern is not important for pricing data, since there is no evidence of such alternating patterns in product-level price series.[4]

[4] An alternative approach would be to employ a sequential test, in which the series are treated as ordered, and the estimate of the location of a change point depends only on the data before the estimated location. Since the Kolmogorov-Smirnov statistic requires using data both before and after the hypothesized change point, such a test would require a different statistic. Moreover, the estimated change points would not be invariant if we were to reverse the order of the observations. Early work implemented a rudimentary sequential test based on changes in the set of values observed, where the set size was neither predetermined, nor constant across regimes. This test however yielded estimates that were very sensitive to whether one conducted the test from the first to the last observation, or in reverse.

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 pricing series are notoriously heterogeneous, hence the specification of the test should be robust *across* different types of processes. Hence, I assume that the true data generating process for product-level prices is a mixture of processes and I use simulations to determine a single critical value to be used across all of the simulated processes.

Given the iterative nature of the test, the critical value determines only how soon the algorithm stops in its search for change points: 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.

2.2.1 Simulated Processes

I simulate processes that represent both recent theoretical models of price-setting and the most commonly observed pricing patterns in grocery data. Specifically, I generate the following four processes: 1) sequences of *single sticky prices* (such as those generated by Calvo 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 (consistent with the assumptions of a rolling mode price filter and with the price-setting model of Kehoe and Midrigan, 2009); 3) sequences of *downward-flex plans*, which consist of sticky prices accompanied by frequent, flexible downward deviations (consistent with the assumptions of a v-shaped sales filter and with models such as the dynamic version of Guimaraes and Sheedy, 2009); and 4) sequences similar to the frozen juice series shown in the introduction, labeled *rigid multi-price plans*, each consisting of a small number of distinct prices alternating over the life of the pricing plan. Figure 4 shows sample series with multiple regime breaks for each of the four simulated processes.

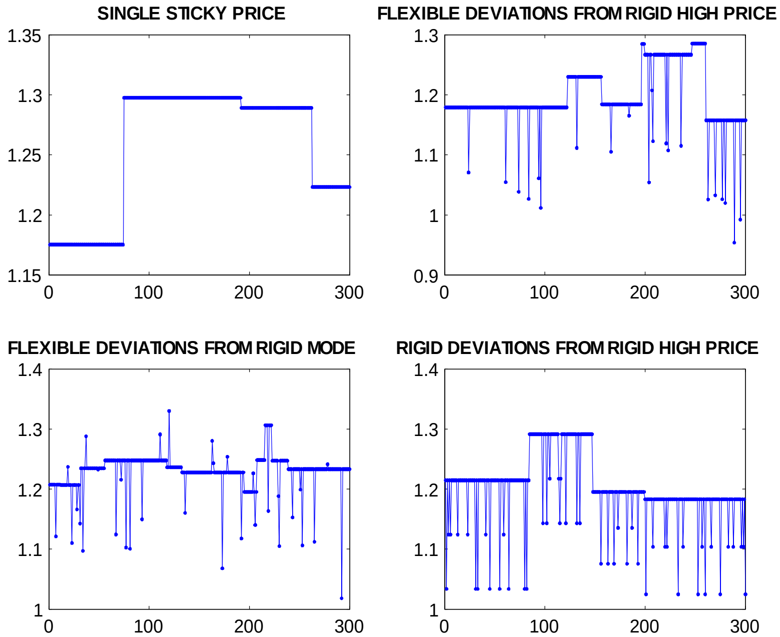

Figure 4. Samples for the four simulated processes.

The processes are calibrated to roughly match the volatility of the Dominick's data set: 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%. Table 1 summarizes the specification of the four processes, discussed further below.

Table 1. Specification of simulated processes.

	I.	II.	III.	IV.
Proba of regular Δp	$\beta \in (0,1)$	$\beta \in (0,1)$	$\beta \in (0,1)$	$\beta \in (0,1)$
Shock to regular price	$\varepsilon \sim N\left(\mu, \sigma^2\right)$	$\varepsilon \sim N\left(\mu, \sigma^2\right)$	$\varepsilon \sim N\left(\mu, \sigma^2\right)$	$\varepsilon \sim N\left(\mu, \sigma^2\right)$
Initial regular price	$p_0^R = \exp\left(\varepsilon_0\right)$	$p_0^R = \exp\left(\varepsilon_0\right)$	$p_0^R = \exp\left(\varepsilon_0\right)$	$p_0^R = \exp\left(\varepsilon_0\right)$
Proba of transitory Δp	$\delta = 0$	$\delta \in (0,1)$	$\delta \in (0,1)$	$\delta \in (0,1)$
Shock to transitory price	$\varepsilon^T = 0$	$\varepsilon^T \sim N(\mu_T, \sigma_T^2)$	$\varepsilon^T \sim N(\mu_T, \sigma_T^2)$	$\varepsilon^T \sim N(\mu_T, \sigma_T^2)$
Additional constraint	-	$\mu_T = 0$	$\mu - 3\sigma > \mu_T + 3\sigma_T$	$\mu - 3\sigma > \mu_T + 3\sigma_T$
Additional step	-	-	$l_\delta \geq 1$	quantize s.t. $n_$distinct $= 3$
Initial price	$p_0 = p_0^R$	$p_0 = p_0^R$	$p_0 = p_0^R$	$p_0 = p_0^R$

