Vanderbilt University Department of Economics Working Paper Series
Nashville, TN 37235-1819

Vanderbilt University Department of Economics Working Papers 13-00015

PRICE DISPERSION AND DEMAND UNCERTAINTY: EVIDENCE FROM US SCANNER DATA

Benjamin Eden
Vanderbilt University

Abstract

I use a flexible price version of the Prescott (1975) hotels model to explain variations in price dispersion across goods sold by supermarkets in Chicago. The main finding is that price dispersion measures are positively correlated with proxies for demand uncertainty. I also find that price dispersion measures are negatively correlated with the average price but are not negatively correlated with the revenues from selling the good (across stores and weeks) and with the number of stores that sell the good.

Citation: Benjamin Eden, (2013) "PRICE DISPERSION AND DEMAND UNCERTAINTY: EVIDENCE FROM US SCANNER DATA", *Vanderbilt University Department of Economics Working Papers*, VUECON-13-00015.
Contact: Benjamin Eden - ben.eden@vanderbilt.edu.
Submitted: September 27, 2013. **Published:** October 03, 2013.
URL:http://www.accessecon.com/Pubs/VUECON/VUECON-13-00015.pdf

PRICE DISPERSION AND DEMAND UNCERTAINTY: EVIDENCE FROM US SCANNER DATA

Benjamin Eden*

Vanderbilt University

August 2013

I use a flexible price version of the Prescott (1975) hotels model to explain variations in price dispersion across goods sold by supermarkets in Chicago. The main finding is that price dispersion measures are positively correlated with proxies for demand uncertainty. I also find that price dispersion measures are negatively correlated with the average price but are not negatively correlated with the revenues from selling the good (across stores and weeks) and with the number of stores that sell the good.

Key Words: Price Dispersion, Demand Uncertainty, Search Frictions, Sequential Trade.

JEL Classification: D50, D80, D83, L10

---

* I would like to than Maya Eden and Saul Lach for useful comments on an earlier draft and Vivian Ying Jiang for excellent research assistance.

1. INTRODUCTION

In the absence of aggregate demand uncertainty, getting price dispersion in equilibrium is a challenge. Diamond (1971) was the first to point out the difficulty. In his model the equilibrium price distribution is degenerate and all firms post the monopoly price. Diamond assumed that buyers sample one firm at a time. Burdett and Judd (1983) allowed for sampling more than one selling offer per period and show that price dispersion will arise if the probability of sampling more than one seller is between zero and one. If however the probability of sampling more than one seller goes to one we will converge to a single price equilibrium in which all firms post the competitive price. If the probability of sampling more than one seller goes to zero we will converge to a single price equilibrium in which all firms post the monopoly price (as in the Diamond model).

While getting equilibrium price dispersion is a challenge for search models, I argue in Eden (2013), that getting single price equilibrium is a challenge for Prescott type models that assume uncertainty about aggregate demand. The original Prescott (1975) model assumed that prices are set in advance and cheaper goods are sold first. In Eden (1990) I describe a sequential trade process that is consistent with Prescott's assumption. Buyers arrive at the market place sequentially. Each buyer sees all available offers, buys at the cheapest available price and disappears. Sellers must make irreversible selling decisions before they know the aggregate state of demand and in equilibrium they are indifferent between prices that are in the equilibrium range because the selling probability is lower for higher prices. Sellers in the model make time consistent plans and do not have an incentive to change prices during the trading process. Prices are thus completely flexible. The special case of no uncertainty about aggregate demand yields a single price equal to the competitive price. This is similar to the special case in which all buyers sample more than one price in the Burdett-Judd model.

There are versions of the Prescott model that assume price rigidity. See for example, Dana (1998, 1999) and Deneckere and Peck (2012). For the positive implications of the theory it does not matter whether a flexible price or a rigid price version of the model is employed. But sometimes the rigid price versions of the Prescott model tends to be lumped together with menu costs models that have very different empirical implications. See Eden (2001) and Baharad and Eden (2004). To distinguish the model from the menu cost literature I use here the flexible price version.

I follow Bental and Eden (1993) by allowing for storage and use the model to explain variations in price dispersion among goods sold by supermarkets. The model says that price dispersion is higher for (a) goods with more demand uncertainty and (b) goods with higher cost for delaying trade, like storage cost, depreciation and interest cost. I find strong support for the first hypothesis.

2. THEORY

As was said above, price dispersion in Prescott type models requires aggregate demand uncertainty. Here I use a version of the model in Bental and Eden (1993) to get a relationship between a specific measure of price dispersion and a specific measure of demand uncertainty.

<u>Sellers</u>:

The economy lasts forever. There are many goods and many sellers who can produce the goods at a constant unit cost. The unit cost of producing good j is λ_j. Production occurs at the beginning of each period before the beginning of trade. The seller knows the distribution of demand but at the time production decisions are made he does not know the realization.

Selling is uncertain. The representative seller faces a tradeoff between the probability of making a sale and the price: The lower the price, the higher is the probability of making a sale. In each period, sellers of good j have to choose between Z_j price tags: $P_{1j} < ... < P_{Zj}$. Posted prices do not change over time and therefore I drop the time index. I also drop the good index and consider a good with prices $P_1 < ... < P_Z$.[2]

The seller takes the probability that he can sell at each of the Z prices as given. The probability of making a sale at the price P_i is q_i, where $1 = q_1 > ... > q_Z > 0$. Goods that are not sold are carried as inventories to the next period. A unit stored can be used to reduce production next period and the value of a unit of inventories is therefore $\beta\lambda$, where $0 < \beta < 1$ reflects the cost of delay, storage cost and depreciation.

Sellers will post the price P_i on a strictly positive and finite number of units only if:

$$q_i P_i + (1 - q_i)\beta\lambda = \lambda \tag{1}$$

The arbitrage condition (1) is key. The left hand side of (1) is the expected revenues from putting the price tag P_i on one unit. With probability q_i the seller will get the quoted price and with probability $1 - q_i$ he will get the value of inventories. The right hand side is the unit production cost. The seller will put the price tag P_i on $0 < x < \infty$ units, only if the two are equal. Otherwise, if $q_i P_i + (1 - q_i)\beta\lambda > \lambda$ he will produce and

[2] There is no incentive in equilibrium to announce a price $P_i < p < P_{i+1}$ because the probability of making a sale at this price is the same as the probability of making a sale at the price P_{i+1}.

put the price tag on infinitely many units and if $q_i P_i + (1-q_i)\beta\lambda < \lambda$ he will not put the price tag on any unit.

Buyers:

Buyers arrive at the market place after sellers have already made their production decisions. Upon arrival they see all available offers and each buyer buys one unit at the cheapest available price. (Thus buyers' reservation price is sufficiently high).

The number of buyers that arrive in the market place in a typical period ($\tilde{N}$) is an iid discrete random variable that is uniformly distributed on the interval $[x, Zx]$, where both x and Z are good specific parameters. The random variable $\tilde{N}$ may take Z possible realizations: $N_1 < ... < N_Z$. State s occurs when $\tilde{N} = N_s$ with probability $\Pr ob(\tilde{N} = N_s) = \pi = 1/Z$. The difference between two consecutive realizations is constant and is given by: $N_s - N_{s-1} = N_1 = x$ for all s.

Buyers arrive in a sequential manner. The first batch of x buyers buys in the first market at the price P_1. If $s = 1$ and no more buyers arrive trade is over for the period. If $s > 1$ an additional batch of x buyers arrive and buys in the second market at the price P_2. Again, if $s = 2$ no more buyers arrive and trade is over for the period. Otherwise, if $s > 2$ a third batch arrives and buys in the third market at the price P_3 and so on. We thus have Z hypothetical markets that open sequentially. When $\tilde{N} = N_s$, the first s markets open and the goods allocated to these markets are sold. The goods allocated to the last $Z - s$ markets are not sold and are carried as inventories to the next period.

Equilibrium:

Using x_i to denote the supply to market i, I define equilibrium as follows.

Equilibrium is a vector of prices $(P_1,...,P_Z)$, a vector of probabilities $(q_1,...,q_Z)$ and a vector of supplies $(x_1,...,x_Z)$ such that (a) the probability that market i will open and goods with price tag P_i will be sold is: $q_i = \Pr ob(\tilde{N} \geq N_s) = (Z-i+1)\pi$, (b) the arbitrage condition (1) is satisfied and (c) the supply to market i is equal to the potential demand: $x_i = x$ for all i.

Thus in equilibrium markets that open are cleared. Note that we may describe sellers in this model as "contingent price takers". They assume that they can sell any amount at the price P_i if market i opens. Note also that production in each period is $Zx - I$, where I is the beginning of period inventories. In equilibrium production is strictly positive because some goods are sold in each period and therefore some production is required to keep the available supply at the level Zx. See Eden (2013) for a formal analysis and for the efficiency of the equilibrium outcome.

<u>Empirical implications</u>:

In state s, when exactly s markets open, sx units are sold and $(Z-s)x$ units are carried as inventories to the next period. The maximum amount sold over weeks is: $H = Zx$. The minimum amount sold over weeks is: $L = x$. We can therefore obtain an estimate for $Z = 1/\pi$ by observing the amount sold over a sufficiently long period. Using the maximum weekly amount sold as an estimate of H and the lowest weekly amount sold as an estimate of L, I compute the ratio $HLU = H/L$ and use it as an estimate of Z. (HLU stands for High-Low Units).

To compute the ratio of the highest to lowest price in a typical week, I use (1) to get:

(2) $$P_i = \beta\lambda + (1-\beta)\frac{\lambda}{q_i}$$

Since the probability that all the Z markets will open is $q_Z = \pi$, in any given week the highest price is:

(3) $$P^H = P_Z = \beta\lambda + (1-\beta)\frac{\lambda}{q_Z} = \beta\lambda + (1-\beta)\frac{\lambda}{\pi}$$

Since the probability that the first market will open is 1, the lowest price in any given week is:

(4) $$P^L = P_1 = \lambda$$

Dividing (3) by (4) leads to:

(5) $$HLP = \frac{P^H}{P^L} = \beta + (1-\beta)\frac{1}{\pi}$$

Using HLU as an estimate for $Z = 1/\pi$ leads to: $HLP = \beta + (1-\beta)HLU$ which is equivalent to:

(6) $$HLP - 1 = (1-\beta)(HLU - 1)$$

The left hand side of (6) is the percentage average difference between the highest and the lowest price. This is proportional to the percentage difference between the highest and the lowest sale week. I also use the following log approximation because of measurement issues that will be discussed later.

(6') $$\ln(HLP) = (1-\beta)\ln(HLU)$$

<u>Cost shocks</u>:

I now allow for cost shocks. I assume that at the time the seller makes the production decisions in week t, he knows the unit cost for this period, λ_t, and the distribution of the unit cost next period. The next period's cost is a random variable, $\tilde{\lambda}_{t+1}$, and its expected value is denoted by: $\lambda^e_{t+1} = E(\tilde{\lambda}_{t+1})$. Since a unit of inventories can be used to cut next period's production, the value of inventories is the expected discounted cost in the next period, $\beta\lambda^e_{t+1}$. We can therefore modify the arbitrage condition (1) as follows.

