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

Vanderbilt University Department of Economics Working Papers 14-00011

Price dispersion and demand uncertainty: Evidence from US scanner data

Benjamin Eden
Vanderbilt University

Abstract

I use the Prescott (1975) hotels model to explain variations in price dispersion across goods sold by supermarkets in Chicago. I extend the theory to accounts for the monopoly power of chains and for non-shoppers. The main empirical finding is that the effect of demand uncertainty on price dispersion is highly significant and quantitatively important: More than 50% of the cross sectional standard deviation of log prices is due to 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.

This is a revised version of a working paper by the same title (VUECON-13-00012, July 2013). I would like to thank PJ Glandon, Maya Eden and Saul Lach for useful comments on an earlier draft and to Vivian Ying Jiang for excellent research assistance
Citation: Benjamin Eden, (2014) ''Price dispersion and demand uncertainty: Evidence from US scanner data'', *Vanderbilt University Department of Economics Working Papers*, VUECON-14-00011.
Contact: Benjamin Eden - ben.eden@vanderbilt.edu.
Submitted: October 03, 2014. **Published:** October 06, 2014.
URL:http://www.accessecon.com/Pubs/VUECON/VUECON-14-00011.pdf

PRICE DISPERSION AND DEMAND UNCERTAINTY: EVIDENCE FROM US SCANNER DATA

Benjamin Eden*

Vanderbilt University

September 2014

I use the Prescott (1975) hotels model to explain variations in price dispersion across goods sold by supermarkets in Chicago. I extend the theory to accounts for the monopoly power of chains and for non-shoppers. The main empirical finding is that the effect of demand uncertainty on price dispersion is highly significant and quantitatively important: More than 50% of the cross sectional standard deviation of log prices is due to 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, Sequential Trade.

JEL Classification: D50, D80, D83, L10

* ben.eden@vanderbilt.edu

This is a revised version of a working paper by the same title (VUECON-13-00012, July 2013). I would like to thank PJ Glandon, Maya Eden and Saul Lach for useful comments on an earlier draft and to Vivian Ying Jiang for excellent research assistance.

1. INTRODUCTION

The fact that a distribution of prices exists for many homogeneous goods is a challenge for the understanding of market-based economies. Among the explanation proposed are price discrimination, search frictions and sticky prices. Here I use scanner data on supermarket prices to test the hypothesis that price dispersion arises as a result of demand uncertainty.

I find that on average, more than 50% of the cross sectional standard deviation of log prices is due to demand uncertainty. This finding lends credibility to Prescott (1975) type models that focus on demand uncertainty as the reason for price dispersion.

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.[2]

I choose versions of the Prescott model because of the focus on uncertainty about aggregate demand.[3] To adapt the model to the market for food I attempt two

---

[2] 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).

[3] In contrast, there is no uncertainty about aggregate demand in search models of price dispersion and therefore getting price dispersion in search models 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

extensions. I consider the case in which there are chains with limited monopoly power. I also attempt a distinction between shoppers and non-shoppers. These extensions do not change the main prediction about the relationship between aggregate demand uncertainty and price dispersion and thus demonstrate its robustness.

Section 2 is about the underlying theory. Section 3 discusses implementation issues. Section 4 describes the data. Section 5 is the estimation results. Section 6 is a robustness checks. Section 7 assesses the quantitative importance of demand uncertainty. Concluding remarks are in the last section.

2. THEORY

I start with a simplified version of Bental and Eden (1993) and derive a relationship between specific measures of price dispersion and specific measures of demand uncertainty.

Sellers

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 of production he does not know the realization.

---

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). For other search models of price dispersion, see Reinganum (1979), Rob (1985) and Stahl (1989).

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 $P_1 < ... < P_Z$.[4]

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:

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

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 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.

[4] 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}.

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 active buyers that arrive in the market place in a typical period ($\tilde{N}$) is an *iid* discrete random variable that may take Z realizations: $0 < N_1 < ... < N_Z$. For notational convenience I use $N_0 = 0$. All realizations occur with equal probabilities: State s occurs when $\tilde{N} = N_s$ with probability $\pi = 1/Z$. The difference between two consecutive realizations is denoted by: $N_i - N_{i-1} = \Delta_i > 0$.

Buyers arrive in a sequential manner. The first batch of Δ_1 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 Δ_2 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.

It is useful to think of 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 hypothetical 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: $x_i = \Delta_i$ 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 $\sum_i x_i - I$, 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 $\sum_i x_i$.

Empirical implications

In state s, when exactly s markets open, $\sum_{i=1}^{s} x_i$ units are sold and $\sum_{i=s+1}^{Z} x_i$ units are carried as inventories to the next period. The maximum amount sold over weeks is: $H = \sum_{i=1}^{Z} x_i = Z\bar{x}$, where $\bar{x}$ is the average supply per market. The minimum amount sold over weeks is: $L = x_1$. 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 $HLU = {}^{H}\!/_{L}$ (HLU stands for High-Low Units) that is proportional to Z:

(2) $$HLU = \frac{Z\bar{x}}{x_1} = Z\alpha \text{ or } Z = \frac{HLU}{\alpha}$$

where $\alpha = {}^{\bar{x}}\!/_{x_1}$ is a constant equal to the ratio of the average supply per market to the first market's supply. Note that $\alpha = 1$ when the number of buyers is uniformly distributed, but may be different from one if the distribution is not uniform.

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

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

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

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

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

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

Dividing (4) by (5) leads to:

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

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

(7) $$HLP - 1 = (1-\beta)\left(\frac{HLU}{\alpha} - 1\right)$$

Using $\ln(HLP)$ as a proxy for $HLP-1$ and $\ln(HLU) - \ln(\alpha)$ as a proxy for $\frac{HLU}{\alpha} - 1$ leads to:

(8) $$\ln(HLP) = -(1-\beta)\ln(\alpha) + (1-\beta)\ln(HLU)$$

Cost shocks

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.

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

Using $\psi_t = \dfrac{\lambda_{t+1}^e}{\lambda_t}$, we can write (9) as:

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

which leads to:

(11) $$\frac{P_{it}^H}{P_{it}^L} = \beta_i \psi_t + Z_i (1-\beta_i \psi_t)$$

Taking the average of (11) over weeks and assuming that the average of ψ_t over weeks is approximately 1 leads to a relationship that is similar to (8).[5] The required modification is that now we should compute HLP as the average ratio of the highest to lowest price over weeks.

2.1 The standard deviation measures

I use both the range measure and the standard deviation measure of dispersion. The relationship between the range measures of dispersion was derived under the assumption that the realizations of demand occur with the same probability (π) but we allowed variation in the number of buyers across batches. Here I assume that the number of buyers is the same across batches and $\Delta_s = x$ for all $s = 1,...,Z$. As a result the fraction of stores that post the price P_s is the same for all s and is given by $\pi = \dfrac{x}{Zx} = \dfrac{1}{Z}$. The probability of making a sale at the price P_s is: $q_s = 1 - \dfrac{s-1}{Z} = 1-(s-1)\pi$. Assuming that storage is not possible and $\beta = 0$, (3) implies:

(12) $$P_s = \frac{\lambda}{q_s} = \frac{\lambda}{1-(s-1)\pi} = \frac{\lambda Z}{Z+1-s}$$

Since λ and Z are constants, the variance of the log of price is:

(13) $$Var(\ln P) = Var\big(\ln(Z+1-s)\big) = Var(\ln s)$$

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

The number of units sold in state s is: $U_s = sx$. Since x is a constant, the variance of the log of units is $Var(\ln s)$ and is equal to (13). Thus in this example there is a perfect correlation between the standard deviation dispersion measures.

The theory up to now was rather abstract, but it is sufficient to describe the empirical findings. A reader who is mostly interested in the empirical findings can therefore jump to the implementation section. But some readers are not comfortable with this level of abstraction and would like to see a model that is more closely related to grocery stores. In an attempt to get a better fit between the model and the industry, I now consider two extensions. In section 2.2, I relax the price taking assumption and allow for the existence of chains each specializing in one of our hypothetical markets. In section 2.3, I distinguish between shoppers and non-shoppers. The relationship between price dispersion and demand uncertainty survives these extensions.

2.2 Chains with monopoly power

Grocery stores typically belong to chains. Here I assume that all stores that post the same price belong to the same chain and the chain has a monopoly power in one of our hypothetical markets. I use the last section in Eden (1990) and the example in Dana (1999) but here the monopoly power is limited to one hypothetical market.

As before, I assume that the number of active buyers $\tilde{N}$ is an iid random variable and that the number of buyers that arrives in batch i is: $N_i - N_{i-1} = \Delta_i$. I assume further that the reservation prices of buyers that arrive in batch i is distributed uniformly on the interval $[0, a_i]$ and at the price $P_i \leq a_i$, a fraction P_i/a of the buyers that arrive in batch i will not buy the good.

I assume that $a_i \geq a_{i-1}$ so that buyers who arrive late have on average higher reservation price. This assumption is not necessary for the main result but it captures some aspects of the distinction between shoppers and non-shoppers in Salop and Stiglitz (1977), Shilony (1977) and Varian (1980). Shoppers who spend more time shopping are more likely to arrive early and get the good at a relatively cheap price. These shoppers are well informed about prices and buy only if the available price is relatively cheap. An extreme example occurs at the beginning of the Christmas shopping season (black Friday) where bargain hunters often wait in line hours before the opening of the stores. A different interpretation of the distinction between shoppers and non-shoppers is in the next section.

The amount sold in hypothetical market i (at the price $P_i \le a_i$) is:

(14) $$x_i = \Delta_i \left(1 - \frac{P_i}{a_i}\right)$$

The inverse demand function in market i is therefore:

(15) $$P_i = a_i - b_i x_i ,$$

where $b_i = {}^{a_i}\!/\!{}_{\Delta_i} > 0$. The chain chooses the supply to hypothetical market i (x_i) by solving the following problem:

(16) $$\max_{x_i} q_i(a_i - b_i x_i)x_i + (1-q_i)\beta\lambda x_i - \lambda x_i$$

Note that the unit cost is λ regardless of whether the unit comes from production or from inventories. The solution to this problem is:

(17) $$x_i = \frac{q_i a_i + (1-q_i)\beta\lambda - \lambda}{2q_i b_i}$$

Substituting (17) in (15) leads to:

(18) $$P_i = (½)\left(a_i + \beta\lambda + \frac{\lambda(1-\beta)}{q_i}\right)$$

Note that since $a_i \ge a_{i-1}$ and $q_i < q_{i-1}$, $P_i > P_{i-1}$.