For process I, the price series is given by:

$$p_{t+1} = B_{t+1} \exp\{\varepsilon_{t+1}\} + (1 - B_{t+1}) p_t \tag{5}$$

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

For the multi-price processes, the full price series is given by:

$$p_{t+1} = B_{t+1} \exp\{\varepsilon_{t+1}\} + (1 - B_{t+1}) D_{t+1} p^R_t \exp\left(\varepsilon^T_{t+1}\right) + (1 - B_{t+1})(1 - D_{t+1}) p^R_t \tag{6}$$

where the sequence $\{D_t\}$ is a Bernoulli trial with probability of success δ.

For process III, transitory price cuts can last one or more periods, with the maximum length of a transitory price parameterized by $l_\delta \geq 1$.

The sequences of discrete, rigid multi-price plans (process IV) are generated by collapsing the simulated values inside each regime to three bins, each corresponding to a unique price quote, such that each regime consists of only three distinct prices.

I eliminate from simulations all regimes that last exactly one period: for a particular realization, the break test (or any other test) cannot identify single-period regimes as distinct from transitory price changes, hence I require all regimes to last at least two periods.

2.2.2 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 change-points at any $t \in T_n$,

$$S_n \to \widetilde{K} \equiv \sup_{u\in[0,1]} \sup_{v\in[0,1]} |B(u,v)| \tag{7}$$

where $B(\cdot,\cdot)$ is the two-dimensional Brownian bridge on $[0,1]$.[5] Hence equation (7) provides asymptotic critical values for the test of a single break at an unknown location: the null is rejected at level α if $S_n > \widetilde{K}_\alpha$, where $\widetilde{K}_\alpha$ is found from $\mathbf{P}(\widetilde{K} \leq \widetilde{K}_\alpha) = 1 - \alpha$. They find that the asymptotic critical values provide a reasonable approximation for sample sizes of at least 50 observations.

If prices are correlated over time (and hence they contain less information than a same-sized sample of independent observations), higher critical values would be required for the same level of significance. Conversely, if prices are drawn from discrete distributions, using these critical values would make the test conservative, therefore requiring a lower critical value. Testing for potentially multiple breaks per series by applying the test iteratively also means that I can no longer rely on the asymptotic distribution of the test statistic derived by Deshayes and Picard (1986). Hence, I determine the appropriate critical value using simulations in which I compare the results of the break test with the true location of change points.

For simplicity, I use a single critical value across all sample sizes. This 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 contains a true change point; it is reported as a fraction of the number of true change points in the simulation. Obtaining a low value for *good_ reject* implies that the test is not sensitive enough such that many true change points 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 true change point, and it is reported as a fraction of the number of estimated change points. A high *bad_ reject* implies that the test yields too many false positives, hence the critical value needs to be increased.

The tension in choosing the critical value is between process I and the multi-price processes: for the single sticky price process, a critical value arbitrarily close to zero ensures that all change points are identified (even those near the edges of a sample), 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 multi-price processes: for process III, in the case when transitory prices can last more than one period ($l_\delta > 1$), too low of a critical value leads to mislabeling transitory prices as new regimes; similarly,

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

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 provided by Deshayes and Picard (1986), the break test correctly finds 84% of the simulated breaks. It fails to identify particularly short regimes, and hence underestimates the average duration of regimes by as much as eight periods in simulations for process IV. 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: $K = 0.61$ 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 substantially increases the incidence of false positives.

At this 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 relatively weak power in identifying particularly short regimes (lasting two to four periods).

Table 2. Break test critical value determination.

Critical value	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	+7.3	+5.8	+4.7	+3.1	+2.9	−0.2	−4.9

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 $T_2 \subset T_1$.

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. This statistic is almost entirely driven by the presence of very short regimes in the simulations and results in overestimating the average regime length by three periods. Constraining the simulations to generate only regimes of at least four periods would result in identification of virtually all regimes and would eliminate the bias in the estimated average regime length.