(7) $$q_i P_{it} + (1-q_i)\beta\lambda^e_{t+1} = \lambda_t$$

Using $\psi_t = \dfrac{\lambda^e_{t+1}}{\lambda_t}$, we can write (7) as:

(8) $$P_{it} = \beta_i \psi_t \lambda_t + (1-\beta_i\psi_t)\frac{\lambda_t}{q_i}$$

We can now follow the steps (2) to (6) to get the following relationship.

(9) $$\frac{P^H_{it}}{P^L_{it}} = \beta_i\psi_t + Z_i(1-\beta_i\psi_t)$$

Taking the average of (9) over weeks and assuming that the average of ψ_t over weeks is approximately 1 leads to a relationship that is similar to (6).[3] The required modification is that now we should compute HLP as the average ratio of the highest to lowest price over weeks.

[3] The average ψ is approximately 1 if the cost shocks are iid and small.

Regressions to be estimated:

I now add a classical measurement error term to (6) and (6'):

(10) $$HLP - 1 = (1-\beta)(HLU - 1) + \varepsilon$$

(10') $$\ln(HLP) = (1-\beta)\ln(HLU) + \varepsilon'$$

where ε (ε') is independent of $HLU - 1$ ($\ln[HLU]$) and has zero mean, HLP is computed as the average ratio over weeks and HLU is the ratio of the highest to lowest aggregate sales. Note that HLP is an average of extreme points while HLU uses two extreme points of aggregate sales.

In the UST model price dispersion requires uncertainty about demand and therefore there is no intercept in (10) and (10'). But the hypothesis of zero intercept is rejected by the data suggesting that as in search models, there is price dispersion even in the absence of demand uncertainty. I therefore add to the regression (10') three variables suggested by search theory: The average price, total revenues and the number of stores that sold the good. The average price was used by Pratt et.al (1979) in an earlier study. Sorensen (2000) used the purchase frequency and the average wholesale price. Here I have data only from the sellers' side and I therefore use aggregate revenues to capture the importance of the goods in the buyers' budget (aggregate revenues = aggregate spending). The number of stores that offer the good is related to the number of price offers sampled by the typical buyer; a variable that plays a key role in the Burdett-Judd model. I also use category dummies and size variables to capture the difference in the cost of not selling across products.

I assume that the average (over weeks) of the log difference between the highest and the lowest price for good i, $\ln(HLP_i)$, is described by the following equation.

(11) $$\ln(HLP_i) = b_0 + b_1 \ln(HLU_i) + b_2 \ln(\mathrm{Re}\, v_i) + b_3 \ln(AvP_i) + b_4 (\#Stores_i) + \sum_j d_{ji} CD_{ji} + \sum_j s_{ji} SD_{ji} + e_i$$

where b are parameters, $\ln(\mathrm{Re}\, v)$ is the log of total revenues (over stores and weeks), $\ln(AvP)$ is the log of average price (averaged over stores and weeks), $\#Stores$ is the number of stores that sold the product, CD are category dummies ($CD_j = 1$ if product i belong to category j and $CD_j = 0$ otherwise), SD are category specific normalized size measures and e is an error term. The size variables will be described later. They are included in the regression as a proxy for shelf space and the cost of trade delays.

There is of course the problem of using extreme observations that may be the result of measurement errors. For this reason I ran (11) after replacing the range measures by standard deviations measures of dispersion. But this did not change the main results.

Unit surprise measures

Price surprises are the residuals in a regression of prices on information available to the buyer before he gets to the marketplace, like the identity of the store and the date. Should we attempt to explain the dispersion of price surprises, as in Lach (2002), or the dispersion of actual prices? The answer depends on the underlying model. To illustrate this point, let us consider an extreme case in which all prices are perfectly predicted by the identity of the store. In the Burdett-Judd model this is equivalent to the assumption that all buyers see all prices and this leads to a degenerate price distribution equal to the competitive price. In the UST model buyers see all prices and it does not matter whether they can predict prices ahead of their arrival time. Therefore the UST model is a theory of price dispersion and it is not about the dispersion of price surprises.

Another reason for controlling for "store effects" is that different stores may provide different services. Kaplan and Menzio (2013) find that about one-sixth of price dispersion can be attributed to store-level quality and amenity differences. Unfortunately, in the UST model (and in the Burdett-Judd model) it is difficult to distinguish between a store that is indifferent among all prices in the equilibrium range but consistently chooses to be at the low price range to a store that is in the low price range because it provides low services. It is also hard to believe that the difference between the average price dispersion of milk and the average price dispersion of hot dogs occurs because there is more dispersion in the services provided by stores that sell hot-dogs. In our samples, stores that sell hot dogs also sell milk and it is unlikely that store effects drive the results. I therefore choose not to control for store effect when constructing measures of price dispersion.

The above reasoning does not apply to unit dispersion: If aggregate demand is perfectly predictable then all the stores in the UST model will choose a single price (the Walrasian price). This suggests that we should take out a UPC specific seasonal element in aggregate demand. For example, the demand for cold drinks may be higher during the summer and the demand for hot dogs may be higher in the 4th of July.

To get a cleaner measure of demand uncertainty, I use $U_{i,t-L}$ to denote the aggregate number of units sold from good i in week $t-L$ and ran the following regressions:

(12) $$\ln(U_{it}) = a_i + b_{i52}\ln(U_{i,t-52}) + \varepsilon_{it}$$

(12') $$\ln(U_{it}) = a_i + b_{i52}\ln(U_{i,t-52}) + b_{i1}\ln(U_{i,t-1}) + b_{i2}\ln(U_{i,t-2}) + b_{i3}\ln(U_{i,t-3}) + \varepsilon_{it}$$

Note that in (12) there is only one lag of 52 weeks designed to capture seasonality. In (12') I added the most recent 3 lags. I then look at the difference between the highest and the lowest residuals from the regression and define $HLRU_i = \varepsilon_i^H - \varepsilon_i^L$, where $\varepsilon_i^H = \max_t\{\varepsilon_{it}\}$ is the highest value of the residual in (12) and $\varepsilon_i^L = \min_t\{\varepsilon_{it}\}$ is the lowest value of the residual. I use $HLRU$ (high-low residual unit) as a range measure of demand uncertainty. The residual standard deviation measure of uncertainty, $SDRU_i$, is the standard deviation of ε_i.

3. DATA

I use a large weekly data set from Information Resources, Inc. (IRI). These scanner data contain weekly observations of the revenues from each good and the number of units sold. The data cover 31 categories in 50 different markets and contain both grocery stores and drug stores from several different chains during the years 2001-2007. A full utilization of this huge data set is beyond the scope of this paper. Here I look at the sample of grocery stores in Chicago during the years 2004 and 2005. I identify a product with a Universal Product Code (UPC) and obtain prices by dividing revenues by the number of units sold.

I exclude from the sample store-UPC combinations (cells) with zero revenues in some of the sample's weeks, UPCs that were sold by less than 10 stores and categories with less than 10 UPCs. The first exclusion is applied to get a reliable measure of the number of stores that sold the good. The second is aimed at reliable measures of cross sectional price dispersion, and the last allows for within category comparison and economizes on the number of category dummies and size variables. The result of these exclusion are "semi balanced" samples in which the number of stores vary across UPCs but stores that are in the sample sold the product in all of the sample's weeks. After implementing the exclusions, I get 1084 UPCs for the 2005 sample and 665 UPCs for the 2004 sample. I also use the combined 04-05 sample with 104 weeks. This combined sample has only 324 UPCs because a store-UPC cell is included only if the cell's revenues were positive in all weeks.

3.1 The Week Starting on January 17, 2005

I start with a description of the data for a randomly chosen week: The week starting on January 17, 2005. Looking at a single week provides information about the relationship between the search variables and price dispersion but not about the relationship between demand uncertainty and price dispersion. Nevertheless I start with a description of the within week correlations to get a sense of the data without the above exclusions.

In the chosen week, 8602 UPCs were sold by more than one store. The average ratio (actual ratio - not the log difference) of the highest to lowest price over all UPCs is 1.36 and its standard deviation is 0.47. The highest ratio of $HLP = 15$ occurs in a UPC that is sold by 2 stores. For 94% of the UPCs the ratio HLP is less than 2.

Scatter plot diagrams are sometimes used to get a visual description of the data, but these diagrams are not useful when there are many observations. Here I use shares in totals diagrams that are based on the Lorenz curve. Unlike the Lorenz curve I plot several variables in the same diagram. The following example illustrates.

An example:

There are two groups of 6 individuals. The income distribution is the same in both groups and is described in the second column of Table 1. The age distribution is different. The age distribution of group 1 is in the third column of Table 1 (age 1) and the age distribution of group 2 is in the last column (age 2). In group 1 the correlation between income and age is -1. In group 2 the correlation between income and age is 0.83.

Table 2 uses Table 1 to compute the accumulated shares in income and age. The poorest 17% makes about 5% of the income, the poorest 33% makes 14% of the income and so on. In group 1 the age of the poorest 17% is 29% of the total age (which is the same as total income in this example and is equal to 210). In group 2 the age of the poorest 17% is only 5% of the total age. Figure 1 plots the data in Table 2. The accumulated share of age curve is above the diagonal for group 1. For group 2 it is below the diagonal. It coincides with the Lorenz curve initially and then departs from it. We can see that the correlation is 1 within the poorest 14% (the first two observations) and is 1 within the top 5% (the last two observations) but this perfect correlation is spoiled by the middle two observations (the correlation between age and income in the middle income group is -1). Figure 1 thus provides more information than the correlation coefficients.[4]

[4] Yitzhaki (2003) argues that Gini related measures of variability and correlations may be superior to standard measures when the distribution is not normal. Here I use a Lorenz curve type graph as a substitute for scatter diagrams and use standard statistics.

Table 1*: An income by age example

	Income	age 1	age 2
1	10	60	10
2	20	50	20
3	30	40	40
4	40	30	50
5	50	20	30
6	60	10	60

* The first column is a serial number, the second is income, the third is the age in group 1 and the last is the age of group 2.

Table 2*: The accumulated shares (example)

Fraction	acc. Income	acc. age 1	acc. age 2
0.17	0.05	0.29	0.05
0.33	0.14	0.52	0.14
0.5	0.29	0.71	0.33
0.67	0.48	0.86	0.57
0.83	0.71	0.95	0.71
1	1	1	1

*The first column is the fraction of the population, the second is the share of income that is made by the fraction in column 1 (thus for example a third of the population makes 14% of total income). The third column is the share in total age in group 1 (thus for example the poorest third accounts for 50% of the total age) and the last column is the share in total age of group 2.