We can write the problem (16) as:

(19) $$\max_{x_i} (a_i - b_i x_i)x_i - \frac{\big(\lambda - (1-q_i)\beta\lambda\big)x_i}{q_i} = (a_i - b_i x_i)x_i - \frac{\lambda_i^* x_i}{q_i}$$

The term $(a_i - b_i x_i)x_i$ is total revenues. The first order condition for this problem can therefore be written as marginal revenue = "unit cost", where the unit cost term, $\frac{\lambda_i^*}{q_i} = \beta\lambda + \frac{\lambda(1-\beta)}{q_i}$, is increasing in i (because the probability of sale, q_i is decreasing in i). We can therefore describe the solution to the problem (19) using the familiar diagram in Figure 1, where (x_i^M, P_i^M) denote the monopoly's solution for market i. For comparison I also have the competitive solution denoted by (x_i^C, P_i^C).

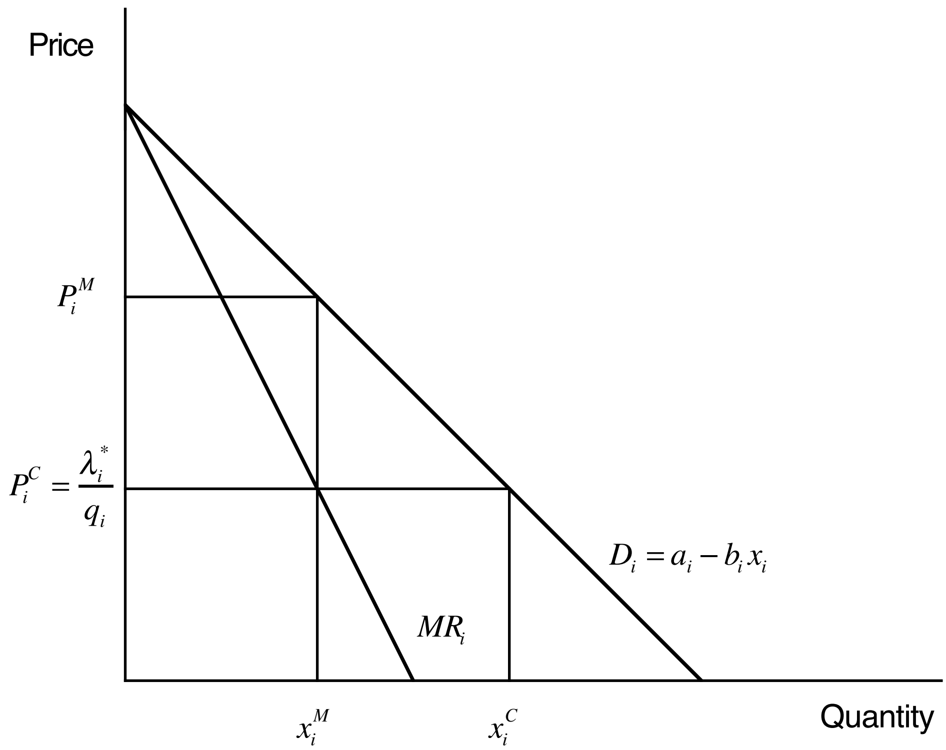


Figure 1: Prices and quantities in market i under monopoly and competition

The lowest price is obtained by substituting $q_1 = 1$ in (18). This leads to:

(20) $$P^{LM} = P_1 = (½)(a_1 + \lambda)$$

Substituting $q_Z = \pi$ in (18) leads to the highest price:

(21) $$P^{HM} = P_Z = (½)\left(a_Z + \beta\lambda + \frac{\lambda(1-\beta)}{\pi}\right) = (½)\left(a_Z + \beta\lambda + \lambda(1-\beta)Z\right)$$

Dividing (21) by (20) leads to:

(22) $$HLP^M = \frac{P^{HM}}{P^{LM}} = \frac{P_Z}{P_1} = \frac{a_Z + \beta\lambda + \lambda(1-\beta)Z}{a_1+\lambda} = \frac{a_Z + \beta\lambda}{a_1+\lambda} + \frac{\lambda(1-\beta)}{a_1+\lambda}Z$$

Substituting (2) in (22) and rearranging leads to:

(23) $$HLP^M - 1 = \frac{a_Z - a_1}{a_1 - \lambda} + \frac{\lambda(1-\beta)}{a_1-\lambda}\left(\frac{HLU^M}{\alpha} - 1\right)$$

Using the log approximation leads to:

(24) $$\ln(HLP^M) = \frac{a_Z - a_1 - \lambda(1-\beta)\ln(\alpha)}{a_1 - \lambda} + \frac{\lambda(1-\beta)}{a_1-\lambda}\ln(HLU^M)$$

Similar to (8) this is a linear relationship between price dispersion and unit dispersion.

Note that we can write (24) as:

(24’) $$\ln(HLP^M) = \frac{1}{(a_1-\lambda)/\lambda}\left\{\frac{a_Z - a_1}{\lambda} + (1-\beta)\ln\left(\frac{HLU^M}{\alpha}\right)\right\}$$

Price dispersion therefore depends on the following three measures:

$(a_1 - \lambda)/\lambda$ is a measure of monopoly power,

$(a_Z - a_1)/\lambda$ is a measure of discrimination power,

HLU^M/α is a measure of demand uncertainty.

Note that price dispersion is decreasing in our measure of monopoly power, increasing in our measure of discrimination power and in our measure of demand uncertainty. There may be a problem in distinguish empirically between monopoly power and discrimination power.[6]

2.3 Shoppers and non-shoppers

The assumption that buyers buy the good at the cheapest available price is of course not realistic. In an attempt to get the model closer to reality I assume now that some buyers (non-shoppers) buy at the most convenient location, say at the supermarket closest to home.

[6] In studies of price dispersion in the airline industry, Borenstein and Rose (1994) found that price dispersion is greater on city-pair routes that are served by a larger number of carriers. Gerardi and Shapiro (2009) argue for the opposite result.

Although non-shoppers do not search across stores the quantity they buy depends on the price. This may be interpreted as search over time. If the price is low the non-shoppers buy a relatively large quantity and store part of it. If the price is high, they buy a relatively small amount and consume out of storage. This is consistent with the findings of Pesendorfer (2002) and Hendel and Nevo (2013) who argue that storage by buyers is important.[7]

As in section 2.1, I require here that all stores will make the same expected profits. This is achieved in equilibrium by varying the number of stores that supply to each market. This is different from the assumption used in section 2.2 where each of the hypothetical market was monopolized by a chain. This difference in assumption is not critical. We can impose a zero expected profit requirement in both cases. But the objective is to demonstrate that the relationship between price dispersion and demand uncertainty is robust.

As before, the total number of buyers that arrive is a random variable $\tilde{N}$ that can take Z possible realizations: $N_1 < N_2 < ... < N_Z$. A fraction ϕ of the buyers in each batch are non-shoppers and a fraction $1-\phi$ of the buyers are shoppers. The demand of each active buyer at the price P is given by the decreasing function $D(P)$.

Shoppers see all price offers and buy at the cheapest available price. Non-shoppers go to one store only and buy at the price posted by the store if the store is not stocked out. Otherwise, if the store is stocked out they go home empty handed. In

[7] A downward sloping demand assumption may also capture storage behavior by shoppers. To see how this may work, consider the case in which a shopper makes it to the first market. Realizing that he is lucky and may not make the first market in the next period, he buys some units for storage and future consumption. Shoppers who are not as lucky and arrive late, buy mostly for current consumption. As a result the demand function of each buyer will depend on the amount of storage he has as well as on some taste parameters. From the point of view of the sellers, there will be buyers of different types where each type is characterized by the probability of becoming active and by the demand function. I attempt such a model recently in Eden (2013). It is more complicated than the model used here but I think that the qualitative results about price dispersion are the same.

a more general model we should allow buyers to buy a close substitute in the case of a stock-out but this is not considered here.

The store chooses the stock-out state. It can choose to stock out only when demand is at its highest possible realization (state Z). It can also choose to stock out at state $s < Z$ with a higher probability. A store that stock out sells its entire supply and therefore a high stock out probability is a benefit. To stock out with high probability the store must post a low price and the tradeoff is therefore between the stock out probability and the price. This is different from the tradeoff we had before (between the probability of making a sale and the price) because now the store will always sell part of its supply to non-shoppers.

As before there are Z hypothetical markets and the prices in these markets are: $P_1 < P_2,...,P_{Z-1} < P_Z$. A store in market i posts the price P_i and stocks out (sells its entire supply) if $s \geq i$. The number of stores in market i is n_i. The total number of stores is: n. The number of stores is a variable that will be varied smoothly and will be treated as a continuous variable.

The number of non-shoppers per store in state s is $\frac{\phi N_s}{n}$ and the amount sold by a store in market i (that posts the price P_i) to the non-shoppers is:

(25) $$\frac{\phi N_s}{n} D(P_i) \text{ if } s \leq i \text{ and zero otherwise.}$$

The amount sold to the shoppers is:

(26) $$\frac{(1-\phi)\Delta_i}{n_i} D(P_i) \text{ if } s \geq i \text{ and zero otherwise.}$$

Since a store in market i plans to stock out in state i, it will produce $\left(\frac{\phi N_i}{n} + \frac{(1-\phi)\Delta_i}{n_i}\right) D(P_i)$ units. As in section 2.2, I assume that storage is not possible ($\beta = 0$). The expected profits for a store that supplies to market i is therefore:

(27) $$V_i = V_i(P_i) = P_i D(P_i)\left\{q_i \frac{(1-\phi)\Delta_i}{n_i} + \sum_{s \leq i} \pi_s \frac{\phi N_s}{n}\right\} - \lambda D(P_i)\left(\frac{\phi N_i}{n} + \frac{(1-\phi)\Delta_i}{n_i}\right)$$

There are m stores that service only the non-shoppers and charge the price p^M that solves the following problem.