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 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 sub-section: 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. For 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 change points and the true change points among non-synchronized breaks would decline to 1.5 periods.

Table 3. Break test 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

Thus far, I have simulated sequences that contained regime breaks. If I simulate the four processes described above but introduce no regime breaks ($\beta = 0$ in Table 1), the break test correctly finds no breaks.

3 Regimes in the Data

This section discusses the properties of good-level pricing series viewed through the lens of infrequently updated regimes. I apply the break test to pricing series from Dominick's Finer Foods stores. Grocery prices are highly volatile and, more importantly, exhibit precisely the sharp, yet transitory, price swings that have recently come to the forefront of the price dynamics literature.

The data set is available from September 1989 until May 1997 and includes weekly prices for a large number of products in 29 categories, including various household supplies and non-perishable food. Dominick's policy is to change prices only once a week (on Wednesdays), hence we are not missing any intra-week price changes. Another advantage of this data, despite its fairly narrow coverage, is the relatively long length of series, for example when compared with series in the BLS database of products underlying the US CPI. The median (average) length of the series is 148 (175) weeks per UPC. The sample used contains only price series that have at least 52 observations and on average at most one missing observation per year. I further restrict the sample to a single store, selected as the store with the most overall observations from the control group.[6] The total size of the restricted sample is $690,578$ observations for $3,951$ unique UPCs, and 257 manufacturers, in 27 product categories (see Appendix A for more summary statistics of the resulting sample).

3.1 Regime Changes vs. Price Changes

Regime changes are estimated to occur infrequently: while the raw weekly frequency of price changes is 21.5% for the median product, implying a median price duration of four weeks, the median implied duration[7] of regimes of 31 weeks across all products. There is considerable underlying heterogeneity, both within and across categories, particularly above the median duration of each product category. Figure 5 shows the median implied regime duration for each category, ordered from lowest to highest, as well as the interquartile range.[8]

[6] Preliminary work suggests virtually identical results across the different control stores. During the period in which the data were collected, pricing experiments were conducted at some of the stores within the chain. To avoid spurrious findings, I only use data from the control stores throughout. See Hoch et al (1994) for a description of the experiments.

[7] Due to the large number of trucated series, 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 $d = -1/\ln(1 - freq)$.

[8] The overall median is towards the high end due to the relatively larger number of observations for products with longer regimes.

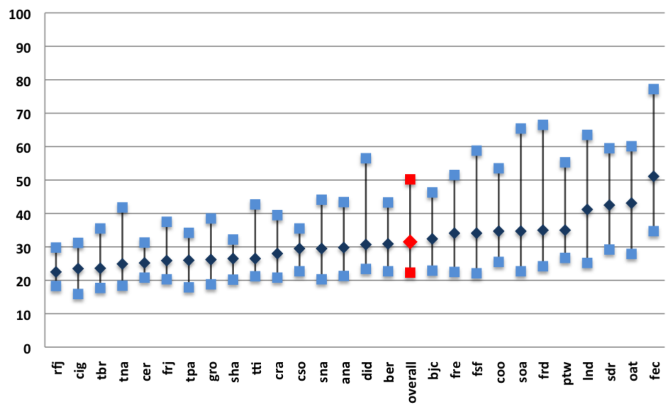

Figure 5. Median regime duration and interquartile range for each category.

Despite their relatively long duration, regimes are characterized by a small number of distinct prices: among regimes that consist of more than 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 regimes, prices change often: across all regimes, the median weekly frequency of within-regime price changes is 29.3%, and it increases to 31.6% for regimes that include more than two distinct prices. Hence, 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.

3.2 Within-Regime Pricing Strategies

Firstly, consistent with existing empirical evidence, the break test recovers high rigidity at the modal price per regime: the most commonly quoted price within a regime occurs with median frequency of 73.6%. Moreover, 99% of regimes have at least one rigid price. In fact, all products in the data set can be grouped into three categories: products characterized exclusively by single sticky price regimes (where I also include regimes in which I observe a single deviation from the modal price, suggesting that transitory price changes are not an important aspect of price setting for this product), 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, *rigid multi-price* products, which contain regimes in which at least two prices are revisited over the life of the regime.

The single sticky price series correspond to the most commonly used models of pricing, in which the firm follows a single-price strategy, employing either time-dependent or state-dependent price adjustment. However, I find that this pattern characterizes only 5.2% of products in the Dominick's data. For these products, the median regime duration is 44 weeks, suggesting that these are products that face a low volatility of costs and demand and/or a relatively high cost of adjusting their nominal price.