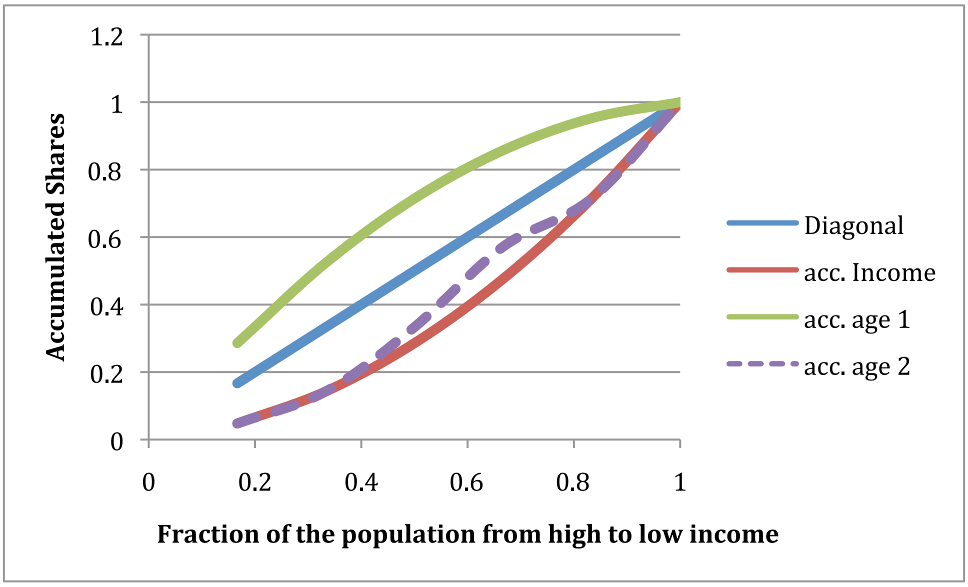


Figure 1: A plot of the data in Table 2 (example)

UPCs instead of people:

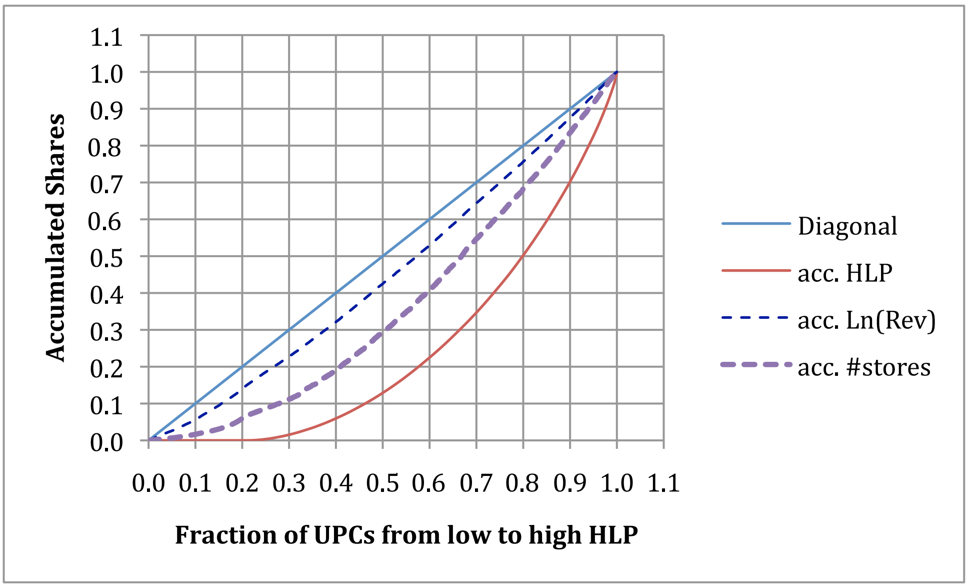
I now turn to the IRI data for the randomly chosen week. In Figure 2, the graph "acc. HLP" is analogous to a Lorenz curve where UPCs play the role of people. The UPCs are ordered by price dispersion from low to high. The graph denoted by "acc. HLP" is the sum: $\sum_{i=0}^{m} NHPL_i$ where m varies from zero (the UPC with the lowest ratio) to one (the UPC with the highest ratio) and $NHLP_i = \frac{\ln(HLP_i)}{\sum_j \ln(HLP_j)}$ is the normalized $\ln(HLP)$.

The graph "acc. HLP" indicate a substantial dispersion in (the log of) HLP across UPCs. In Figure 2A, the share of the lowest 20% is zero and the share of the highest 20% is 50%. (This is analogous to the statement that the share of the poorest 20% in national income is zero and the share of the top 20% is 50%).

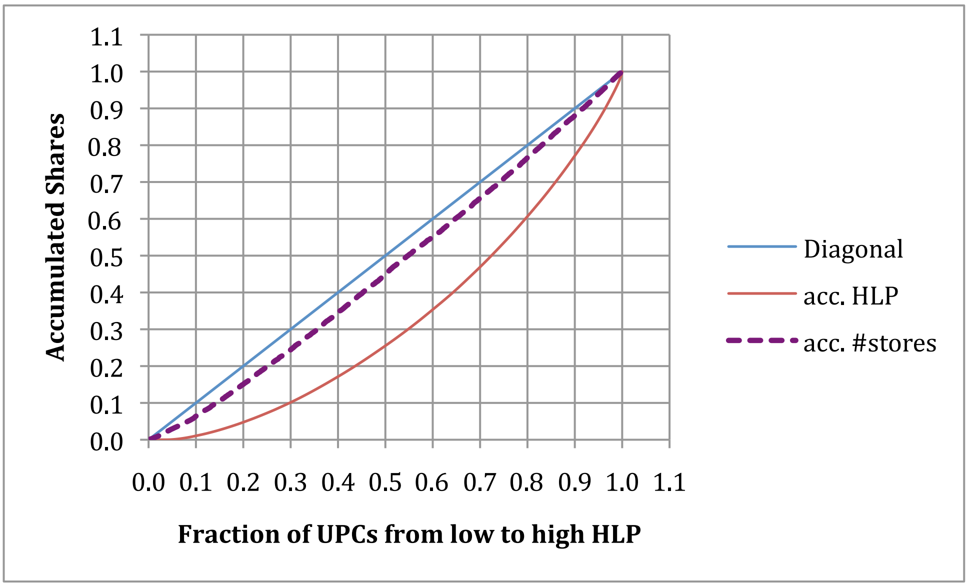
The graph "acc. #stores" describes the accumulated share in the total number of UPC-store cells and is analogous to the acc.age graph in Figure 1. The fact that it is on the right of the diagonal suggests a positive relationship between the number of stores and price dispersion. Indeed, the standard correlation between $\ln(HLP)$ and the number of stores is 0.53. To get more information we may compare the slopes of the two curves. We see that the slope of the "acc. HLP" graph is initially zero and then increases gradually. The slope of the "acc. #stores" graph is also increasing gradually. This suggests a positive correlation within most segments of the UPC population. This is apparent when computing the following conditional averages. The average ratio of high to low price is 1.11 for UPCs sold by 2 stores, 1.18 for UPCs sold by less than 10 stores and 1.52 for UPCs sold by more than 10 stores.

The "acc. Ln(Rev)" graph is the cumulative share in total revenues. This graph is also to the right of the diagonal suggesting a positive relationship between price dispersion and revenues. The share of UPCs with less than the median amount of price dispersion in total revenues is 40%. This is more than the share in the total number of stores (30%) but still the curve is below the diagonal. Also here the slope of the curve increases gradually as in the "acc. HLP" curve suggesting that the correlation occurs within most segments of the UPC population. The standard correlation between log HLP and log revenues is 0.46.

Figure 2B describes the subsample of 4537 UPCs that were sold by more than 10 stores during the week of January 17, 2005. Here 90% of the UPCs have a ratio of high to low price below 2. The maximum ratio is 10 and is less than the maximum ratio of 15 in the larger sample. The "acc. HLP" graph shows that the fraction of UPCs with no price dispersion is small. The "acc. #stores" curve is to the right of the diagonal suggesting a positive correlation between HLP and the number of stores. The standard correlation between the number of stores and $\ln(HLP)$ is 0.33 and is less than in the larger sample. Unlike Figure 2A here I did not plot the "acc.Ln(Rev)" graph because it was too close to the diagonal. But the correlation between ln(Rev) and $\ln(HLP)$ is still positive and is equal to 0.25. Comparing the graphs in Figures 2A and 2B suggest more variability of ln(HLP) in the larger sample. The difference in the standard deviation is however small. The standard deviation of $\ln(HLP)$ is 0.26 in the large sample of 8602 UPCs and 0.25 in the smaller sample of 4537 UPCs. The Gini coefficient in the larger sample is 0.26 while it is 0.17 in the smaller sample.



A. The Sample of 8602 UPCs sold by more than 1 store.



B. The sample of 4537 UPCs sold by more than 10 stores

Figure 2: Cumulative shares for the week starting January 17, 2005. UPCs are ordered from low to high price dispersion (HLP)

3.2 Applying the exclusions and the construction of the main variables

To construct measures of aggregate demand uncertainty, I use the 3 samples described above: The 2005 sample with 1084 UPCs, the 2004 sample with 665 UPCs and the combined 04-05 sample with 324 UPCs.

To economize on space I provide summary statistics in Table 3 only for the largest 2005 sample. The first column is the category name. The second is the number of UPCs in each category. There are for example, 56 UPCs in the beer category. The third is the average (maximum, minimum) number of stores per UPC. The average number of stores in the beer category is 21, the maximum number of stores is 35 and the minimum number of stores is 11. The next four columns provide the averages of the main variables.

The columns ln(HLU) and SDU are unit dispersion measures used as proxies for aggregate demand uncertainty. With the risk of repetition I now describe the construction of the main variables in detail. The variable HLU_i is constructed as follows. I use U_{it} to denote the aggregate amount (over all stores) of UPC i sold in week t, $H_i = \max_t\{U_{it}\}$ to denote the maximum weekly amount sold during the year (or during the sample period when the combined sample of 2 years is used) and $L_i = \min_t\{U_{it}\}$ to denote the minimum weekly amount sold during the year. $HLU_i = H_i / L_i$ is the ratio between the amount sold in the highest sale week and the lowest sale week. The fourth column in Table 3 is the average of the log of this variable, $\ln(HLU)$, over the UPCs in the category. For beer the average log difference is 1.01 implying an average ratio between the highest and the lowest week of HLU = 2.73.

To construct the variable SDU let SDU_i denotes the standard deviation of $\ln(U_{it})$ over weeks. Column 5 is the average of SDU_i over the UPCs in the category. For beer the average is 0.25.

The columns ln(HLP) and SDP are price dispersion measures.

The variable HLP is constructed as follows. Let P_{it}^H (P_{it}^L) denote the highest (lowest) price of UPC i in week t. $HLP_{it} = P_{it}^H / P_{it}^L$ is the ratio in week t and $\ln(HLP_i)$, is the average of the log of this ratio over 52 weeks. The average reported in column 6 is over all the UPCs in the category.

The variable SDP was constructed as follows. Let P_{its} denote the price of UPC i in week t store s and SDP_{it} denote the standard deviation of $\ln(P_{its})$ over stores. The variable SDP_i is the average of SDP_{it} over weeks. In column 7 we have the average of SDP_i over the UPCs in the category. For beer the average standard deviation is 0.06.

I also attempted to include proxies for the cost of not selling $(1-\beta_i)$ that is the proportionality constant in (6). As was said above $1-\beta$ represents the cost of delaying revenues, storage cost and depreciation. Ideally we would therefore like to have information on the shelf life of each UPC and the shelf space that it takes. It also matters whether the good needs to be refrigerated or not. In the data there is only a size measure that may serve as a proxy for "shelf space". But the size measures are not comparable across categories. They are in terms of a fraction of a "regular pack" and the size of a "regular pack" is sometimes in units of volume (for example, rolls for toilet paper) sometimes in terms of square feet (100 square feet is the regular pack for paper towel) and sometimes in units of weights (the regular pack of beer is 288 oz). For this reason I constructed 18 "size dummy" variables. The "size dummy" for beer was constructed as follows. First I normalized the size of all the 56 UPCs in the beer category so that the largest size is 1. I then assigned the value of zero to UPCs that are not in the beer category and the normalized beer size to UPCs within the beer category. Similar treatment was applied to other categories. The last column in Table 3 is the average normalized size. The maximum is 1 by construction. The minimum normalized size is in parentheses. For example, the average size in the beer category is 0.47 implying that on average the size of a UPC is about half the size of the largest UPC in the category.