(28) $$V^M = \max_{p,j} pD(p)\sum\nolimits_{s\le j}\pi_s \frac{\phi N_s}{n} - \lambda D(p)\frac{\phi N_j}{n}$$

In equilibrium shoppers will always be able to buy at a price that is cheaper than p^M. Thus,

(29) $$P_{i-1} < P_i \le p^M \quad \text{for all } i.$$

The maximum number of buyers serviced by a store in market i is: $Q_i = \frac{\phi N_i}{n} + \frac{(1-\phi)\Delta_i}{n_i}$. A store in market i that chooses a price $P_{i-1} < P \le P_i$ will not affect its stock out probability. I require that in equilibrium such a deviation will not increase expected profits. Thus,

(30) $$V_i(P_i) \ge \max_{Q,P} PD(P)\left\{q_i Q + \sum\nolimits_{s<i}\pi_s \max\left(Q, \frac{\phi N_s}{n}\right)\right\} - \lambda D(P)Q$$

$$\text{s.t. } P_{i-1} < P \le P_i$$

Equilibrium is a non negative vector $(P_1,...,P_Z, p^M; n_1,...,n_Z, m; V, V_M)$ that satisfies (29), (30) and

(a) $n_i > 0$

(b) $V_i = V$ for all $i = 1,...,Z$

(c) $V_M \le V$ with equality if $m > 0$

(d) $\sum_{i=1}^{Z} n_i = n - m$

An example

To get a sense of how this model works, I assume a Unit demand case:

(31) $$D(P) = 1 \text{ if } P \le a \text{ and zero otherwise.}$$

I now show that when a is large there exists an equilibrium in which a large number of stores post the monopoly price and service only the non-shoppers and a small "large" number of stores service the shoppers.

In the proposed equilibrium $V_M = V$, $p^M = a$ and since a is large a store that specializes in the non-shoppers will produce to satisfy the highest realization of demand, $\frac{\phi N_Z}{n}$. In state s the store will sell only $\frac{\phi N_s}{n}$ units and its expected profits are:

(32) $$V = -\lambda \frac{\phi N_Z}{n} + a \sum_s \pi_s \frac{\phi N_s}{n} \geq 0$$

In (32) the first term is the cost of production and the second is the expected revenue.

A store in market Z will sell its entire supply in the highest aggregate demand state, but it will sell to both shoppers and non-shoppers. In state $s < Z$ the store sells $\frac{\phi N_s}{n}$ units to non-shoppers. In state Z it sells $\frac{\phi N_Z}{n}$ to non-shoppers and $\frac{(1-\phi)\Delta_Z}{n_Z}$ to shoppers. The expected profit of a store in market Z is:

(33) $$P_Z \left(\frac{q_Z (1-\phi)\Delta_Z}{n_Z} + \sum_s \pi_s \frac{\phi N_s}{n} \right) - \lambda \left(\frac{\phi N_Z}{n} + \frac{(1-\phi)\Delta_Z}{n_Z} \right) = V$$

The first terms in (33) is the expected revenue and the second is the cost.
We choose (n_Z, P_Z) that satisfy (33) and $P_Z < a$. Equating (33) with (32) leads to:

(34) $$(a - P_Z) \sum_s \pi_s \frac{\phi N_s}{n} = (q_Z P_Z - \lambda) \frac{(1-\phi)\Delta_Z}{n_Z}$$

The left hand side of (34) is the loss of profits from servicing the non-shoppers that a store in market Z will suffer relative to a store that posts the monopoly price a. The right hand side is the gain from servicing the shoppers when demand is at its highest possible realization.

I choose a small n_Z and therefore the quantity per store that is demanded by the last batch of shoppers, $\frac{(1-\phi)\Delta_Z}{n_Z}$, is large. It follows that $q_Z P_Z - \lambda$ must be small. Thus,

(35) $$q_Z P_Z \approx \lambda$$

I now proceed by induction, taking (P_{i+1}, n_{i+1}) as given and assuming $q_{i+1} P_{i+1} \approx \lambda$.

Since a store in market i makes the same expected profits as a store in market $i+1$ we get:

(36)
$$(q_i P_i - \lambda)\frac{(1-\phi)\Delta_i}{n_i} - (q_{i+1}P_{i+1} - \lambda)\frac{(1-\phi)\Delta_{i+1}}{n_{i+1}} =$$
$$(P_{i+1} - P_i)\sum_{s<i+1}\pi_s \frac{\phi N_s}{n} + \pi_{i+1}P_{i+1}\frac{\phi N_{i+1}}{n} - \lambda\frac{\phi\Delta_{i+1}}{n}$$

The left hand side is the change in expected profits from servicing the shoppers. The right hand side is the change in expected profits from servicing the non-shoppers. Since $(P_{i+1} - P_i)\sum_{s<i+1}\pi_s \frac{\phi N_s}{n} + \pi_{i+1}P_{i+1}\frac{\phi N_{i+1}}{n} - \lambda\frac{\phi\Delta_{i+1}}{n} + (q_{i+1}P_{i+1} - \lambda)\frac{(1-\phi)\Delta_{i+1}}{n_{i+1}}$, is finite it follows that if we choose n_i that is small we will get:

(37)
$$q_i P_i \approx \lambda$$

Thus, in the proposed equilibrium prices are close to the case in which all active buyers are shoppers as is typically assumed in UST models.

Sellers that service shoppers are large relative to sellers that specialize in non-shoppers. In the theory the seller is defined by the price: a seller is in market i if he posts the price P_i. One possible interpretation of the above example is that the large stores that service the shoppers are chains that post the same price.

The computation of the prices in the above example also illustrates that the equilibrium is not unique. We have more equations than unknown because we can choose both the number of sellers in each market and the price. I now turn to a prediction that seems more robust.

The relationship between price dispersion and unit dispersion.

There is an equilibrium relationship between price (P), expected capacity utilization (ECU) and expected per unit profits (PUP). To derive this relationship, let $CAP_i = \frac{\phi N_i}{n} + \frac{(1-\phi)\Delta_i}{n_i}$ denote the capacity (amount offered for sale) of a store in market i, $ECU_i = \frac{1}{CAP_i}\left(q_i \frac{(1-\phi)\Delta_i}{n_i} + \sum_{s\le i}\pi_s \frac{\phi N_s}{n}\right)$ denote expected capacity utilization and $PUP_i = \frac{V}{CAP_i}$ denote expected per unit profits. Then the requirement $V_i = V$, leads to:

(38) $$P_i(ECU_i) = \lambda + PUP_i$$

Assuming that profits per unit are small we can use the approximation: $P(ECU) \approx \lambda$ that leads to:

(39) $$\frac{P_Z}{P_1} = \frac{1}{ECU_Z}$$

Assuming that $\pi_s = \pi$ leads to:

$$ECU_Z = \frac{\pi}{CAP_Z}\left(\frac{(1-\phi)\Delta_Z}{n_Z} + \sum_{s \le i} \frac{\phi N_s}{n}\right)$$

Substituting in (39) leads to:

(40) $$\frac{P_Z}{P_1} = \kappa Z,$$

where κ is a constant. This prediction is consistent with (6).

Thus the requirement that expected profits is the same across stores and the assumption of *iid* shocks leads to a relationship between average capacity utilization and price and between price dispersion and demand uncertainty.[8]

3. IMPLEMENTATION

The positive correlation between price dispersion and unit dispersion is the main prediction of the above versions of the Prescott model. Under certain assumptions the relationship is linear as in (8), (13), (24) and (40). I assume her that the relationship is linear and start by adding a classical measurement error to the

[8] Capacity utilization has been used to explain price dispersion in the airline industry. See Escobari and Li (2007), Escobari (2012) and Escobari and Lee (2013) and Cornia, Gerardi and Shapiro (CGS, 2012). Escobari et al. focus on within flight correlation between price dispersion and capacity utilization: Flights that are relatively empty tend to have less price dispersion. CGS find a negative correlation between average capacity utilization and price dispersion: Routes with low average capacity utilization tend to have more price dispersion. In a previous version of this paper, I argue here that both observations are consistent with the UST model.

derived relationship between ln(HLP) and ln(HLU). I then add variables suggested by other models: 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 may be a proxy for monopoly power and is analogous to the number of airlines in the route used by Gerardi and Shapiro (2009) when studying price dispersion in the airline industry. I also use category dummies and size variables to capture the difference in the cost of not selling (or the value of inventories) 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.

$$\text{(41)} \quad \ln(HLP_i) = b_0 + b_1 \ln(HLU_i) + b_2 \ln(\mathrm{Re}\, v_i) + b_3 \ln(AvP_i) + b_4 (\#Stores_i) + \sum_j d_{ji} CD_{ji} + \sum_j s_{ji} SD_{ji} + e_i$$

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.

To check for robustness, I ran (41) after replacing the range measures by standard deviations measures of dispersion.

Unit surprise measures

The model assumes *iid* demand shocks. We should therefore take out a UPC specific seasonal element in aggregate demand. For example, the demand for cold drinks and hot dogs may be higher during the 4th of July week.

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:

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

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

Note that in (42) there is only one lag of 52 weeks designed to capture seasonality. In (42') 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 (42) 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.

Price surprises or just prices?

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 (1983) 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.

Price inertia and the services provided by the store

Another reason for controlling for "store effects" is that different stores may provide different services. Kaplan and Menzio (2014) find that about 10% of price dispersion can be attributed to "store effect". 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. Stores that sell hot dogs also sell milk and it is unlikely that store effects drive the results. I therefore start with measures of price dispersion that do not control for store effect. I control for store effect later, in the robustness checks section.

Sales

The distinction between sale prices and regular prices has been a major issue in the literature on price rigidity. See for example, Eichenbaum, Jaimovich and Rebelo (2011), Guimaraes and Sheedy (2011) and Kehoe and Midrigan (2010). The distinction arises because the "menu cost" of a temporary price reduction seems less than the "menu cost" of a regular price change. As a result some researchers have chosen to remove sale prices and consider only regular prices. Here I include all prices (regular and sale) in the measure of price dispersion. One reason for this choice is that from the consumer's point of view sale prices matters because a large amount of purchases are done in sale prices. More importantly, this choice is consistent with the UST model.