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 in Section 2.2. For example, Kehoe and Midrigan (2010) and the dynamic extension of Guimaraes and Sheedy (2010) generate *one-to-flex* patterns in which transitory prices last one period. I find that this pattern only accounts for 18.4% of products, of which, for 4.3% of products, transitory prices last only one period, and for the remaining 14.1% of products, transitory price changes can last more than one week. For these products, the median implied regime duration is 30 weeks.

The remaining 76.4% of products all exhibit some degree of within-regime rigidity among non-modal prices, in which at least two prices are revisited over the life of the regime. The median implied duration of regimes for these products is 32 weeks. It is important to note that in computing this statistic, I do not require all regimes in the series to be multi-rigid: it seems reasonable to 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 rigid multi-price regime are labeled as pursuing a rigid multi-price strategy. Under this assumption, we would expect single-price and one-to-flex regimes that are part of rigid multi-price series to 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 44 weeks, the average length of *all* single-price regimes is 19 weeks, indicating that single-price regimes included in the one-to-flex and rigid multi-price series are much shorter. Similarly, the average length of one-to-flex regimes in purely one-to-flex series is 36 weeks, versus 27 weeks for all one-to-flex series, indicating that one-to-flex regimes in the rigid multi-price series are significantly shorter. 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.

Table 4 summarizes these results.

Table 4. Breakdown of data by pricing strategy.

	Single price	One-to-flex	Rigid multi-price
% of regimes[9]	33.0	30.9	35.1
% of products[10]	5.2	18.4	76.4
avg. regime length overall, weeks	19	27	49
avg. regime length at product level, weeks	44	30 (36)	32 (49)

3.3 Volatility of Rigid Multi-Price Products

One-to-flex products differ meaningfully from rigid multi-price products in terms of within-regime volatility: as shown in Table 5, the within-regime frequency of price changes is 13.0% for one-to-flex products and 31.3% for rigid multi-price products. Moreover, the within-regime absolute size of price changes is 7.2% for one-to-flex products and 12.7% for rigid multi-price products. A potential interpretation of this finding is that for products for which frequent transitory adjustment is desirable, the seller finds it optimal – possibly as a way to economize on information acquisition – to design a pricing strategy that consists of a small number of prices, among which to alternate over the course of the regime.

[9] 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 the regime are unique.

[10] The numbers in this row add up to 100% because all series exhibit some form of rigidity in prices.

Table 5. Within-regime volatility.

	One-to-flex products	Rigid multi-price products	All products
frequency of price changes	13.0%	31.3%	29.3%
absolute size of price changes	7.2%	12.7%	12.0%

3.4 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 6 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.

Table 6. Relationship between regime length and number of distinct prices observed.

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

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

How much rigidity is there in the pricing 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, but often occurs more than five times; 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.

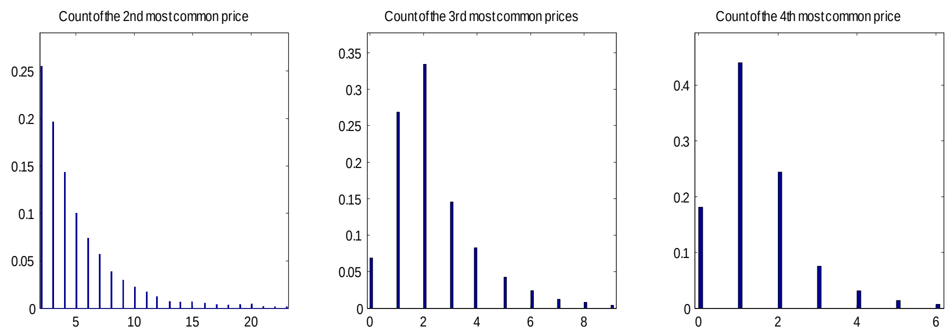

Figure 6. Count of top prices across regimes for multi-rigid products.

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.

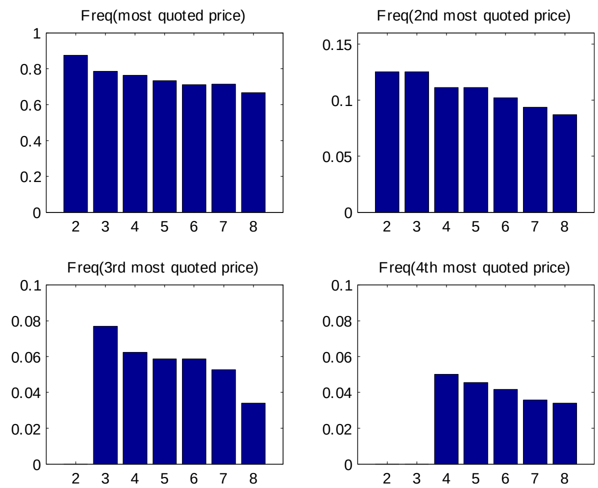

Figure 7. Frequency with which common prices are charged as a function of cardinality.