The last row in Table 3 is the average per category. On average, a category has 60 UPCs and 20 stores per UPC.

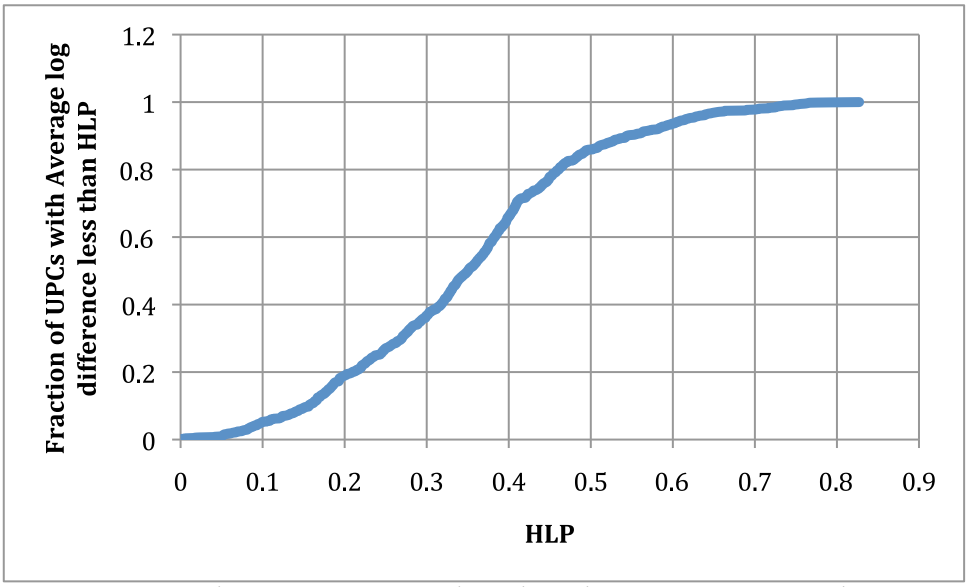
I use HLU and SDU as proxies for demand uncertainty and HLP and SDP as proxies of price dispersion. As can be seen there is substantial variations in these measures across categories. The lowest HLU is for milk (ln[HLU] = 0.78) implying that for an average UPC in the milk category the aggregate (over stores) amount of milk sold in the highest sale week is 2.18 higher than the aggregate amount sold in the lowest sale week. The highest HLU is for hot dogs (ln[HLU] = 2.36) implying that for an average UPC in this category, the aggregate amount of hotdogs sold in the highest sale week is 10.6 times the aggregate amount sold in the lowest sale week.

Table 3*: Summary Statistics for the 2005 sample

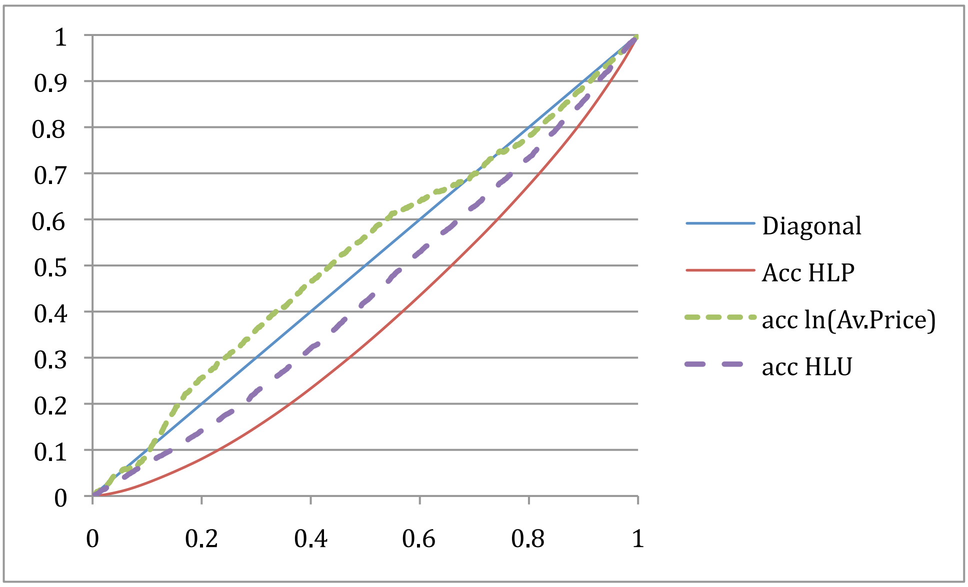
	# UPC	# stores Avg (max,min)	ln(HLU)	SDU	ln(HLP)	SDP	Av. Size (Min)
paper towels	19	20(31,11)	0.95	0.21	0.15	0.05	0.31(0.13)
beer	56	21(35,11)	1.01	0.25	0.19	0.06	0.46(0.07)
facial tissue	18	18(26,11)	1.54	0.38	0.24	0.08	0.29(0.1)
frozen dinners/entrees	75	16(28,11)	1.61	0.36	0.32	0.1	0.62(0.41)
milk	64	22(34,11)	0.78	0.16	0.32	0.1	0.50(0.13)
mustard & ketchup	21	20(32,11)	1.59	0.36	0.33	0.1	0.32(0.13)
salty snacks	120	22(35,11)	1.26	0.3	0.3	0.1	0.47(0.16)
toilet tissue	19	21(34,11)	1.51	0.35	0.32	0.1	0.32(0.04)
frozen pizza	53	18(29,11)	1.49	0.32	0.36	0.11	0.52(0.18)
peanut butter	24	21(31,14)	1.3	0.26	0.34	0.11	0.61(0.30)
yogurt	152	23(35,11)	1.16	0.26	0.31	0.11	0.36(0.13)
carbonated beverages	144	23(35,11)	1.55	0.37	0.37	0.12	0.38(0.04)
mayonnaise	19	23(32,11)	1.29	0.3	0.39	0.12	0.63(0.25)
soup	74	19(35,11)	2.06	0.49	0.39	0.12	0.51(0.40)
spaghetti/Italian sauce	32	16(29,11)	1.37	0.31	0.38	0.13	0.55(0.29)
cold cereal	133	21(34,11)	2.03	0.49	0.45	0.15	0.59(0.21)
margarine/butter	40	25(35,11)	1.22	0.27	0.49	0.15	0.37(0.17)
hotdog	21	20(34,11)	2.36	0.56	0.43	0.16	0.96(0.75)
total	1084						
average	60.2	20(32,11)	1.45	0.33	0.34	0.11	0.49(0.22)

* The first column is the category name. The second is the number of UPCs in the category. The third is the average number of stores per UPC in the category (maximum and minimum in parentheses). The next two columns are measure of demand uncertainty and the following two columns are measures of price dispersion. The average (minimum) normalized size is in the last column. Categories are sorted by SDP. The last row is the average across categories. Thus for example, there are on average 60 UPCs per category.

Figure 3A is the cumulative frequency distribution of ln(HLP) in the 2005 sample. The maximum ln(HLP) is about 0.8 implying HLP=2.2. Recall that HLP is the average ratio of weeks and therefore the maximum HLP is much lower than the maximum in the randomly selected week. About 70% of the UPCs have ln(HLP) less than 0.4 (HLP=1.5). Figure 3B describes share in totals where UPCs are ordered (from low to high) by HLP. As can be seen the slopes of the “acc.HLU” curve are similar to the slopes of the “acc.HLP” curve. Consistent with this observation the correlation between ln(HLP) and ln(HLU) is 0.43. The slopes of the “acc.ln(Av.Price)” graph are not similar to the slopes of the “acc.HLP” graph and the correlation between the log of average price and ln(HLP) is -0.07. The correlation between ln(HLP) and the number of stores is 0.34 and the correlation between ln(HLP) and the log of revenues is 0.27. These correlations are similar to the correlations in the week of January 17.



A. The Cumulative Frequency Distribution of the log difference between the highest and the lowest price averaged over weeks (ln[HLP]).



B. Cumulative Shares in totals. UPCs are ordered from low to high price dispersion (HLP)

Figure 3: Price dispersion in the sample of 1084 UPCs sold by more than 10 stores in all the weeks of 2005

The correlations between the main variables in the 3 samples are in Table 4. The correaltions between the price dispersion measures ln(HLP) and SDP and between the unit dispersion measures ln(HLU) and SDU are both very high (in the range 0.95-0.97). The correlation between the price dispersion measures and the unit dispersion measures (HLU&HLP, SDU&HLP, HLU&SDP, SDU&SDP) are in the range of 0.43-0.60.

Table 4*: Correlation between the main variables

2005	ln(HLU)	SDU	ln(HLP)	SDP
ln(HLU)	1.000			
SDU	0.957	1.000		
ln(HLP)	0.431	0.451	1.000	
SDP	0.480	0.499	0.958	1.000
# of UPCs	1084			
2004	ln(HLU)	SDU	ln(HLP)	SDP
ln(HLU)	1.00			
SDU	0.96	1.00		
ln(HLP)	0.56	0.59	1.00	
SDP	0.57	0.60	0.97	1.00
# of UPCs	665			
04-05	ln(HLU)	SDU	ln(HLP)	SDP
ln(HLU)	1.00			
SDU	0.97	1.00		
ln(HLP)	0.47	0.51	1.00	
SDP	0.50	0.53	0.97	1.00
# of UPCs	324			

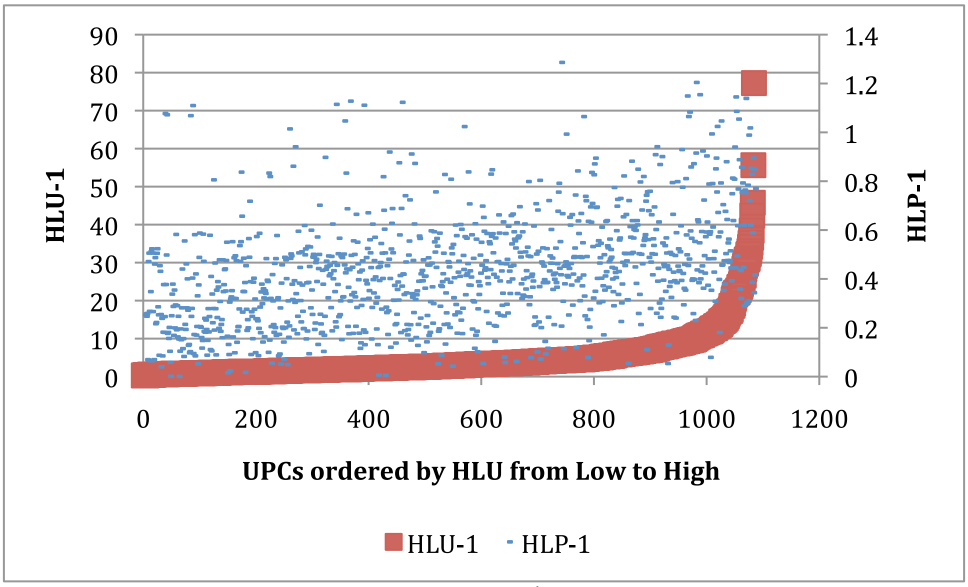
* This Table contains 3 correlation matrices followed by the number of UPCs. The first matrix is for the 2005 sample with 1084 UPCs, the second is for the 2004 sample with 665 UPCs and the last is for the 04-05 sample with 324 UPCs. The variables are the log difference between the highest and lowest weekly aggregate sales ln(HLU), the standard deviation of the log of aggregate sales (SDU), the average log difference between the highest and the lowest price ln(HLP) and the average cross sectional standard deviation of log prices (SDP). See the text for detailed definitions.