In the UST model there are no menu costs. Since sellers are indifferent about prices in the equilibrium range they can adopt various strategies. They may adopt for example an (S,s) type strategy which looks like temporary price reductions. See Eden (1994) and Head et al. (2012) for further analysis.

Increasing marginal cost, price changes and store effect

When demand is iid and the marginal cost is flat, the equilibrium price distribution does not change over time and it can be maintained without price changes. We can therefore have equilibrium with extreme price inertia in which each store stick to a single price that is consistent with (1). But in the data controlling for store effect eliminates only about 15% of the price dispersion.

One possible remedy is to assume that stores change their prices randomly without any reason. Another possibility is to assume correlated demand shocks. Here I assume increasing marginal cost. This is enough to generate changes in the equilibrium price distribution because changes in inventories cause changes in production and marginal cost.

I follow Bental and Eden (1993) and consider the case in which the cost of production is given by a strictly increasing and strictly convex function: $C(y)$. As before I use x_s to denote the aggregate supply to market s (over all sellers that post the price P_s). Let $\alpha(I)$ denote the first market price as a function of the beginning of period inventories. The arbitrage condition (1) can now be written as:

$$\frac{q_{i+1}}{q_i}P_{i+1} + \left(1 - \frac{q_{i+1}}{q_i}\right)\beta\alpha\left(I' = \sum_{s=i+1}^{Z} x_s\right) = P_i \tag{43}$$

This says that when market i opens sellers are indifferent between supplying to market i and supplying to market $i+1$. If they supply to market i, they get P_i per unit. If they supply to market $i+1$ they get P_{i+1} if the market opens and the value of inventories otherwise. The probability that market $i+1$ will open given that market

i opened is: $\frac{q_{i+1}}{q_i}$. The value of inventories in case market $i+1$ does not open is $\beta\alpha(I')$ where $I' = \sum_{s=i+1}^{Z} x_s$ is the beginning of next period inventories if exactly i markets open in the current period. Production in the BE model is determined by equating the marginal cost to the first market price:

(44) $$C'(y) = P_1$$

When each active buyer demands one unit and is willing to pay a high price for it, total supply $\sum_{s=1}^{Z} x_s$ is constant and total production $\sum_{s=1}^{Z} x_s - I$ depends on the beginning of period inventories. This and (44) imply that prices depend on the beginning of period inventories: When inventories are high, production and prices are low.

An Example

To see why temporary price changes may be used to maintain equilibrium, I consider the following 2 states example. The number of active buyers is 100 or 200 with equal probabilities. Each buyer demands one unit of the good and is willing to pay a high price for it. There are 200 sellers, the cost of production for each seller is $C(y) = (\frac{1}{2})y^2$, the marginal cost is $C'(y) = y$ and $\beta = \frac{1}{2}$. When demand is high everything is sold and we start the following period with $I = 0$. When demand is low only 100 units are sold and we start the following period with $I = 100$.

In periods with $I = 0$, production is 1 unit per seller and $P_1(0) = 1$. In periods with $I = 100$ production is 0.5 units per seller and $P_1(100) = 0.5$. Using the arbitrage condition (43) we can find the price in the second market when $I = 0$, by:

(45) $$P_1(0) = 1 = (\frac{1}{2})P_2(0) + (\frac{1}{2})\beta(\frac{1}{2}) = (\frac{1}{2})P_2(0) + (\frac{1}{4})\beta$$

This leads to: $P_2(0) = 1.75$.

When $I = 100$ the arbitrage condition is:

(46) $$P_1(100) = \frac{1}{2} = (\frac{1}{2})P_2(100) + (\frac{1}{4})\beta$$

This leads to: $P_2(100) = 0.75$.

In periods with no inventories 100 sellers supply to the first market (at the price of 1) and 100 sellers supply to the second market (at the price of 1.75). In periods with $I = 100$, there are 100 sellers who hold inventories and with their production they have 150 units for sale. The sellers with no inventories have 50 units for sale. Some adjustment in posted prices is necessary to maintain equilibrium.

One possibility is that sellers adopt one of the following 3 price strategies: Low, High and Switch. The low price sellers sell in the first market. The high price sellers sell in the second market. The switchers sell in the second market when inventories are low and in the first market when inventories are high. Table 1 describes this possibility over 3 periods. It is assume that demand was high in $t = 0$, low in $t = 1$ and high in $t = 2$. As a result inventories are zero at $t = 1$, 100 at $t = 2$ and zero at $t = 3$.

In this example temporary price reductions occur whenever inventories are high. Price dispersion measures are lower in the high inventories period and therefore eliminating “sale prices” in $t = 2$ will increase the average measures of price dispersion and may introduce a difference between goods that use temporary price reduction strategies to goods that use other strategies to maintain equilibrium. This is a problem from a theoretical point of view because the theory says that price dispersion should not depend on the way in which the equilibrium price distribution is achieved.

Table 1*: "Sales" when inventories are high

time	1	2	3
Inventories	0	100	0
1st mark. price	1	0.5	1
2nd mark. price	1.75	0.75	1.75
Price by seller type			
Low (100)	1	0.5	1
High (66.6)	1.75	0.75	1.75
Switch (33.3)	1.75	0.5	1.75
Supply per seller			
Low	1	0.5	1
High	1	1.5	1
Switch	1	1.5	1
Supply to the 1st market			
Low	100	50	100
High	0	0	0
Switch	0	50	0
Agg. supply	100	100	100
Supply to the 2nd market			
Low	0	0	0
High	66.6	100	66.6
Switch	33.3	0	33.3
Agg. supply	100	100	100

*This Table describes the changes in the distribution of prices and the changes in the prices of individual stores over 3 periods. The first row is the time index. The second is aggregate inventories. We start with no inventories at $t = 1$. This is followed by a period with 100 units of inventories and a period with no inventories. The third row is the first market price. The fourth is the second market price. The price strategies follow. There are 100 low price sellers who post the low price. There are 66.6 high price sellers who post the high price and there are 33.3 switchers who post the high price when inventories are low and the low price when inventories are high. Then we have the supply by individual sellers (production plus inventories). The high price sellers and switchers may have inventories while the low price sellers do not and therefore the supply per seller depends on the type. The rows that follow are supply to the first market by type. This is followed by supply to the second market by type.

4. 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.

4.1 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 2. The age distribution is different. The age distribution of group 1 is in the third column of Table 2 (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 3 uses Table 2 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 2 plots the data in Table 3. 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 2 thus provides more information than the correlation coefficients.[9]

Table 2*: 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 3*: 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.

UPCs instead of people:

I now turn to the IRI data for the randomly chosen week. In Figure 3, 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

[9] 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.

(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 3A, 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 2. 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 3B 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 3A 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 3A and 3B 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.

4.2 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 4 for the largest 2005 sample in detail and the averages across UPCs for the other two samples. 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 4 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_{it} denote the standard deviation of $\ln(U_{it})$ in week t and let SDU_i denotes the average 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 4 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 3 rows are averages across all UPCs. In the 2005 sample the average UPC is sold in 20 stores, has ln(HLU) of 1.46, SDU of 0.34, ln(HLP) of 0.35, SDP of 0.11 and the average size is 0.49. As can be seen these statistics do not vary much across samples.

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 4*: Summary Statistics

	# 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)
average 2005		20	1.46	0.34	0.35	0.11	0.49
Average 2004		15	1.61	0.38	0.38	0.13	0.50
Av. 04-05		15	1.62	0.35	0.37	0.13	0.45

* The statistics about individual categories use the 2005 sample. 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 three rows are the averages across UPCs in the three samples used.

Figure 4A 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 4B 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.

The correlations between the main variables in the 3 samples are in Table 5. 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 5*: 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.

4.3 Unit surprises

I used the combined 04-05 sample with 324 UPCs to run (24)-(24’) 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 (24’). The results are almost identical to the results when using (24).

5. ESTIMATIONS

As described in the data section, I use two measures of dispersion: The range measure and the standard deviation measure. Here I report the results when using the range measures. The regressions that use the standard deviation measures are reported in the Appendix.

I start with running price dispersion on unit dispersion for categories with more than 50 UPCs and for the samples as a whole. As can be seen from Table 6, 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.

Table 6*: Running ln(HLP) on ln(HLU) for selected categories.

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.

Table 7 uses the samples with all categories. It 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 7 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 (41) 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.

6. 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).

Using the residual unit dispersion measure

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 10 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 (23) 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 (41) as follows.

(47)
$$\ln(HLP_i) = b_{0i} + b_{1i}^H \ln(H_i) - b_{1i}^L \ln(L_i) + ... +$$
$$= b_{0i} + b_{1i}^H \ln(HLU_i) + (b_{1i}^H - b_{1i}^L)\ln(L_i) + ... +$$

The specification (41) is a special case of (47) that assumes: $b_{1i}^H = b_{1i}^L$. Table 11 provides the results when running (47). The coefficient of $\ln(L_i)$ is not significantly different from zero, thus supporting the specification (41).

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.

Store effect

As was illustrated by the example of Table 1, identical stores may choose different price strategies and therefore controlling for “store effect” may reduced the measure of price dispersion. In the extreme case in which shocks are *iid* and stores do not change prices, controlling for “store effect” will reduce the measure of price dispersion to zero. On the other hand it is possible that stores change prices often but provide different services.[10] In this case if we do not control for "store effect" we will overestimate price dispersion.

In general, "store effect" captures price inertia and services. Ideally we would like to capture only the services part but this is not possible. We can however, estimate an upper bound on the services component. To do that, I used the 04-05 sample with 324 UPCs and 104 weeks and ran 324 regressions of price logs on store dummies.

(48) $$\ln(P_{ijt}) = a_i + b_{ij}(store - dummy) + e_{ijt}$$

Here i is a UPC index, j is a store index and t indexes the week. Note that the "store effect" is UPC specific. The reason is that product placement is important. One store can consistently place UPC i in a relatively visible location.