It is important to note that Dominick's data provides a lower bound on the within-regime rigidity of price quotes for these particular products, due to the fact that the reported prices are averages of prices paid by customers with and without loyalty cards. Hence, even small changes in the composition of customers from one week to the next would lead to a change in the reported price even though the sticker price has not changed.

The panels in Figure 8 show histograms for the frequency with which top prices are charged across all regimes for all products. The histograms illustrate more clearly 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.

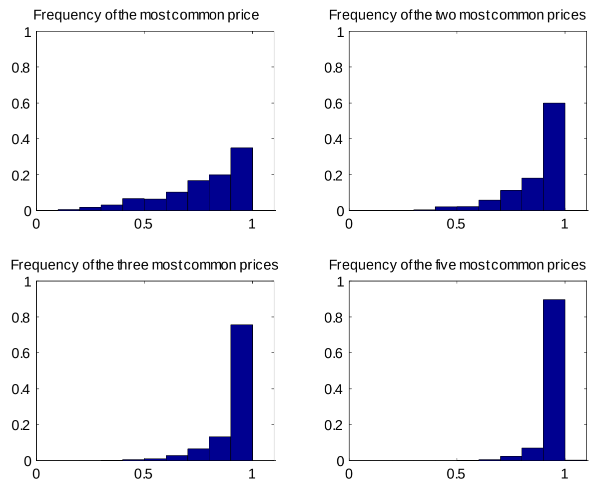

Figure 8. Histograms for the frequency with which top prices are charged.

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.

Table 7. "Top" prices for rigid multi-price regimes. Note, Klenow & Malin numbers are for monthly data.

Median % (Mean %)	**Top Price**	**Top 2 Prices**	**Top 3 Prices**	**Top 4 Prices**	**Top 5 Prices**	**# Distinct Prices**
Multi-Rigid Regimes	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	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	31.4 (37.6)	50.9 (53.2)	62.7 (61.3)	70.1 (66.2)	n.a. (n.a.)	n.a. (n.a.)

The statistics at the regime level suggest that models which assume that only the mode price is rigid, with all other prices being adjusted flexibly may miss an additional source of rigidity in terms of price levels.

The evidence presented in this section is consistent with the proposed price plans of Eichenbaum, Jaimovich and Rebelo (2009), according to which firms are assumed to choose from a small set of prices that is updated relatively infrequently, subject to a cost. Generating such pricing plans endogenously remains an open topic in the literature. One approach is Matejka's (2009) model of rational inattention, in which information-constrained firms facing shocks drawn from a uniform distribution respond by optimally choosing a pricing strategy that consists of a small number of distinct prices[11].

However, the resulting distributions from this baseline rational inattention model are too flat. The information constraint requires that all signals are on average equally informative about the true state, subject to the degree of asymmetry in the firm's profit function and in the density of the shocks. Under a symmetric objective and a uniform prior, all signals would be equally valuable, and hence would be observed with equal probabilities. Under an asymmetric objective function, precision is less valuable in the relatively flat region of the objective function, hence this area will be more coarsely represented by the signalling mechanism. In a standard pricing model in which the firm's profit function is flatter at high prices, this translates into a higher unconditional probability of observing the high price. However, for a wide range of parametrizations, the asymmetry in the profit function of a standard constant markup model of monopolistic competition is not sufficient to generate the degree of asymmetry observed in the data. For instance, in a parameterization of Matejka's (2009) model with uniform shocks, moderate curvature of the profit function[12], and an information cost low enough to generate five distinct prices, the resulting frequency of the top price is only 40% (versus 74% in the data) and the second most common price is charged 21% of the time (versus 12% in the data).

Hence, rational inattention alone generates too flat of a distribution compared with the histograms in Figure 7, while the one-to-flex models generate no mass points beyond the top price.

[11] Discreteness arises in the solution to a rational inattention problem if the density over possible states approaches zero faster than the rate at which losses from the imprecision in the signal become arbitrarily large. This is necessarily the case for states drawn from bounded distributions. See Fix (1978) and Rose (1994) for results in information theory, and Matejka and Sims (2010) for results in economics.