3.3 Percentage difference and log difference

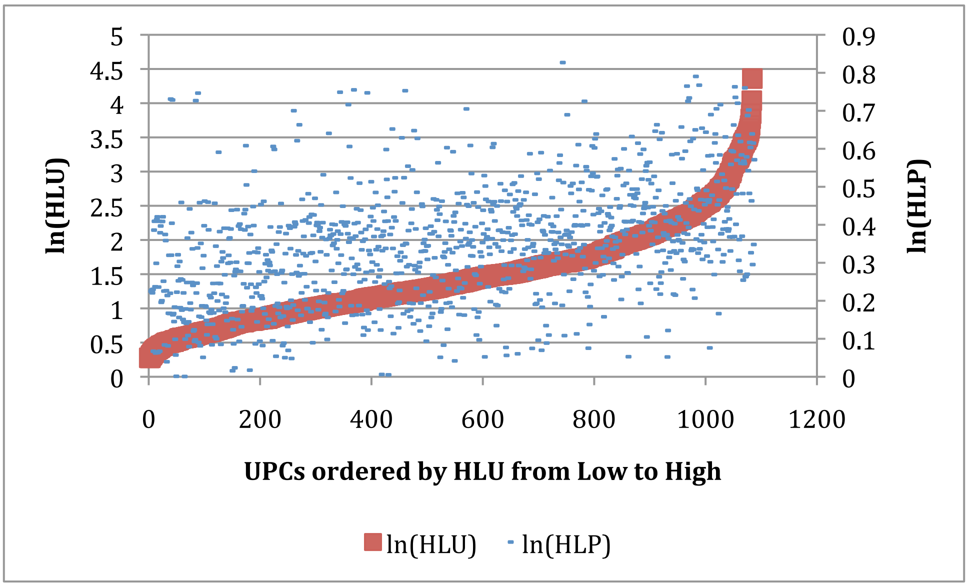
The choice between (6) and (6') is not trivial: Why use an approximation rather than the relationship implied by theory?

Many researchers use the log difference approximation to deemphasize (smooth) outliers. The problem of outliers is evident from Figure 4. Figure 4A is a scatter diagram of the percentage differences HLP-1 and HLU-1. Close to 90% of the UPCs have HLU<10. But there are extreme values of HLU, the highest being HLU=78. Figure 4B describes the log difference approximations ln(HLP) and ln(HLU). The approximation works well for relatively small differences and pull outliers to the bulk of the data. This is especially true for the unit dispersion measure. The range of the variable HLU-1 is 0.32 to 77 while the range of ln(HLU) is 0.28 to

4.36. The correlation between the percentage difference measures HLU-1 and HLP-1 is 0.34 which is lower than the correlation between the log difference measures ln(HLU) and ln(HLP).



A. Percentage difference



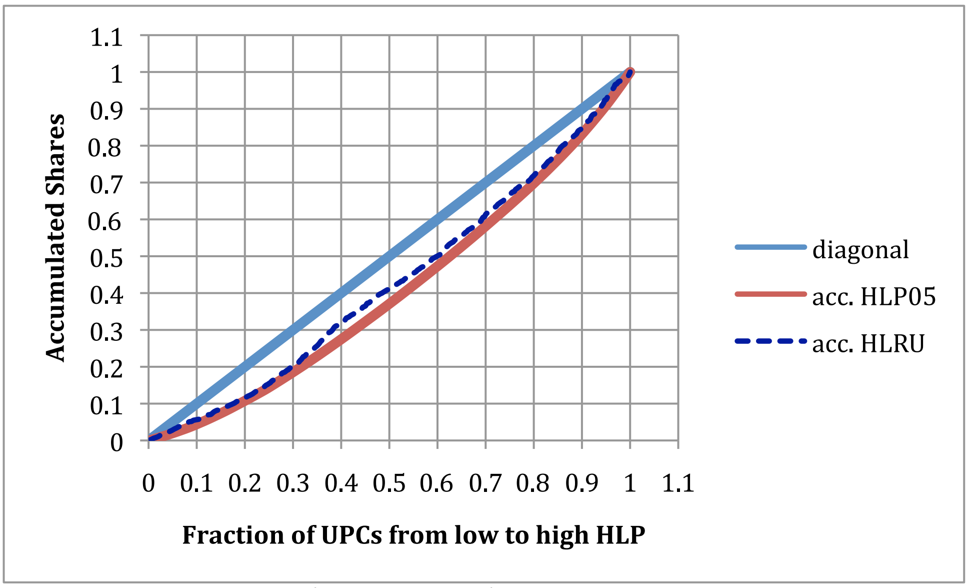
B. Log differences

Figure 4: Percentage and log differences.

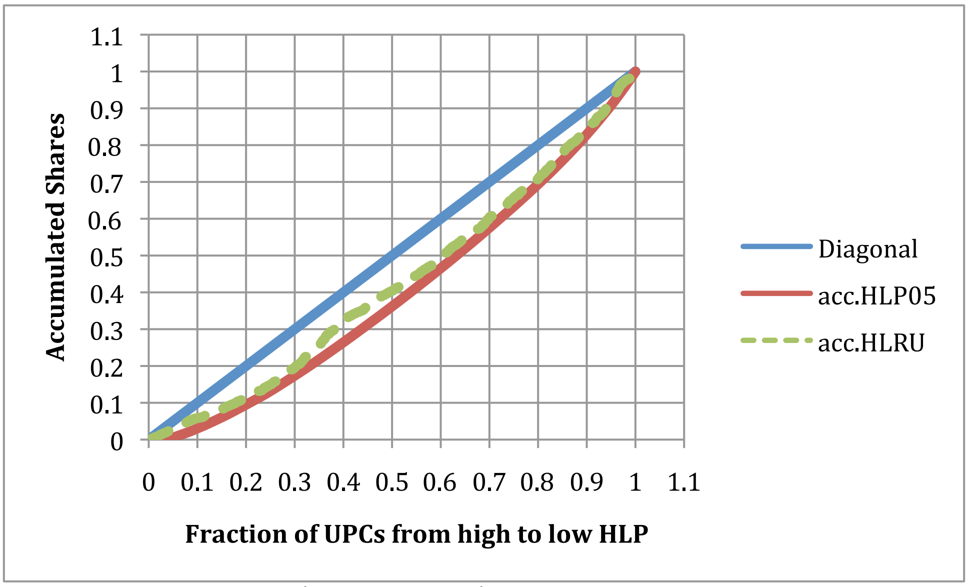
3.4 Unit surprises

I used the combined 04-05 sample with 324 UPCs to run (12)-(12') and get the unit surprise measures $HLRU$ and $SDRU$. I then look at the difference between the highest and the lowest residuals from this regression and define $HLRU_i = \varepsilon_i^H - \varepsilon_i^L$ as the residual range measure of demand uncertainty. The residual standard deviation measure of uncertainty, $SDRU_i$, is the standard deviation of ε_i.

Figure 5A is a shares in totals graph when using the residuals from the regression (12). The two curves are almost on top of each other except for the segment in which the normalized ln(HLP) is between 0.3 to 0.6. The correlation between the two variables is 0.49. The correlation when looking at UPCs with dispersion below the 40th percentile is 0.55. Figure 5B uses the residuals from the regression (12'). The results are almost identical to the results when using (12).



A. Using the residuals from (12)



B. Using the residuals from (12')

Figure 5: Cumulative shares in totals. UPCs are ordered from low to high price dispersion in 2005 (HLP05). HLRU is the residual range measure of unit dispersion

3.5 The ratio of price dispersion to unit dispersion

In the model, the value of inventories is $\beta\lambda$ and we may think of β as the fraction of the wholesale price that the manager of the store will pay for a unit that will be on the shelf for a week with probability 1. This is lower than the fraction that he will pay for a unit that will be delivered in a week because of depreciation and the shelf space cost.

The ratio of price dispersion to unit dispersion is of interest because it provides an unbiased estimate of β. To show this claim, I divide both sides of (10) by $HLU-1$ to get:

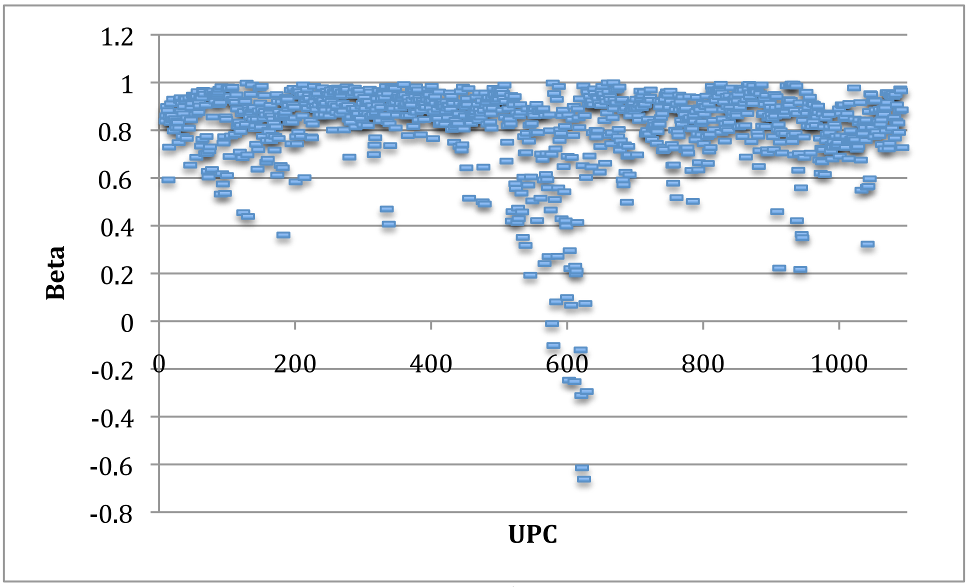
$$\frac{HLP-1}{HLU-1}=1-\beta+\frac{\varepsilon}{HLU-1} \tag{13}$$

Since ε is independent of $HLU-1$ and $E(\varepsilon)=0$ it follows that

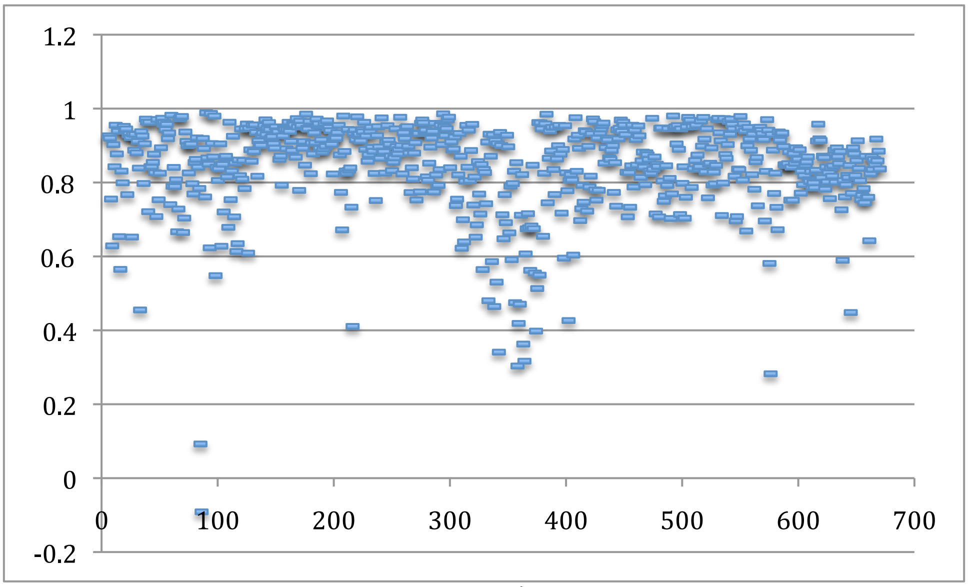
$$(14) \qquad E\left(\frac{HLP-1}{HLU-1}\right)=1-\beta$$

We can therefore obtain an unbiased estimate of $1-\beta$ by computing a simple average of the ratio $\frac{HLP-1}{HLU-1}$ over UPCs.