I then replace the log of prices by the residuals from (48). In detail:

$e_{it}^H = \max_j(e_{ijt})$ = the highest residual for UPC i in week t

$e_{it}^L = \min(e_{ijt})$ = the lowest residual for UPC i in week t

The log difference of the price residuals is: $\ln(HLRP_{it}) = e_{it}^H - e_{it}^L$. The variable $\ln(HLRP_i)$ is the average of $\ln(HLRP_{it})$ over 104 weeks. The column ln(HLRP) reports the average of $\ln(HLRP_i)$ over all the UPCs in the category. The average log difference, reported in the last row of Table 12 is 0.31. This is 6 percentage points less than the average log difference when not controlling for a store effect reported in the last row of the column ln(HLP). Similarly, $SDPR_{it}$ is the standard deviation of e_{ijt} over stores (j) in week t and $SDRP_i$ is the average of $SDPR_{it}$ over weeks. The last column in Table 12, labeled "SDRP" is the average of $SDPR_i$ over all the UPCs in the category. The average standard deviation over all UPCs is 0.11 (in the last row of

[10] The two possibilities are not easy to distinguish even in principle. The "quality" of the store may be judged by the variety of the product it offers and more generally, by the probability of a stock-out: At the same price, a buyer prefers a store that he can find everything that is on his shopping list. But according to our model, the probability of a stock-out is higher for a low price store. We should therefore think of a store quality as attributes like location, cleanliness, average length of the line at the exit and parking availability.

the last column). This is 2 percentage point less than the average standard deviation when not controlling for store effect.

The upper bound on the amount of price dispersion caused by difference in services is thus 16% (= 6/37) if we use the range measure and 15% (= 2/13) if we use the standard deviation measure. This is an upper bound because as I said before the reduction in the price dispersion measures due to removing the “store effect” may occur because of price inertia rather than difference in services.

Table 13 is a correlation matrix that repeats the correlations in Table 5 for the 04-05 sample and adds the correlations with the new price dispersion measures that control for “store effects”. As we can see the correlation between the “old” and the “new” measures is about 0.9. The correlation between the “new” measures ln(HLRP) and SDRP is 0.98 and is similar to the correlation between the “old” measures. What is striking is that the correlation between the unit dispersion measures and the “new” price dispersion measures is higher than the correlation between the unit dispersion measures and the “old” dispersion measures. For example, the correlation between SDU and SDP is 0.53 and this is less than the correlation between SDU and SDRP that is 0.62. This suggests that the “store effect” dummy captures some differences in services and not merely price inertia.

Table 14 reports the results of running ln(HLRP) on ln(HLU). Comparing with the bottom of Table 6, it reveals higher coefficients of ln(HLU). Table 15 is the results of running ln(HLRP) on ln(HLU) and other variables. Comparing these results to the bottom rows of Table 7 reveals that the coefficient of ln(HLU) is now higher and the coefficient of ln(avgPrice) is lower (higher in absolute value). Table 16 is the results of running the standard deviation measure SDRP on SDU and other variables. Comparing these results with the bottom rows of Table A1 reveals that now the coefficient of SDU are higher and the coefficients of ln(avgPrice) are somewhat lower.

Table 12: Summary statistics for the 04-05 sample with and without fixed "store effect"

	#UPC	ln(HLU)	SDU	ln(HLP)	SDP	ln(HLRP)	SDRP
beer	20	1.27	0.27	0.18	0.06	0.13	0.04
carbbev	58	1.80	0.39	0.39	0.13	0.34	0.12
coldcer	53	2.31	0.51	0.45	0.17	0.40	0.14
fzpizza	12	1.66	0.32	0.40	0.14	0.30	0.10
margbutr	18	1.53	0.32	0.53	0.19	0.41	0.14
milk	23	0.85	0.17	0.36	0.12	0.25	0.08
peanbutr	11	1.44	0.28	0.35	0.12	0.24	0.08
saltsnck	42	1.50	0.32	0.31	0.11	0.29	0.10
soup	22	1.99	0.43	0.40	0.13	0.29	0.09
yogurt	65	1.29	0.27	0.35	0.13	0.32	0.11
Total	324						
average		1.62	0.35	0.37	0.13	0.31	0.11

* The first column is the category name. The second is the number of UPCs in the category. The next two columns are measure of demand uncertainty and the following two columns are measures of price dispersion that are comparable to the measures in Table 4 for the 05 sample. The last two columns are measures of price dispersion that control for a "store effect". The last row is the average across all the 324 UPCs.

Table 13*: Correlations between unit dispersion measures, "old" price dispersion measures and "new" price dispersion measures

	lnHLU	SDU	lnHLP	SDP	lnHLRP	SDRP
lnHLU	1					
SDU	0.97	1				
lnHLP	0.47	0.51	1			
SDP	0.5	0.53	0.97	1		
lnHLRP	0.55	0.6	0.91	0.9	1	
SDRP	0.58	0.62	0.87	0.9	0.98	1

* This correlation matrix uses the 04-05 sample with 324 UPCs. The "new" price dispersion measures (lnHLRP and SDRP) use the residuals from the regression of price log on store dummies. The "old" price dispersion measures (lnHLP and SDP) use price logs and do not control for store effects. The unit dispersion measures (lnHLU and SDU) are based on aggregate amounts sold.

Table 14: Running ln(HLRP) on ln(HLU)

I	Intercept	lnHLU	#UPC	Adj. R^2
carbbev	0.266***	0.044**	58	0.144
coldcer	0.087	0.134***	53	0.472
yogurt	0.242***	0.060*	65	0.077
All	0.161***	0.095***	324	0.306

Table 15*: Running ln(HLRP) on ln(HLU) and other variables.

	1	2	3	4	5	6
lnHLU	0.103*** (0.010)	0.097*** (0.009)		0.091*** (0.009)		0.096*** (0.009)
ln(Rev)			0.046*** (0.009)	0.035*** (0.008)	0.040** (0.013)	0.008 (0.012)
#Stores					0.002 (0.003)	0.009** (0.003)
ln(avgPrice)		-0.133*** (0.022)	-0.169*** (0.024)	-0.142*** (0.021)	-0.165*** (0.025)	-0.123*** (0.022)
Adj. R^2	0.497	0.550	0.437	0.580	0.436	0.590

* This is the same as Table 6 with a different dependent variable.

Table 16*: Running SDRP on SDU and other variables.

	1	2	3	4	5	6
SDU	0.179*** (0.012)	0.171*** (0.012)		0.165*** (0.012)		0.169*** (0.012)
ln(Rev)			0.013*** (0.003)	0.006* (0.002)	0.017*** (0.005)	0.001 (0.004)
#Stores					-0.001 (0.001)	0.002 (0.001)
ln(avgPrice)		-0.033*** (0.007)	-0.048*** (0.008)	-0.035*** (0.007)	-0.050*** (0.009)	-0.031*** (0.007)
Adj. R^2	0.619	0.646	0.440	0.652	0.440	0.654

*This is the same as Table A1 with a different dependent variable.

7. QUANTITATIVE IMPORTANCE

We have seen that the coefficients of the measures of demand uncertainty are statistically significant and relatively stable across samples. To get a sense of the economic significance I ask what will be the average price dispersion in a hypothetical world in which there is no demand uncertainty and the aggregate amount sold per week is constant over time.

Ideally we should focus on the dispersion of services adjusted prices or "true" prices defined by:

(49) $$P^*_{ijt} = \frac{P_{ijt}}{S_{ij}}$$

where P^*_{ijt} is the "true" price, P_{ijt} is the observed price and S_{ijt} is a measure of services provided by the store. Thus a price of 1 dollar in store 1 is equivalent to a price of 1.1 dollars in store 2 if store 2 provides 10% more services. I assume that the "true" price is the relevant price for both the buyer and the seller.[11]

The "true" price has a regular price component (R_{ij}) and a temporary element (T_{ijt}):

(50) $$P^*_{ijt} = R_{ij}T_{ijt}$$

The regular price reflects the choice between average capacity utilization and price. The temporary element may reflect changes that are required to achieve a UST equilibrium (this cannot be achieved if all stores stick to their regular price - see the example in Table 1). It may also reflect the desire to discriminate between shoppers and non-shoppers as in Varian (1980).

I therefore assume that the temporary element in the price has two components: one that is required to achieve UST equilibrium (u_{ijt}) and one that reflects discrimination (d_{ijt}). I assume $T_{ijt} = u_{iji}d_{ijt}$ and write (50) as:

(51) $$P^*_{ijt} = R_{ij}u_{iji}d_{ijt}$$

In an hypothetical world with no demand uncertainty the first two components in (51) are constants and we can write

(52) $$P^{*H}_{ijt} = k_i d_{ijt}$$

where k_i is a constant and the discrimination component d_{ijt} varies over stores and time.

[11] Thus a buyer in the UST model will choose to buy at the cheapest available "true" price. The seller will first choose the amount of services provided by the store. This cannot be easily changed and may be treated as a constant in the short run. He then chooses the price taking into account the probability of making a sale that is determined by the "true" price.

The range measure

Using (52) I write the ratio of the high to low hypothetical "true" prices as:

$$HLP^{*}H_{it} = \frac{\max_j \{P_{ijt}^{*H}\}}{\min_j \{P_{ijt}^{*H}\}} = \frac{\max_j \{d_{ijt}\}}{\min_j \{d_{ijt}\}} \tag{53}$$

Using (51) I write the ratio of the high to low "true" price as:

$$HLP_{it}^{*} = \frac{\max_j \{P_{ijt}^{*}\}}{\min_j \{P_{ijt}^{*}\}} = \frac{\max_j \{R_{ij}u_{ijt}d_{ijt}\}}{\min_j \{R_{ij}u_{ijt}d_{ijt}\}} \tag{54}$$

Dividing (53) by (54) leads to:

$$\frac{HLP^{*}H_{it}}{HLP_{it}^{*}} = \frac{\max_j \{d_{ijt}\}}{\min_j \{d_{ijt}\}} \frac{\min_j \{R_{ij}u_{ijt}d_{ijt}\}}{\max_j \{R_{ij}u_{ijt}d_{ijt}\}} \tag{55}$$

The ratio (55) is a measure of the importance of demand uncertainty: The lower this ratio is the more important is demand uncertainty in determining price dispersion. Unfortunately, we do not observe the "true" price and its components. But under certain conditions we can use the observed price and its components to estimate (55).