[12] Given CES preferences with elasticity of demand ε, yielding demand $q(p) = p^{-\varepsilon}$, and the firm's profit function, $\pi(p) = pq(p) - \omega q(p)^{\gamma}$, the optimal price is a constant markup over the marginal cost. The only source of uncertainty is the cost shock, ω, drawn from a known distribution, $g(\omega)$. The numbers quoted in the text are generated by solving the rational inattention problem using $\varepsilon = 6$, $\gamma = 1.5$, and ω drawn from a uniform distribution.

3.6 Sticky i.i.d. Plans?

In this section, I evaluate the extent to which the pattern of prices inside regimes can be approximated by independent draws from a fixed distribution. Suppose that in the data, regimes looked nothing like those that would be generated via i.i.d. draws. This would suggest that the *ordering* of prices inside regimes is important and that the change points identified on the i.i.d. simulations would not be well synchronized with those in the data. Hence this exercise essentially evaluates to what extent the timing of price changes is important for the identification of regime changes. Klenow and Malin (2010) found limited evidence of i.i.d. plans in monthly US CPI prices. I compare statistics from the data to those generated when prices inside regimes are sampled without replacement from the regime's CDF and do not find strong evidence against the i.i.d. hypothesis. I also compare results to those generated when distinct prices inside regimes are clustered, behaving as sequences of single-price regimes interrupted by transitional periods between regimes, and find weak alignment between the data and artificial series constructed in two different ways.

For the i.i.d. simulation, suppose that for a particular product series, $\{p_t\}$, the algorithm identified regimes r_j, $j = 1, ..., J$. For each regime r_j, I perform independent draws from the distribution of prices inside the regime and concatenate all simulated regime sub-samples into a simulated product series, $\{\widetilde{p}_t\}$. I then rerun the test to identify change points on the new series and compute resulting statistics.

Table 8. Simulation results under assumption of i.i.d. multi-price plans

Samples:	**1 = data 2 = simulation**	**1 = simulation 2 = simulation**
Success (% of sample 1)	99.6	99.9
Synch'd breaks, exact (% of sample 1)	80.4	70.6
Synch'd breaks, +/-1 wk (% of sample 1)	88.2	83.7

The i.i.d. simulation is very close to the actual data in terms of identification and location of the breaks: the simulation rejects the null in nearly 100% of the cases in which the data also rejected the null. Moreover, 80% of the time, breaks in the simulation are found in exactly the same place as those in the data, and synchronization increases to 88% when allowing for a window of $+/-1$ week. In Table 8, I report the comparison between data and simulation alongside a comparison between two simulations, as a measure of how high a synchronization can be expected if the two series both consist of i.i.d. regimes. The fact that synchronization is higher in the first column suggests that volatility is mostly interior, rather than near the edges of each regime.

For the single-sticky price view, first suppose that a particular regime r_j contains – in order of first occurrence – k unique prices, $\{p_k\}$, which occur with frequencies $\{q_k\}$. I rearrange the series such that when a unique price is first observed, all subsequent realizations of that price in regime r_j follow it. For example, the series $\{2, 1, 3, 1, 2\}$ becomes $\{2, 2, 1, 1, 3\}$. This experiment provides a way to evaluate the extent to which rigid multi-price regimes are estimated to occur frequently predominantly due to the relative weak power of the break test on particularly short regimes, such that most of the within-regime rigidity measured in the previous sub-section is due to this lack of identification. I find that such a pattern of essentially single-price regimes is not a strong fit for the actual data.

Table 9 reports that the singletons simulation finds 72% more change points than the data.

Table 9. Simulation results under assumption of sequences of single sticky prices

Samples:	**1 = data 2 = simulation**	**1 = simulation 2 = simulation**
Success (% of sample 1)	171.7	-
Synch'd breaks, exact (% of sample 1)	45.8	-
Synch'd breaks, +/-1 wk (% of sample 1)	82.2	-

As another alternative - the "experimental view" - I consider the possibility that single sticky price regimes are followed by a period of volatility, before the firm settles on a new sticky price to charge. 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.

Figure 9 contrasts the i.i.d version of two regimes with this single-sticky + volatility version.

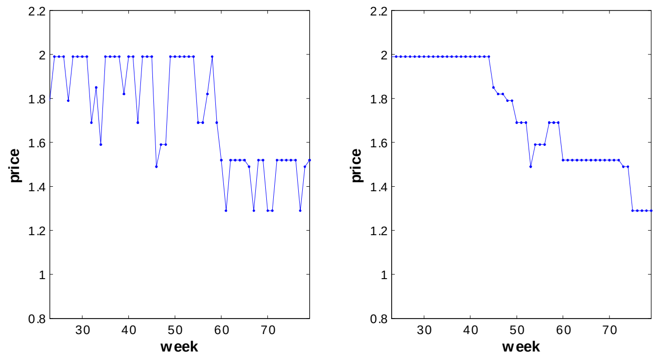

Figure 9. Pure i.i.d regimes vs. experimental.