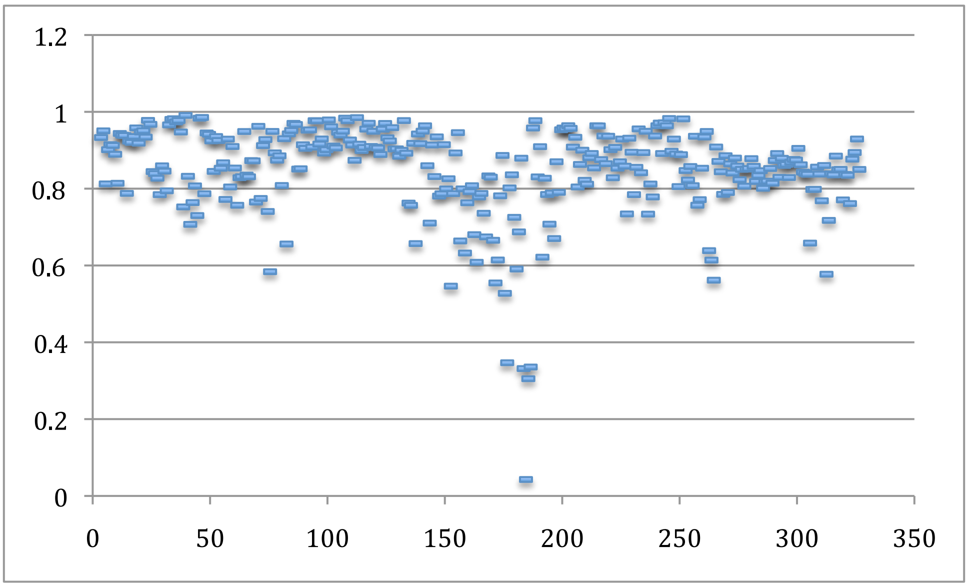
Figure 6 describes the UPC specific estimates of β. The average β across all UPCs (standard deviation in parentheses) is 0.82 (0.18) for the 2005 sample and 0.85 (0.12) for both the 2004 sample and for the combined 04-05 samples. There are no negative β in the combined sample, only one negative β in the 2004 sample. There are 9 negative β in the 2005 sample that are less than 1% of the observations (9 out of the 1084 UPCs). The data thus support the hypothesis that $HLU > HLP$ and that the estimated β is in the unit interval.



A. 2005 sample with 1084 UPCs



B. 2004 sample with 665 UPCs



C. The combined 04-05 sample with 324 UPCs

Figure 6: UPC specific $\beta = 1 - [(HLP - 1)/(HLU - 1)]$

I also used the log approximation (10') to estimate β. The average β across all UPCs (standard deviation in parentheses) is 0.73 (0.17) for the 2005 sample; 0.74 (0.12) for both the 2004 sample and the combined sample. These estimates are lower than the estimates under (10) by about 0.1, possibly because the log transformation smoothes measurement errors and reduces the attenuation bias.

Figure 7 uses the residual from (12') to compute $\beta = 1 - [\ln(HLP05)/HLRU]$. As can be seen the estimated β are smaller with an average of 0.62. There are 9 (out of 324) UPCs with negative β: 5 in the milk category and 4 in the yogurt category.

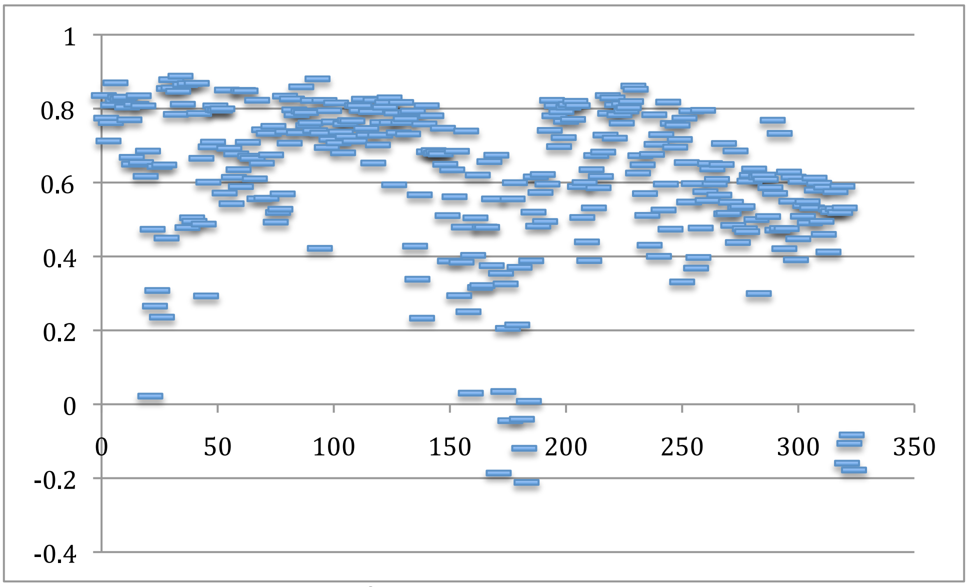


Figure 7: UPC specific $\beta = 1 - [\ln(HLP05)/HLRU]$ where $HLRU$ were computed from the residuals in (12')

Table 5 reports the average β by category. There are 18 categories in the 2005 samples, 17 in the 2004 sample and 10 in the combined sample. The number of UPCs in each category is more than 10 by construction. In the 2005 sample the number of UPCs per category is 60 on average ranging from 18 for facial tissues to 152 for yogurt. In the 2004 sample the average is 39 UPCs per category and the range is 12 to 94. In the combined sample the average is 32 and the range is 11 to 65.

The first 3 columns in Table 5 are based on (10) and the following 3 columns are based on the log approximation (10'). The lowest average β is for milk. The average cost of keeping milk on the shelf for a week is between 54% and 35% of its wholesale price. The other extreme is facial tissues. The average cost of keeping facial tissues on the shelf for a week is between 7% and 18% of the wholesale price.

Table 5*: Average β by category

	$\beta = 1-[(HLP-1)/(HLU-1)]$			$\beta = 1-[\ln(HLP)/\ln(HLU)]$			Using $HLRU$
	2005	2004	04-05	2005	2004	04-05	04-05
beer	0.86	0.84	0.91	0.79	0.78	0.85	0.77
carbbev	0.83	0.82	0.87	0.72	0.71	0.75	0.65
coldcer	0.89	0.90	0.92	0.77	0.77	0.80	0.76
factiss	0.92	0.93		0.84	0.82		
fzdinent	0.90	0.90		0.79	0.77		
fzpizza	0.84	0.87	0.85	0.74	0.75	0.73	0.63
hotdog	0.93	0.91		0.81	0.78		
margbutr	0.69	0.79	0.78	0.57	0.66	0.63	0.52
mayo	0.76	0.75		0.65	0.64		
milk	0.50	0.65	0.63	0.46	0.56	0.55	0.28
mustketc	0.86	0.88		0.76	0.77		
paptowl	0.89			0.84			
peanbutr	0.78	0.81	0.81	0.69	0.70	0.71	0.60
saltsnck	0.84	0.86	0.88	0.75	0.77	0.79	0.70
soup	0.90	0.88	0.89	0.78	0.76	0.78	0.62
spagsauc	0.81	0.86		0.70	0.73		
toitisu	0.84	0.83		0.75	0.71		
yogurt	0.80	0.83	0.83	0.71	0.73	0.71	0.51
All	0.82	0.85	0.85	0.73	0.74	0.74	0.62

* The first 3 columns (after the category name) use (10). The following 3 columns use the log approximation (10'). The last column use $\beta = 1-[\ln(HLP05)/HLRU]$ where $HLRU$ are the residuals of (12'). There are 18 categories in the 2005 sample, 17 in the 2004 sample and 10 in the combined sample.

4. FORECASTING PRICE DISPERSION

In search models price dispersion may arise even in the absence of uncertainty about aggregate demand. I therefore ran (10') with an intercept. Table 6 reports the results for categories with more than 50 UPCs and for the samples as a whole.

Consistent with search models, all the intercepts are positive and significant. In the 2005 sample there are 9 categories with more than 50 observations. 8 out of the 9 coefficients of ln(HLU) are positive and 6 out of the 8 are significant. In the 2004 sample there are 4 such categories. 3 out of the 4 coefficients are significant and positive. In the 04-05 sample there are 3 categories all the coefficients are positive and 2 are significant. The estimates do not change much when we replace ln(HLU) in the 04-05 sample with HLRU. When using all the observations in the samples, the coefficients of ln(HLU) are around 0.1. It is no longer clear that we can interpret the coefficient of ln(HLU) as an estimate of $1-\beta$. But the elasticity itself is of interest.

Table 6*: Running ln(HLP) on ln(HLU) with intercept.

2005 sample	Intercept	ln(HLU)	#UPC	Adj R^2
beer	0.165***	0.023	56	0.005
carbbev	0.308***	0.040**	144	0.049
coldcer	0.190***	0.127***	133	0.321
fzdinent	0.217***	0.063*	75	0.044
fzpizza	0.247***	0.074**	53	0.141
milk	0.343***	-0.024	64	-0.012
saltsnck	0.059*	0.194***	120	0.454
soup	0.311***	0.040*	74	0.059
yogurt	0.283***	0.027	152	0.001
All	0.209***	0.095***	1084	0.185
2004 sample	Intercept	ln(HLU)	#UPC	Adj R^2
carbbev	0.411***	-0.005	86	-0.011
coldcer	0.151***	0.149***	93	0.561
saltsnck	0.107***	0.138***	94	0.457
yogurt	0.229***	0.068*	92	0.060
All	0.207***	0.106***	665	0.318
04-05 sample	Intercept	ln(HLU)	#UPC	Adj R^2
carbbev	0.346***	0.022	58	0.031
coldcer	0.154**	0.130***	53	0.490
yogurt	0.287***	0.052**	65	0.091
All	0.244***	0.080***	324	0.219
04-05 sample	Intercept	HLRU	#UPC	Adj R^2
carbbev	0.348***	0.030*	58	0.058
coldcer	0.164***	0.147***	53	0.601
yogurt	0.385***	0.002	65	-0.016
All	0.276***	0.088***	324	0.230

* One star (*) denotes p-value of 5%, two stars (**) denote p-value of 1% and three stars (***) denote p-value of 0.1%. The first 10 rows are the results when using the 2005 sample. The following 5 rows are the results when using the 2004 sample and the last 4 rows are the results when using the 04-05 sample.