The "true" price is the price net of services. I define a gross price that includes services by $P_{ijt}^{*}S_{ij}$. The ratios of the gross prices that are analogous to (53) and (54) are:

$$HLPH_{it} = \frac{\max_j \{P_{ijt}^{*H}S_{ij}\}}{\min_j \{P_{ijt}^{*H}S_{ij}\}} = \frac{\max_j \{d_{ijt}S_{ij}\}}{\min_j \{d_{ijt}S_{ij}\}}; \quad HLP_{it} = \frac{\max_j \{R_{ij}u_{ijt}d_{ijt}S_{ij}\}}{\min_j \{R_{ij}u_{ijt}d_{ijt}S_{ij}\}} \tag{56}$$

And the "importance" measure analogous to (55) is:

$$\frac{HLPH_{i}}{HLP_{i}} = \frac{\max_j \{d_{ijt}S_{ij}\}}{\min_j \{d_{ijt}S_{ij}\}_i} \frac{\min_j \{R_{ij}u_{ijt}d_{ijt}S_{ij}\}}{\max_j \{R_{ij}u_{ijt}d_{ijt}S_{ij}\}} \tag{57}$$

The measure (57) is the same as the ideal measure (55) if S_{ij} is a constant that does not vary across stores. It is also the same as (55) if variation in S_{ij} are important and can be "factored out" in the following way.

$$\max_j \{x_j S_j\} = \max_j \{x_j\} \max_j \{S_j\}; \ \min_j \{x_j S_j\} = \min_j \{x_j\} \min_j \{S_j\} \tag{58}$$

where x_j denotes other components of the store's price.[12] This says that when services dominates the store with the highest price has the highest level of services and the store with the lowest price has the lowest level of services.

The "factoring out" property (58) may also hold when we have a large number of stores. To illustrate, I consider the case in which the x_j and S_j are independently distributed and each may take two possible realizations: high (H) and low (L) with equal probability of occurrence. With 20 stores (the average number of stores in the 2005 sample) the probability that the maximum of the product xS occurs when both random variables realize the high realization is:

$$\Pr ob\{\max(xS = HH)\} = \left(1-(½)^{20}\right)^2 = 0.999998$$

When the number of stores is 11 (the minimum number of stores in our samples) this probability drops to 0.998. When the number of stores is sufficiently large we can therefore use the following approximation.

$$(59)\quad \frac{HLPH_i}{HLP_i} \approx \frac{\max_j\{d_{ijt}\}\max_j\{S_{ij}\}}{\min_j\{d_{ijt}\}\min_j\{S_{ij}\}}\frac{\min_j\{R_{ij}u_{ijt}d_{ijt}\}\min_j\{S_{ij}\}}{\max_j\{R_{ij}u_{ijt}d_{ijt}\}\max_j\{S_{ij}\}} = \frac{HLP^*H_i}{HLP_i^*}$$

I also use the price net of store effect. This is not the "true" price because controlling for "store effect" eliminates variation in both services and the regular price, rather than just variations in services.

The ratio of prices net of store effect that is analogous to (53) and (54) is:

$$(60)\qquad HLRH_i = \frac{\max_j\{d_{ijt}\}}{\min_j\{d_{ijt}\}}\ ;\ HLR_i = \frac{\max_j\{u_{ijt}d_{ijt}\}}{\min_j\{u_{ijt}d_{ijt}\}}$$

And the ratio analogous to (55) is:

$$(61)\qquad \frac{HLRH_{it}}{HLR_{it}} = \frac{\max_j\{d_{ijt}\}}{\min_j\{d_{ijt}\}}\frac{\min_j\{u_{ijt}d_{ijt}\}}{\max_j\{u_{ijt}d_{ijt}\}}$$

This will equal (55) only if there is no price inertia and R_{ij} is a constant that does not vary across stores or if we can "factor out" R_{ij} in a way similar to the "factoring out" of S_{ij} in (58).

[12] In general, $\max_j\{x_jS_j\} \le \max_j\{x_j\}\max_j\{S_j\}$ and $\min_j\{x_jS_j\} \ge \min_j\{x_j\}\min_j\{S_j\}$.

The variance measure

To simplify, I assume that the "true" price does not depend on the level of services. A store that provides good services can offer a good deal and have a low "true" price. I also assume that the three components of the "true" price are independently distributed. Thus the probability of a "sale" is the same for "high true regular price" store and for "low true regular price stores". This simplifies the exposition but the main results hold if we allow for a positive correlation between the "true price" and services.

Writing (49) and (51) in log terms leads to:

(62) $$\ln(P_{ijt}) = \ln(S_{ij}) + \ln(R_{ij}) + \ln(u_{ijt}) + \ln(d_{ijt})$$

(63) $$\ln(P^*_{ijt}) = \ln(R_{ij}) + \ln(u_{ijt}) + \ln(d_{ijt})$$

After removing "store effect" we are left with the temporary components of the price:

(64) $$e_{ijt} = \ln(u_{ijt}) + \ln(d_{ijt})$$

The elimination of demand uncertainty will eliminate UST type reasons for price dispersion: The variation in regular true price and in temporary true price due to non-discrimination. We can therefore compute the variances of the hypothetical prices by substituting $Var(\ln R_{ij}) = Var(\ln u_{ijt}) = 0$ in (62)-(64). This leads to:

(65) $$Var(\ln P^H_{ijt}) = Var(\ln S_{ij}) + Var(\ln d_{ijt})$$

(66) $$Var(\ln P^{*H}_{ijt}) = Var(\ln e^H_{ijt}) = Var(\ln d_{ijt})$$

The ideal measure of the importance of demand uncertainty is:

(67) $$\frac{Var(\ln P^{*H}_{ijt})}{Var(\ln P^*_{ijt})} = \frac{Var(\ln d_{ijt})}{Var(\ln P^*_{ijt})}$$

The measure that does not control for "store effect" is:

(68) $$\frac{Var(\ln P^H_{ijt})}{Var(\ln P_{ijt})} = \frac{Var(\ln S_{ij}) + Var(\ln d_{ijt})}{Var(\ln S_{ij}) + Var(\ln P^*_{ijt})}$$

This measure is the same as the ideal measure (67) when all stores provide the same services and $Var(\ln S_{ij}) = 0$. Otherwise it is higher than the ideal measure.

The measure that controls for "store effect" is:

$$\frac{Var(\ln e_{ijt}^{H})}{Var(\ln e_{ijt})} = \frac{Var(\ln d_{ijt})}{Var(\ln u_{ijt}) + Var(\ln d_{ijt})} \tag{69}$$

This measure is the same as the ideal measure (67) if there is no price inertia and $Var(\ln R_{ij}) = 0$. Otherwise, it is higher than the ideal measure.

Since both (68) and (69) are larger than the ideal measure, it follows that:

$$1 - \frac{SD(\ln P_{ijt}^{H})}{SD(\ln P_{ijt})} < 1 - \frac{SD(\ln P_{ijt}^{*H})}{SD(\ln P_{ijt}^{*})} \text{ and } 1 - \frac{SD(e_{ijt}^{H})}{SD(e_{ijt})} < 1 - \frac{SD(\ln P_{ijt}^{*H})}{SD(\ln P_{ijt}^{*})} \tag{70}$$

The first inequality says that the percentage reduction in the standard deviation of the gross prices is less than the ideal measure of the percentage reduction. The second inequality says that the percentage reduction in the prices net of store effect is less than the ideal measure of the percentage reduction.

Estimation

I now attempt to estimate the effect of eliminating demand uncertainty for an "average" UPC (that is for a UPC with the average unit dispersion and the average price dispersion).

Table 17 focuses on the range measures. The category name is in the first column. The second column (C1) is the coefficient of ln(HLU) in the regression of ln(HLP) on ln(HLU) and other variables (taken from Table 7). This coefficient is 0.094 for the 2005 sample, 0.102 for the 2004 sample and 0.083 for the combined 04-05 sample. The third column is the average of the log of the high to low price ratio in the sample. These are 0.34, 0.38 and 0.37. The average ratios (anti-logs) are 1.4, 1.46 and 1.45. The forth column ln(HLPH) is the hypothetical ratio of high to low price computed as: ln(HLPH) = ln(HLP) - C1*ln(HLU), where HLU is the average high-low ratio in the sample. These are 0.21, 0.21 and 0.24. The hypothetical anti-logs are 1.23, 1.23 and 1.27. The fifth and the sixth columns are the estimated percentage reduction in price dispersion that will follow the elimination of demand uncertainty.

The fifth column uses the logs while the sixth column uses the actual ratio (the anti-logs). As we can see from the last column, the percentage reduction is between 41 and 48 percent.

Table 18 focuses on the standard deviation measure of dispersion. The second column is the coefficient (C2) of SDU in a regression of SDP on SDU and other variables. The third column is the average SDP in the sample. The forth column is the hypothetical SDP calculated as: SDPH = SDP - C2*SDU, where SDU is the average in the sample. The last column is the estimated effect of demand uncertainty on price dispersion. As can be seen the elimination of demand uncertainty will reduce the standard deviation by 39-44 percent.

Tables 19 and 20 repeat the hypothetical experiment after controlling for "store effect". In this case, the elimination of demand uncertainty will reduce price dispersion by 54 percent.

Table 17*: The hypothetical range measure

Sample	C1	$\ln HLP$	$\ln HLPH$	$\frac{\ln HLP - \ln HLPH}{\ln HLP}$	$\frac{HLP - HLPH}{HLP - 1}$
05	0.094 (0.087,0.101)	0.34	0.21 (0.20,0.23)	0.39 (0.37,0.42)	0.44 (0.41,0.47)
04	0.102 (0.095,0.109)	0.38	0.21 (0.20,0.23)	0.43 (0.40,0.46)	0.48 (0.45,0.51)
04-05	0.083 (0.074,0.092)	0.37	0.24 (0.22,0.25)	0.36 (0.32,0.40)	0.41 (0.37,0.45)

*The first column is the sample used. The second is the coefficient of ln(HLU) taken from Table 7. In parenthesis are the lower and upper bounds of the estimated coefficients. Thus for example, in 2005 the estimated coefficient is 0.094 and the standard error is 0.007. The lower bound of the coefficient is 0.094-0.007 = 0.087 and the upper bound is 0.094+0.007 = 0.101. The third column is the average lnHLP in the data. The fourth is the hypothetical lnHLP calculated as: Ln(HLPH)=ln(HLP) - C1*Ln(HLU), where Ln(HLU) is the average of the log HLU in the data. In parenthesis are the calculation when using the lower and upper bound of C1. The fifth and sixth columns is the percentage decline in price dispersion. The fifth is the ratio of the log difference (lnHLP-lnHLPH) to lnHLP and the last column is the ratio of the percentage difference HLP-HLPH to HLP-1. In parentheses are the computation when using the lower and upper bounds of C1.