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: 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.

Table 10. Simulation results under the experimental view.

Samples :	**1=data**
Success (% of sample 1)	153.8
Synch 'd breaks , exact (% of sample 1)	75.3
Synch 'd breaks , +/- 1wk (% of sample 1)	85.0

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%. 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 closer to the actual data.

Table 11. Multi-price Plans.[13]

(Inside regimes)	**Multi-price plans (iid)**	**Data**	**No iid multi-price**
% Uninterrupted prices	69.5	80.0	93.1
% Comeback prices	66.7	33.3	0.0
Median regime duration	31	31	19
Single-price regimes (% of all regimes)	32.0	33.0	55.6
Multi-rigid-price regimes (% of all regimes)	47.9	35.1	1.2

4 Break Test versus Filters

I compare the regimes-based method of analyzing pricing 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 (2009), and the running mode filter of Kehoe and Midrigan (2010), which is similar to that of Chahrour (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.

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

[13] Note, here single-price regimes are defined as regimes consisting of exactly one price quote; multi-price regimes are all regimes with more than two distinct price quotes.

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 2, the break test offers a direct way to evaluate the assumptions invoked by various filtering methods. 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 2.

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 Nakamura and Steinsson (2008). A slightly modified version of their algorithm is reproduced in Appendix B. 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. 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.

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

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 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 as a function of the sale window, J: *breaks_regtest* reports the total number of regime changes found by the break test across all series; *breaks_vshaped* reports the number of change points of the regular price found by the different parametrizations of the v-shaped filter; *frac_vshaped* reports the ratio of breaks found by the v-shaped filter to the break test, computed at the series level and averaged across all series in the data ; *frac_synch_exact* reports the number of breaks exactly synchronized between the two methods as a fraction of the number of breaks found by the regime test (also computed for each series and then averaged across all series); and *tau_gap* reports the median minimum distance between the change points estimated by the two methods, excluding exact synchronizations[14].

I report results for parametrizations of K, L, F that yield the highest degree of synchronization between the v-shaped filter and the regime 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 increases exact synchronization

[14] 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.

from 65% to 80%.

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

J	3	5	7	9	12
breaks_regtest (total)	18,412	18,412	18,412	18,412	18,412
breaks_vshaped (total)	66,424	39,659	32,138	28,852	26,596
duration_vsh (median)	12.4	19.0	23.9	27.5	28.5
% frac_vshaped (average)	360	214	177	164	155
% frac_synch_exact (average)	80	68	64	61	58
tau_gap (median)	4.0	5.0	6.0	8.0	9.0

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 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.

Table 14. Simulation results for v-shaped filter and regime test

# breaks/# true breaks	VSH	BT
I (single sticky)	100%	90%
II (mode-to-flex)	450%	94%
III (downward-flex)	115%	96%
IV (multi-rigid)	122%	91%

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.

4.2 Reference Price Filter

I implement the reference price filter proposed by Eichenbaum et al (2009). 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.

Table 15. Reference price filter on Dominick's.

	EJR (6)	EJR (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%)

I find that 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.

Table 16. Reference price filter on simulated data.

	EJR (w=6)	BT
# breaks/# true breaks	93%*	93%
of which, exact synch with true breaks	17%	94%

4.3 Rolling Mode Price Filter

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. 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, whereas every price change that coincides with a change in the regular price is defined as regular. 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).

Table 17. Simulation results for Kehoe-Midrigan filter

	KM (w=10)	BT
# breaks/# true breaks	94%	93%
of which, exact synch with true breaks	94%	94%
mean distance if not synch'd	2	2
overshoot of median reg length	+2.3	+3

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 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).

5 Concluding Remarks

This paper proposes a new method to characterize product-level pricing series, by splitting series into sequences of regimes and characterizing the distribution of prices within each regime. At the heart of the method is the two-sample Kolmogorov-Smirnov statistic, which tests if two samples are likely to have been drawn from the same distribution. This statistic is intuitive and very appealing since it is robust to different types of changes in the underlying distributions, and is suited to small sample analysis.