Motivated by the finding of a positive intercept, I adopt the more eclectic approach in (11) that uses variables suggested by search theories. I use two measures of dispersion: the range dispersion measures (HLP, HLU and HLRU) and the standard deviation dispersion measures (SDP, SDU and SDRU). The qualitative results are the same for both measures. The regressions that use the standard deviation dispersion measures are reported in the Appendix.

Table 7 reports the results of running the price dispersion measure ln(HLP) on category dummies, “size dummies” and various combinations of the following main variables: The unit range dispersion measure ln(HLU), revenues, the number of stores and the average price. Only the coefficients of the main variables are reported.

The first 5 rows in the Table describe the regression results when using the 1084 observations in the 2005 sample. The regression reported in Column 1 uses only the unit dispersion measures ln(HLU), intercept, category dummies and size variables. As can be seen the coefficient 0.082 is highly significant. This coefficient does not change much when we add other explanatory variables in columns 2-6 and it is in the range 0.078 - 0.094. The coefficient when running (10’) with intercept reported in Table 6 is 0.095 suggesting that the estimated elasticity is not sensitive to the addition of the other variables.

The coefficient of the average price is also consistently significant and it is in the range of -0.089 to -0.55. The coefficients of revenues are positive but not always significant. The coefficients of the number of stores are positive and significant.

The next 5 rows describe the regression results when using the 665 observations in the 2004 sample. Also here the coefficients of the unit dispersion measure are highly significant and stable. The range of the estimated elasticity is 0.097-0.105 and is slightly higher than the range in the 2005 sample. The elasticity reported in Table 6 is 0.106 suggesting that adding the variables does not change the estimated elasticity by much.

The coefficients of the average price in the 2004 sample are significantly negative and are in the range (-0.062 to -0.055). The coefficients of revenues and the number of stores are positive but not always significant.

The last five rows reports the regression results when using the combined 04-05 sample with 104 weeks and 324 UPCs. The coefficients of the unit dispersion measure are in the range (0.078 - 0.089) that is similar to the range in the 2005 sample and slightly less than the range in the 2004 sample. The coefficients of the average price are in the range (-0.142 to -0.103) that is lower than the range in the previous two samples. The coefficients of revenues and the number of stores are positive but not always significant.

On the whole, the estimated elasticity of the range dispersion measure with respect to the unit dispersion measure is close to 0.1 and is not sensitive to adding variables to the regression.

Table 7*: The Main Explanatory Variables; Dependent variable = ln(HLP)

2005	1	2	3	4	5	6
ln(HLU)	0.082*** (0.007)	0.082*** (0.007)		0.078*** (0.006)		0.094*** (0.007)
ln(Revenues)			0.077*** (0.005)	0.074*** (0.004)	0.043*** (0.01)	0.005 (0.009)
#Stores					0.004*** (0.001)	0.009*** (0.000)
ln(Av. Price)		-.059*** (0.013)	-.089*** (0.012)	-.089*** (0.012)	-.072*** (0.013)	-.055*** (0.012)
Adj. R^2	0.3306	0.3432	0.415	0.4851	0.4228	0.5171
2004	1	2	3	4	5	6
ln(HLU)	0.104*** (0.007)	0.105*** (0.007)		0.097*** (0.007)		0.102*** (0.007)
ln(Revenues)			0.052*** (0.007)	0.036*** (0.006)	0.049*** (0.011)	0.008 (0.010)
#Stores					0.001 (0.002)	0.009*** (0.000)
ln(Av. Price)		-.055*** (0.014)	-.061*** (0.015)	-.062*** (0.013)	-.060*** (0.015)	-.056*** (0.013)
Adj. R^2	0.4905	0.5028	0.3746	0.5312	0.3737	0.5393
04-05	1	2	3	4	5	6
ln(HLU)	0.089*** (0.009)	0.083*** (0.009)		0.078*** (0.009)		0.083*** (0.009)
ln(Revenues)			0.040*** (0.007)	0.031*** (0.007)	0.034** (0.013)	0.007 (0.012)
#Stores					0.002 (0.003)	0.008* (0.003)
ln(Av. Price)		-.111*** (0.021)	-.142*** (0.023)	-.119*** (0.021)	-.139*** (0.024)	-.103*** (0.022)
Adj. R^2	0.5351	0.5721	0.4874	0.594	0.4863	0.6015

* This Table reports the results of 6 regressions in 3 different samples. The samples are 2005, 2004 and the combined sample of 04-05. The first column is the name of the explanatory variables. The 6 regressions include different combinations of the explanatory variables. Each column reports the coefficients of a different regression. Standard errors are in parentheses. The dependent variable in all 6 regressions is the average log difference between the highest and the lowest price. All 6 regressions have category dummies (17 + intercept) and 18 size variables. One star (*) denotes p-value of 5%, two stars (**) denote p-value of 1% and three stars (***) denote p-value of 0.1%. The main explanatory variable in regression 1 is the log difference between the aggregate number of units sold in the week of highest sales and the week of lowest sales (HLU). Regression 2 adds the average log of the price. Regression 3 replaces HLU with the log of total revenues. Regression 4 has both HLU and revenues. Regression 5 replaces HLU with the number of stores and regression 6 uses all the variables.

Table 8 reports the regression results when running (11) for each category with more than 50 UPCs. As in Table 6, there were 9 such categories in the 2005 sample, 4 in the 2004 sample and 3 in the combined 04-05 sample. The first row in the Table reports the regression result when using the sample of 56 UPCs in the 2005 beer category. The coefficient of ln(HLU) is positive for all the 9 categories in the 2005 sample, all the 4 categories in the 2004 sample and for 2 out of the 3 categories in the combined sample. The coefficient of ln(HLU) is significant and positive in 12 out of the 16 regressions and the single negative coefficient is not significant. On the whole, the category regressions in Table 8 provide strong support for a positive ln(HLU) coefficient, a somewhat weaker support for a negative average price coefficient and even weaker support for a positive revenues and number of stores coefficients. The results with respect to the size variables are mixed.

Table 8*: Separate regressions for selected categories; dependent variable = ln(HLP)

2005	ln(HLU)	ln(Rev)	# stores	ln(Av.P)	Size	#UPC	Ad. R^2
beer	0.046*	0.025	-0.001	-.105***	0.068	56	0.34
carbbev	0.073***	-.076***	0.015***	0.025	0.060	144	0.29
coldcer	0.121***	0.097**	0.006	-.244***	0.059	133	0.678
fzdinent	0.056*	0.165***	-0.008	-0.009	-0.038	75	0.4116
fzpizza	0.064**	0.037	0.003	-0.155*	0.044	53	0.3854
milk	0.076*	0.040	0.013***	-0.134*	0.259	64	0.563
saltsnck	0.186***	0.004	0.006	-0.068	0.014	120	0.509
soup	0.026	0.001	0.008	-0.001	0.328***	74	0.2425
yogurt	0.092***	0.034	0.009***	0.021	-0.035	152	0.647
2004	ln(HLU)	ln(Rev)	# stores	ln(Av.P)	Size	#UPC	Ad. R^2
carbbev	0.048	-0.001	-0.002	-0.073	-0.030	86	0.074
coldcer	0.102***	0.158***	-0.003	-.258***	0.142*	93	0.7857
saltsnck	0.128***	-0.004	0.016*	0.008	0.006	94	0.5487
yogurt	0.006	0.063***	-0.003*	-.058***	-.145***	92	0.8483
04-05	ln(HLU)	ln(Rev)	# stores	ln(Av.P)	Size	#UPC	Ad. R^2
carbbev	0.069***	-0.058*	0.017*	-0.034	0.094	58	0.2754
coldcer	0.111***	0.118**	0.009	-0.180*	0.027	53	0.7437
yogurt	-0.011	0.035***	0.001	-.110***	0.081	65	0.8142

*This Table reports the results of a regression that was run for each category separately in 3 different samples. The selected categories have more than 50 UPCs. The first column is the coefficient of the unit dispersion measure HLU, and the following 5 columns are the coefficients of the other explanatory variables.

5. ROBUSTNESS CHECKS

Stores may make mistakes in setting prices. These price-setting errors may affect price dispersion and the right hand side variables of the regression. The problem may not be severe because the dependent variable is the average price dispersion over weeks and in large samples price setting mistakes are zero on average. But here the average is over 52 (104 in the combined sample) weeks and there may still be an endogeneity problem. To address this issue I use the combined sample with 104 weeks and compute the independent variables on the basis of the first 52 weeks and the dependent variable on the basis of the last 52 weeks. The results in Table 9 are similar to the results in Table 7 for the combined sample suggesting that endogeneity is not important.

Table 9*: Dependent variable = ln(HLP.05)

	1	2	3	4	5	6
ln(HLU.04)	0.102*** (0.011)	0.095*** (0.011)		0.088*** (0.011)		0.091*** (0.011)
ln(Rev. 04)			0.040*** (0.009)	0.027*** (0.008)	0.035* (0.014)	0.011 (0.013)
#Stores					0.001 (0.004)	0.005 (0.003)
ln(Av. P. 04)		-0.130*** (0.023)	-0.161*** (0.025)	-0.137*** (0.023)	-0.158*** (0.026)	-.127*** (0.024)
Adj. R^2	0.5112	0.557	0.4795	0.5712	0.478	0.5738

* This Table uses the combined 04-05 sample. The dependent variable is based on the last 52 weeks in the sample (in 2005) while the explanatory variables are based on the first 52 weeks (in 2004).

<u>Using the residual unit dispersion measure</u>:

Table 10 replaces the unit dispersion measure in Table 9 with the residual range measure of demand uncertainty that is obtained from running the regressions in (12'). The coefficients of HLRU are very similar to the coefficients of HLU in Table 9 and are in the range of 0.103 to 0.114. The coefficients of the variables suggested by search theory are also in line with the previous estimates.

Table 10*: Dependent variable = ln(HLP.05)

HLRU	0.114*** (0.010)	0.107*** (0.010)	0.103*** (0.010)	0.105*** (0.010)
Ln (Rev. 04)			0.031*** (0.008)	0.012 (0.012)
#Stores				0.006* (0.003)
Ln (Av. P. 04)		-.121*** (0.022)	-.128*** (0.022)	-.117*** (0.022)
Adj. R^2	0.5631	0.6016	0.6207	0.6245

* This Table reports the results of 4 regressions using the combined 04-05 sample. The dependent variable is the range price dispersion measure that is computed on the basis of the last 52 weeks of the sample. The explanatory variable HLRU is the residual unit dispersion measure obtained from (12').