Table 18*: The hypothetical standard deviation measure

Sample	C2	SDP	$SDPH$	$\frac{SDP-SDPH}{SDP}$
05	0.147 (0.139,0.155)	0.11	0.063 (0.061,0.066)	0.44 (0.41,0.46)
04	0.152 (0.143,161)	0.13	0.077 (0.073,0.080)	0.43 (0.40,0.46)
04-05	0.145 (0.133,0.157)	0.13	0.079 (0.075,0.083)	0.39 (0.36,42)

* The second column is the coefficient of SDU taken from Table A1 (upper and lower bounds in parenthesis). The third is the average SDP in the data. The fourth is the hypothetical SDP calculated as: SDPH = SDP - C2*SDU. The last column is the ratio of the difference SDP-SDPH to SDP.

Table 19*: The hypothetical range measure after controlling for “store effects”

Sample	C1	ln(HLRP)	ln(HLRPH)	$\frac{\ln HLRP - \ln HLRPH}{\ln HLRP}$	$\frac{HLRP - HLRPH}{HLRP - 1}$
04-05	0.096 (0.087,0.105)	0.31	0.15 (0.14,0.17)	0.5 (0.45,0.55)	0.54 (0.49,0.59)

* The Table reports the hypothetical experiment results after controlling for “store effect”. The coefficient of ln(HLU) is from Table 15 (lower and upper bounds in parenthesis), the ratio of the residual ln(HLRP) is 0.31 (Table 12), the hypothetical ratio is Ln(HLRPH) = ln(HLRP)-(C1)ln(HLU) = 0.15 and the percentage reduction in the dispersion measures due to the elimination of unit dispersion are 0.5 and 0.54.

Table 20*: The hypothetical standard deviation measure after controlling for “store effects”

Sample	C2	SDRP	SDRPH	$\frac{SDRP-SDRPH}{SDRP}$
04-05	0.169 (0.157,0.181)	0.11	0.051 (0.047,0.055)	0.54 (0.50,0.58)

* The second column is the coefficient of SDU taken from Table 16. The third column is the standard deviation of the residuals taken from Table 12. The fourth column is the hypothetical standard deviation calculated as SDRPH = SDRP - (C2)SDU = 0.051 where SDU is from Table 12. The last column is the percentage reduction in the standard deviation that will follow the elimination of demand uncertainty.

Using the inequalities in (70) we may say that eliminating demand uncertainty will reduce the standard deviation measure of “true” price dispersion by more than 54 percent. This is a big effect.

8. CONCLUDING REMARKS

Consistent with the theory, I find that price dispersion is increasing in measures of unit dispersion. To check for robustness, I include in the regressions three variables suggested by search and discrimination theories: 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 additional variables does not change the unit dispersion coefficient by much. This coefficient 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 additional variables used, the average price is the only one with a stable and significant effect. As in Pratt et. al. (1979), higher average price reduces price dispersion.

The effect of demand uncertainty on price dispersion has economic as well as statistical significance. Our estimates suggest that eliminating demand uncertainty will on average, reduce the cross sectional standard deviation of the price log by more than 50%.

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.

Borenstein, S. and Rose, N., 1994, 'Competition and Price Dispersion in the U.S. Airline Industry,' The Journal of Political Economy, 102(4), pp. 653–683.

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

Cornia Marco & Kristopher S. Gerardi & Adam Hale Shapiro, 2012. "**Price Dispersion Over the Business Cycle: Evidence from the Airline Industry**," Journal of Industrial Economics, Wiley Blackwell, vol. 60(3), pages 347-373, 09.

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.

________ "**Monopoly Price Dispersion under Demand Uncertainty**," International Economic Review, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, vol. 42(3), pages 649-70, August, 2001.

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, Vol. 98, No.6,1293-1306, Dec. 1990.

_________ "**The Adjustment of Prices to Monetary Shocks When Trade Is Uncertain and Sequential**," Journal of Political Economy, University of Chicago Press, vol. 102(3), pages 493-509, June 1994.

________ "**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.

Eichenbaum Martin & Nir Jaimovich & Sergio Rebelo, 2011. "**Reference Prices, Costs, and Nominal Rigidities**," American Economic Review, American Economic Association, vol. 101(1), pages 234-62, February.

Escobari, Diego 2012. "**Dynamic Pricing, Advance Sales and Aggregate Demand Learning in Airlines**," Journal of Industrial Economics, Wiley Blackwell, vol. 60(4), pages 697-724, December.

________ and Li Gan, 2007. "**Price Dispersion under Costly Capacity and Demand Uncertainty**," NBER Working Papers 13075, National Bureau of Economic Research, Inc.

________ and Lee, Jim, 2013. "**Demand Uncertainty and Capacity Utilization in Airlines**," MPRA Paper 46059, University Library of Munich, Germany.

Gerardi Kristopher S. & Adam Hale Shapiro, 2009. "**Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry**," Journal of Political Economy, University of Chicago Press, vol. 117(1), pages 1-37, 02.

Guimaraes Bernardo & Kevin D. Sheedy, 2011. "**Sales and Monetary Policy**," American Economic Review, American Economic Association, vol. 101(2), pages 844-76, April.

Head Allen & Lucy Qian Liu & Guido Menzio & Randall Wright, 2012. "**Sticky Prices: A New Monetarist Approach**," Journal of the European Economic Association, European Economic Association, vol. 10(5), pages 939-973, October.

Hendel Igal & Aviv Nevo, 2013. "**Intertemporal Price Discrimination in Storable Goods Markets**," The American Economic Review, Vol. 103, No. 7. (December), Pages 2722-2789.

Kaplan Greg and Guido Menzio. “The Morphology of Price Dispersion” PIER working paper 14-002, January 2014.

Kehoe Patrick J. & Virgiliu Midrigan, 2010. "**Prices are Sticky After All**," NBER Working Papers 16364, National Bureau of Economic Research, Inc.

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

Pesendorfer Martin, 2002. "**Retail Sales: A Study of Pricing Behavior in Supermarkets**," The Journal of Business, University of Chicago Press, vol. 75(1), pages 33-66, January.

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.

Reinganum, Jennifer F, 1979. "**A Simple Model of Equilibrium Price Dispersion**," Journal of Political Economy, University of Chicago Press, vol. 87(4), pages 851-58, August.

Rob, Rafael, 1985. "**Equilibrium Price Distributions**," Review of Economic Studies, Wiley Blackwell, vol. 52(3), pages 487-504, July.

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.

Stahl, Dale O, II, 1989. "**Oligopolistic Pricing with Sequential Consumer Search**," American Economic Review, American Economic Association, vol. 79(4), pages 700-712, September.

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.

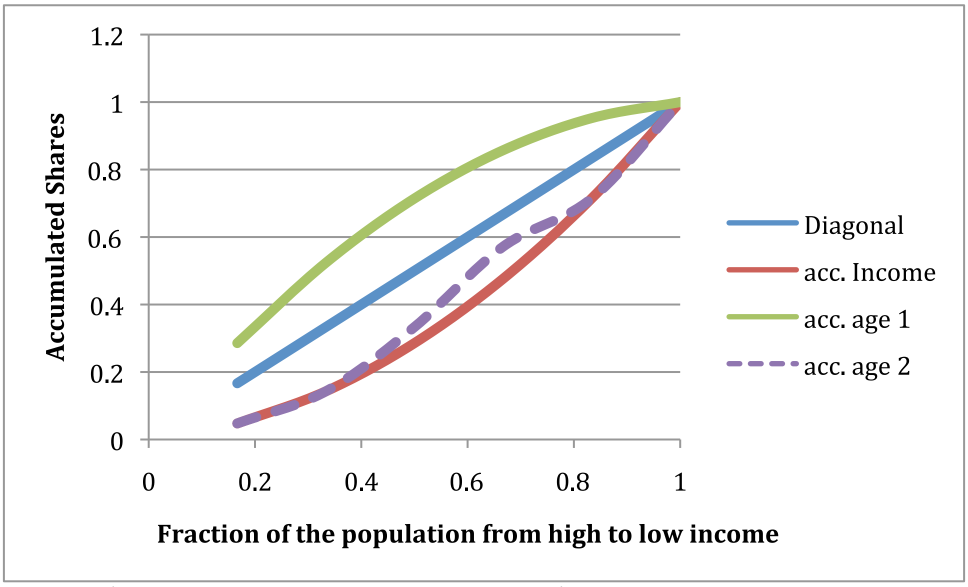
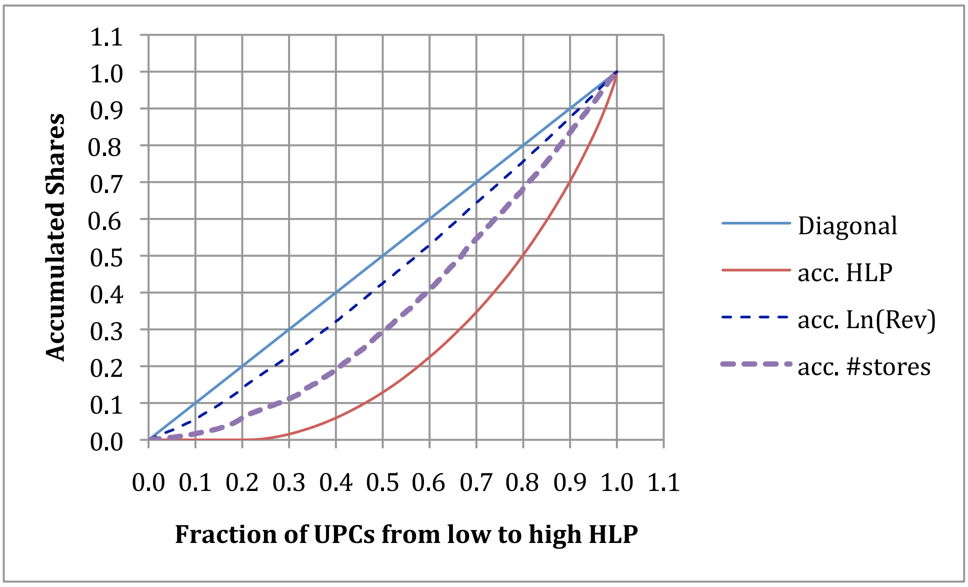
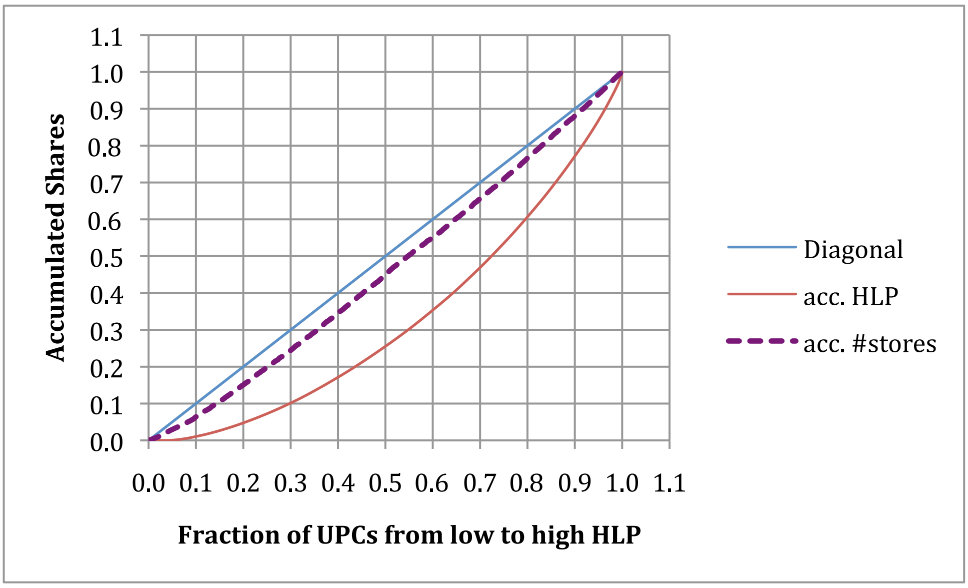