I apply this method to grocery store data and find that regime changes are robustly estimated, with regimes typically lasting 7 months. I find strong evidence against both the single sticky 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 since they either assume diffusion resulting from full flexibility beyond the modal price – as in Kehoe and Midrigan (2010) or Guimaraes and Sheedy (2010) – or they generate too flat of a distribution, as in the case of Matejka (2009). I also find some evidence that pricing regimes can be viewed as sequences of i.i.d. multi-price plans, with statistics not drastically different from simulation-based statistics. Moreover, I find that products characterized by multiple rigid prices per regime are also highly volatile, compared with one-to-flex series. A potential interpretation of this finding 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. Moreover, the fact that multi-rigid regimes last significantly longer than other regimes may be indicative of the fact that the seller uses these multiple prices to adjust to market conditions and can therefore conduct a revision of the entire strategy less frequently. Another 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

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, charged with sharply declining frequency, and in which the highest price is typically the most frequently quoted price. 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.

References

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

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

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

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

Burstein, Ariel and Christian Hellwig (2007), "Prices and Market Shares in a Menu Cost Model," unpublished paper, UCLA and Stanford University.

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

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

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

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

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

Conover, W. J. (1980), Practical Nonparametric Statistics. New York: Wiley.

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.

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo (2009), "Reference Prices and Nominal Rigidities," American Economic Review, forthcoming.

Guimaraes, Bernardo and Kevin D. Sheedy (2010), "Sales and Monetary Policy," American Economic Review, forthcoming.

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

Kehoe, Patrick and Virgiliu Midrigan (2010), "Prices Are Sticky After All," unpublished paper, Princeton University and NYU.

Klenow, Peter J. and Benjamin A. Malin (2010), "Microeconomic Evidence on Price-Setting," Handbood of Monetary Economics, forthcoming.

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

Matejka, Filip. (2008), "Rationally Inattentive Seller: Sales and Discrete Pricing," unpublished paper, Princeton University.

Nakamura, Emi and Jon Steinsson (2008), "Five Facts about Prices: A Reevaluation of Menu Cost Models," Quarterly Journal of Economics, 123(4): 1415-1464.

Wood, L. Constance and Michele M. Altavela (1978), "Large-Sample Results for Kolmogorov-Smirnov Statistics for Discrete Distributions," Biometrika, 65(1): 235-239.

Appendix A

The sample used contains only product series that have at least 52 observations and on average, at most one missing observation per year. The table below reports the breakdown of the resulting sample by category. Note, the total number of unique manufacturers is lower than the column sum because of some repeat manufacturers across categories.

Table A1. Summary of Dominick's data used.

Code	Category	Raw (# obs)	Used (# obs)	# Manufs.	# UPCs	Median # obs/UPC	Mean # obs/UPC
ana	Analgesics	3,060,156	11,292	11	72	133	157
bat	Bath soap	418,097	-	-	-	-	-
ber	Beer	1,970,266	11,310	15	84	138	135
bjc	Bottled Juices	4,325,024	41,760	28	232	132	180
cer	Cereals	4,751,202	45,128	12	231	192	195
cig	Cigarettes	1,810,615	11,532	4	61	181	189
coo	Cookies	7,635,071	56,327	26	310	137	182
cra	Crackers	2,245,703	16,960	22	105	114	162
cso	Canned Soup	5,551,684	47,480	12	248	196	191
did	Dish Detergent	2,183,582	18,236	9	127	98	144
fec	Front-end Candies	4,475,750	46,635	23	228	164	205
frd	Frozen Dinners	1,654,053	12,316	6	88	132	140
fre	Frozen Entrees	7,232,080	68,580	16	368	157	186
frj	Frozen Juices	2,387,420	27,316	14	108	274	253
fsf	Fabric Softeners	2,296,612	18,555	18	116	134	160
gro	Grooming Products	4,065,694	10,589	10	84	107	126
lnd	Laundry Detergents	3,303,174	24,236	13	206	96	118
oat	Oatmeal	981,263	9,650	6	45	301	214
ptw	Paper Towels	948,871	7,800	9	46	167	170
rfj	Refrigerated Juices	2,182,989	23,049	15	115	163	200
sdr	Soft Drinks	10,807,191	89,117	35	457	158	195
sha	Shampoos	4,676,790	4,056	11	47	66	86
sna	Snack Crackers	3,515,192	30,543	11	160	177	191
soa	Soaps	1,835,196	13,904	9	96	112	145
tbr	Toothbrushes	1,854,983	5,406	11	51	113	106
tna	Canned Tuna	2,403,558	14,035	20	95	131	148
tpa	Toothpastes	3,003,392	16,770	9	124	112	135
tti	Bathroom Tissues	1,159,016	7,996	9	47	185	170
Total		**99,545,249**	**690,578**	**257**	**3951**	**148**	**175**

Appendix B - Nakamura and Steinsson (2008) V-shaped Sales Filter

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, $r_t = p_t$.