The computation of Ln (Av. Price) and Ln (Revenues) are computed on the basis of the first 52 weeks in the sample.

Specification search

The specification (11) says that price dispersion is increasing in the ratio of the amount sold in the highest sale week to the amount sold in the lowest sale week. A more general specification may assume that price dispersion is increasing in the amount sold in highest sale week and decreasing in the amount sold in the lowest sale week. We can thus generalize (11) as follows.

(13)
$$\ln(HLP_i) = b_{0i} + b_{1i}^H \ln(H_i) - b_{1i}^L \ln(L_i) + ... +$$
$$= b_{0i} + b_{1i}^H \ln(HLU_i) + (b_{1i}^H - b_{1i}^L)\ln(L_i) + ... +$$

The specification (11) is a special case of (13) that assumes: $b_{1i}^H = b_{1i}^L$. Table 11 provides the results when running (13). The coefficient of $\ln(L_i)$ is not significantly different from zero, thus supporting the specification (11).

Table 11*: Dependent variable = ln(HLP)

	ln(HLU)	ln(L)	ln(Rev)	#Stores	ln(Av.Price)	Ad R^2
2005	.110***	.042	-.035	.008***	-.010	.5159
2004	.085***	-.054	.06	.008***	-.109***	.5409
04-05	.072***	-.032	.037	.008**	-.134**	.6011

* These are the regression results when adding the variable ln(L) to the regression (11), where L is the amount sold in the lowest sale week. The first row is the regression results for the 2005 sample, the second row uses the 2004 sample and the third uses the combined 04-05 sample.

6. CONCLUDING REMARKS

Price dispersion is increasing in proxies of aggregate demand uncertainty. When running price dispersion measures on unit dispersion measures and other variables, the coefficients of the unit dispersion measures are positive and highly significant. This finding supports Prescott type models. But the finding of a significant positive intercept suggests that the model does not capture all the relevant aspects. I therefore include in the regressions three variables suggested by search theory: The number of stores that sell the good, total revenues from selling the good and the average price of the good.

The inclusion of the search variables does not change the estimated elasticity of price dispersion with respect to unit dispersion by much. The estimated elasticity is about 0.1 when using the range measures of dispersion, and about 0.15 when using the standard deviation measures of dispersion.

Out of the variables suggested by search theory the average price is the only one with a stable and significant effect. This is in line with the findings of Pratt et. al. (1979). A possible explanation for the significant negative effect of the average price may rely on the distinction between informed and uninformed buyers as in Salop and Stiglitz (1977), Shilony (1977) and Varian (1980). Roughly speaking, this literature assumes that there are some buyers who pay attention to prices and some who do not. Maybe there are some buyers who in the spirit of Mankiw and Reis (2007) pay attention to high prices but do not pay attention to low prices. This interpretation is not without problems. If this was true than we should find a negative relationship between the size variables and price dispersion. But the coefficients of the size variables are not negative (see Table 8).

REFERENCES

Baharad Eyal and Benjamin Eden (2004). "**Price Rigidity and Price Dispersion: Evidence from Micro Data**," Review of Economic Dynamics, Elsevier for the Society for Economic Dynamics, vol. 7(3), pages 613-641, July.

Bental, Benjamin and Benjamin Eden (1993). "**Inventories in a Competitive Environment**," Journal of Political Economy, University of Chicago Press, vol. 101(5), pages 863-86, October.

Burdett, Kenneth and Kenneth Judd (1983). “Equilibrium Price Dispersion.” Econometrica, 51, 955–70.

Dana James D. Jr. "Advance-Purchase Discounts and Price Discrimination in Competitive Markets" Journal of Political Economy, Vol.106, Number 2, April 1998, 395-422.

_______ “Equilibrium Price Dispersion under Demand Uncertainty: The Roles of Costly Capacity and Market Structure” The RAND Journal of Economics, Vol. 30, No. 4 (Winter, 1999), pp. 632-660.

Deneckere Raymond and James Peck., "Dynamic Competition with Random Demand and Costless Search: A Theory of Price Posting". Econometrica, Vol. 80, No. 3 (May, 2012), 1185–1247.

Diamond, Peter A., 1971. "**A model of price adjustment**," Journal of Economic Theory, Elsevier, vol. 3(2), pages 156-168, June.

Eden, Benjamin. "Marginal Cost Pricing When Spot Markets are Complete" Journal of Political Economy, Dec. 1990. Vol. 98, No.6,1293-1306.

_______ "**Inflation and Price Adjustment: An Analysis of Microdata**," Review of Economic Dynamics, Elsevier for the Society for Economic Dynamics, vol. 4(3), pages 607-636, July, 2001.

________"Price dispersion and efficiency", Vanderbilt University Department of Economics Working Papers, VUECON-13-00012, July 2013.

Kaplan Greg and Guido Menzio. “Five Facts About Price Dispersion” mimeo, Very Preliminary, Feb. 18, 2013.

Lach Saul., “Existence and Persistence of Price Dispersion: An Empirical Analysis” The Review of Economics and Statistics, August 2002, 84(3): 433–444.

Mankiw N. Gregory & Ricardo Reis, 2007. "**Sticky Information in General Equilibrium**," Journal of the European Economic Association, MIT Press, vol. 5(2-3), pages 603-613, 04-05.

Pratt, John W.; Wise, David A.; and Zeckhauser, Richard J. “Price Differences in Almost Competitive Markets.” Q.J.E. 93 (May 1979): 189–211.

Prescott, Edward. C., "Efficiency of the Natural Rate" Journal of Political Economy, 83 (Dec. 1975): 1229-1236.

Salop S., and J. Stiglitz, "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion," Rev. Econ. Stud., Oct. 1977, 44, 493-510.

Shilony, Yuval., "Mixed Pricing in Oligopoly," J. Econ. Theory, Apr. 1977, 14, 373-88.

Sorensen, Alan T. "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs" Journal of Political Economy, Vol. 108, No. 4 (August 2000), pp. 833-850.

Varian, H., "A Model of Sales," American Economic Review 70:4 (1980), 651–659.

Yitzhaki Shlomo "Gini's Mean difference: a superior measure of variability for non-normal distributions" Metron - International Journal of Statistics, 2003, vol. LXI, n. 2, pp. 285-316.

APPENDIX: USING THE STANDARD DEVIATION AS A MEASURE OF DISPERSION

This Appendix replaces the range dispersion measures (HLP, HLU) in Tables 7 - 10 with the standard deviation dispersion measures (SDP, SDU).

Table A1*: The Main Explanatory Variables; Dependent variable = SDP

2005	1	2	3	4	5	6
SDU	0.136*** (0.009)	0.136*** (0.009)		0.129*** (0.008)		0.147*** (0.008)
ln(Revenues)			0.018*** (0.002)	0.016*** (0.001)	0.014*** (0.01)	-0.002 (0.003)
#Stores					0.000 (0.000)	0.002*** (0.000)
ln(Av. Price)		-.016*** (0.004)	-.022*** (0.004)	-.022*** (0.004)	-.021*** (0.004)	-.013*** (0.004)
Adj. R^2	0.4203	0.429	0.3751	0.4963	0.3754	0.5179
2004	1	2	3	4	5	6
SDU	0.159*** (0.009)	0.160*** (0.009)		0.151*** (0.009)		0.152*** (0.009)
ln(Revenues)			0.016*** (0.002)	0.008*** (0.002)	0.024*** (0.004)	0.007* (0.003)
#Stores					-0.002** (0.001)	0.000 (0.001)
ln(Av. Price)		-.018*** (0.004)	-.018*** (0.005)	-.019*** (0.004)	-.020*** (0.005)	-.019*** (0.004)
Adj. R^2	0.57	0.5807	0.3964	0.5912	0.402	0.5907
04-05	1	2	3	4	5	6
SDU	0.154*** (0.012)	0.147*** (0.011)		0.143*** (0.012)		0.145*** (0.012)
ln(Revenues)			0.010*** (0.003)	0.004 (0.002)	0.015*** (0.004)	0.001 (0.004)
#Stores					-0.002 (0.001)	0.001 (0.001)
ln(Av. Price)		-.030*** (0.006)	-.043*** (0.008)	-.031*** (0.006)	-.046*** (0.008)	-.029*** (0.007)
Adj. R^2	0.6402	0.6631	0.5002	0.6654	0.5019	0.6651

* See notes to Table 7.

Table A2*: Separate regressions for selected categories; dependent variable = SDP

2005	SDU	ln(Rev)	# stores	ln(Av. P)	Size	Adj. R^2
beer	0.061*	0.012	-0.001	-0.044***	0.023	0.4571
carbbev	0.075**	-0.017*	0.003***	0.012	-0.009	0.1205
coldcer	0.152***	0.029*	0.001	-0.083***	0.020	0.6525
fzdinent	0.135**	0.028	-0.002	0.007	-0.028	0.3403
fzpizza	0.135**	0.004	0.001	-0.037	0.002	0.3608
milk	0.124*	0.003	0.004***	-0.035	0.085	0.3851
saltsnck	0.251***	0.000	0.001	-0.003	-0.015	0.607
soup	0.147***	-0.022	0.002	-0.012	0.092*	0.3697
yogurt	0.084**	0.015	0.002***	0.012*	-0.044**	0.6344
2004	SDU	ln(Rev)	# stores	ln(Av. P)	Size	Adj. R^2
carbbev	0.044	0.011	-0.005	-0.013	-0.069	0.071
coldcer	0.142***	0.054***	-0.003	-0.096***	0.039	0.8089
saltsnck	0.168***	0.008	0.002	-0.002	0.025	0.6056
yogurt	0.016	0.025***	-0.003***	-0.021***	-0.056***	0.8471
04-05	SDU	ln(Rev)	# stores	ln(Av. P)	Size	Adj. R^2
carbbev	0.094***	-0.017*	0.004	-0.004	0.003	0.2116
coldcer	0.186***	0.028*	0.002	-0.049*	-0.004	0.8177
yogurt	-0.016	0.010***	-0.001*	-0.041***	0.028	0.8227

* See notes to Table 8.

Table A3: Dependent variable = SDP.05

05y-04x	1	2	3	4	5	6
SDU. 04	0.147*** (0.013)	0.140*** (0.012)		0.137*** (0.013)		0.138*** (0.013)
ln(Rev. 04)			0.009** (0.003)	0.003 (0.003)	0.013** (0.005)	0.002 (0.004)
#Stores					-0.002 (0.001)	0.000 (0.001)
ln(Av. P. 04)		-.035*** (0.007)	-.046*** (0.008)	-.035*** (0.007)	-.048*** (0.009)	-.035*** (0.007)
Adj. R^2	0.5839	0.6122	0.4654	0.6125	0.4665	0.6114

* See notes to Table 9.

Table A4*: Dependent variable = SDP.05

SDRU	0.170*** (0.013)	0.162*** (0.009)	0.159*** (0.013)	0.161*** (0.013)
Ln (Av. P. 04)		-.033*** (0.007)	-.035*** (0.007)	-.033*** (0.007)
Ln(Rev. 04)			0.005* (0.002)	0.003 (0.004)
#Stores				0.001 (0.001)
Adj. R^2	0.6178	0.6441	0.6476	0.6471

* See notes to Table 10.