Figure 2: A plot of the data in Table 3 (example)

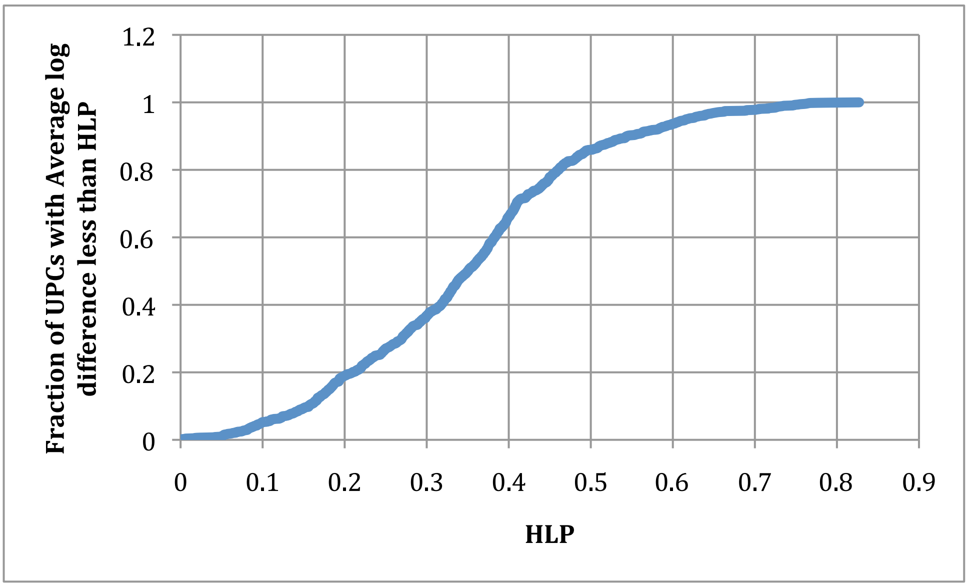


A. The Sample of 8602 UPCs sold by more than 1 store.

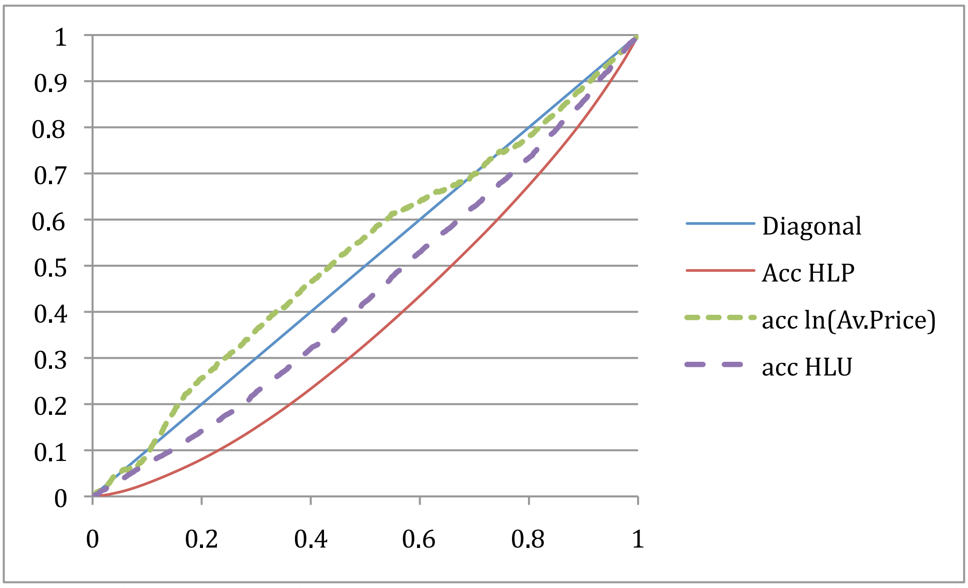


B. The sample of 4537 UPCs sold by more than 10 stores

Figure 3: Cumulative shares for the week starting January 17, 2005. UPCs are ordered from low to high price dispersion (HLP)

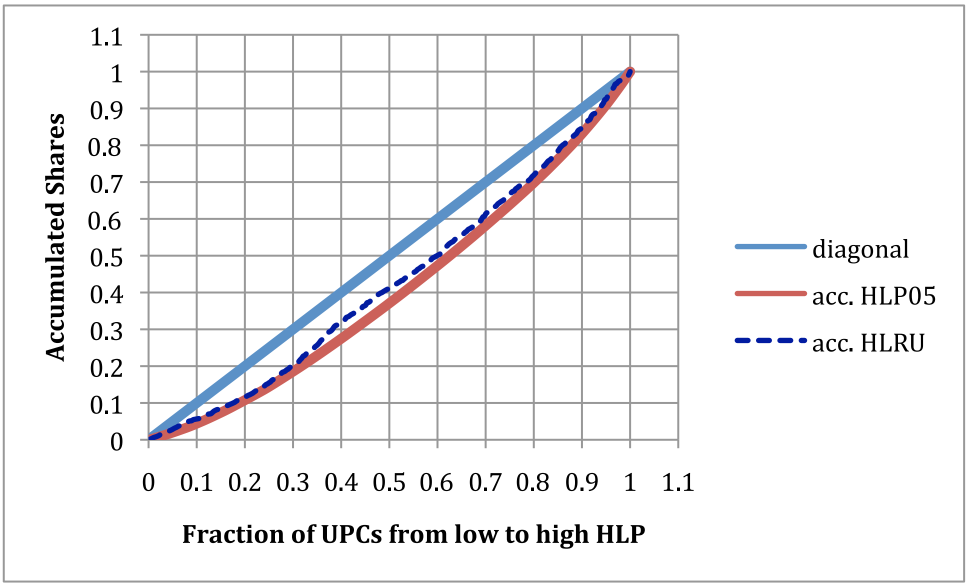


A. The Cumulative Frequency Distribution of the log difference between the highest and the lowest price averaged over weeks (ln[HLP]).

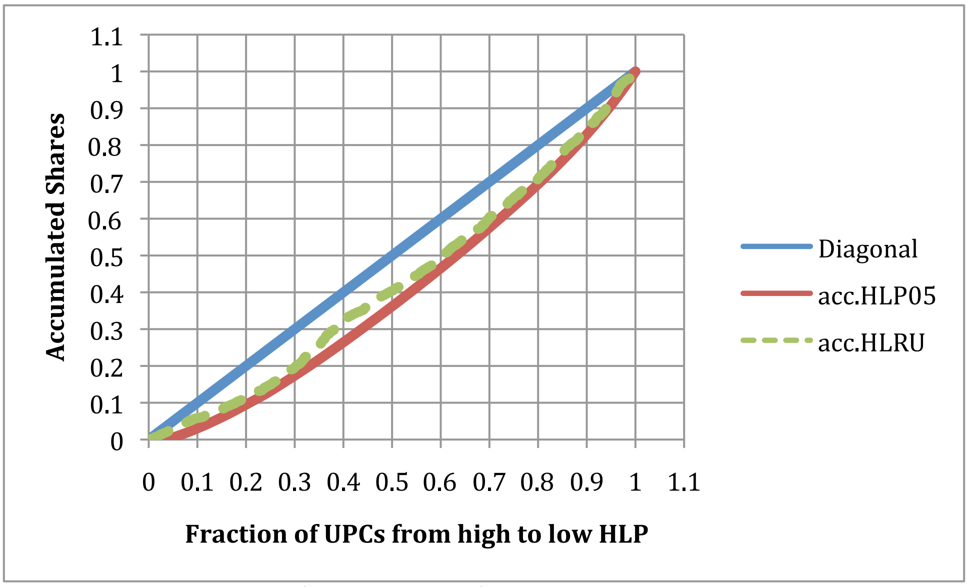


B. Cumulative Shares in totals. UPCs are ordered from low to high price dispersion (HLP)

Figure 4: Price dispersion in the sample of 1084 UPCs sold by more than 10 stores in all the weeks of 2005



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.

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.