Pricing for Data Markets

Avanish Kushal,
Sharmadha Moorthy,
Vikash Kumar

ABSTRACT

The increasing accessability and speed of Internet has enabled emergence of an online market where data providers publish their valuable information and from which customers subscribe and query data of all types. This virtual space in which the supply-demand forces exist and the exchange of information is incentivized by monetary gain is referred to as the data market (e.g. Windows Azure Marketplace DataMarket). The commonly observed pricing models for information goods and data markets are per unit pricing (metered usage pricing) and site licensing (unlimited usage pricing). There hasn't been considerable work on determining a pricing model and parameters for profit maximization for pricing datasets which has great relevance and applicability in the datamarket. The paper determines the optimal prices of datasets for profit maximization from an Economics framework both for unit and step pricing scheme. This paper validates a pricing model against properties such as arbitrage & competition. The paper presents necessary and sufficient conditions for a valid pricing scheme and provides an $O(n^2)$ algorithm to validate multi-step pricing models of existing datasets in the datamarket. The paper also formulates the problem of computing the closest valid pricing model(under L1 Norm) as linear program for invalid pricing models.

1. INTRODUCTION

Technology has dramatically improved the ways in which data can be stored, analyzed and disseminated. The Internet facilitates almost instantaneous access to data and information and plays a key role in enabling data share across businesses, research organizations, governments and individuals alike. Researchers in genomics, for example, rely heavily on publicly available resources such as GenBank [13], a repository of annotated DNA sequences. Every company has reams of data that are immensely valuable: sales data, marketing analytics, financial records, customer insights and intellectual property it has generated in the course of business. Increasing amounts of data as well as extensive data analysis has catalyzed the emergence of data market as a single, consistent solution to the e-commerce challenges.

The Windows Azure Marketplace DataMarket [4] is a cloud service offering which functions as an information marketplace and brokerage business. It provides all the facilities a data provider needs to monetize the intrinsic value of carefully created datasets, which are subscribed by customers. The dataset are priced as monthly subscriptions [6] which are of two types. In the unlimited subscription type, the consumer is charged monthly for access to dataset and the transactions on the dataset are unlimited. For the limited subscription type, each month there is a pre-defined number of transactions that can be executed on the dataset. Each page of results returned from a query uses a single transaction and will count toward the transaction limit. A page of results may return up to 100 records. In most cases, there is a percentage markup to cover the cost of bandwidth, compute, and billing expenses. While some of the datasets are available for free, most of the proprietary datasets follow a multi-step pricing model that vary on the limit of transactions allowed per month. The data providers set pricing and terms and the data market allows access to the data. However, there is little guidance on how each data provider should price his dataset, especially in the presence of competition.

Data-management-in-the-cloud services such as Windows Azure Storage [7], SQL Azure [5], Google App Engine Datastore [3], Amazon SimpleDB [2] and Amazon RDS [1] charge for compute time in addition to storage and bandwidth size. However, such content-time pricing as discussed in [11] is not desirable for data markets. Firstly it gives an undue advantage to experienced users over people who are new to the database, as they would spend less time on the server. Further this would necessitate free systems where people can experiment on the query before migrating to the paid site. It also gives a negative incentive for sellers to improve performance as any optimization would reduce compute time , and thus the seller would require to reset the price frequently. Thus going forward we concentrate on pricing schemes based on the data returned.

Information goods vendor offer different pricing schemes among which the most common and widely studied are the site licensing or subscription based and per unit or pay-as-you-go pricing. The factors that decide the choice of pricing scheme include consumers' usage patterns; and the differentiation from competing firms to alleviate price competition and increase profit. As discussed in [10], the information goods market can be divided into two segments of consumers, one with a declining marginal Willingness to Pay (WTP) (d-type) and the second with a constant marginal WTP (c-type). For example, a firm may have declining marginal

WTP for WolframAlpha Facts dataset when it is used by individual employees for productivity improvements. On the other hand, firms that plan to use Consumer Expenditure Data as part of their enterprise system for market analytics and product development will have a constant and fixed requirement for the dataset.

We establish properties of a good pricing scheme that enforce that the price should vary sub-linearly with increasing transactions; the pricing model is arbitrage free; and the profit returns are consistent with the cumulative WTP of each consumer segment. However, we find these properties do not always hold in practice. For example two pricing schemes in which we found fallacies is shown in the following table (refer to Appendix A.2 for more examples from Azure DataMarket).

Datasets from Windows Azure Marketplace Datamarket			
Company	Dataset	Transaction limit	Price level
ESRI	2010 Key US Demographics	150 100 50 25 10	$49.95 $39.95 $24.95 $19.95 $9.95
Weather Central LLC	Weather Imagery	1000000 100000 10000 2500	$2400 $600 $120 $0

Consider a consumer in need of 60 transactions of the Demographics dataset from ESRI. He would rather buy 2 accounts for 50 and 10 transactions for a cost of 34.90$ rather than buying one account for 100 transactions costing 39.95$. The pricing of the Weather Imagery dataset has a glaring loop-hole. If a customer wishes to buy 10000 transactions, he could buy 4 account of 2500 transactions for free rather than the account for 10000 transactions costing 120$. To avoid arbitrage, in our pricing structures, we derive conditions that guarantee the validity of generalized functions. We come up with alternate optimal multi-step pricing scheme as solutions of linear programs which compute the closest pricing scheme in case of fallacies.

We have extended the site licensing approach in [10], where they consider only one level for step pricing model. In Azure DataMarket, in addition to such single level subscription pricing, there are available multi-step subscription pricing fixed on different sets of transaction limit to access the dataset. We setup a general framework to express the profit for a generalized multi-step licensing pricing against the cumulative WTP of consumer segment and derive conditions for maximizing profit. We show that for c-type buyer segment,restricting the number of steps to one, the maximum profit is in line with the results shown in [10]. We also study other forms of the WTP function and derive conditions for profit maximization.

The rest of the paper is organized as follows. In Section 2, we begin with literature survey of work in database systems and economics which focus on monetization of information. In Section 3, we describe the desirable properties of a pricing scheme. In Section 4 we view pricing from an Economics framework and come up the optimal model and parameters for profit maximization. In Section 5, we present necessary and sufficient conditions to validate pricing models and provide an $O(n^2)$ algorithm for validation. We also formulate the problem of computing the closest valid pricing model(under L1 Norm) as linear program for sample invalid pricing models found in Azure DataMarket. We conclude with our findings and future work in Section 6.

2. RELATED WORK

To the best of our knowledge, there hasn't been any work on the pricing of datasets in the data markets. We therefore set to survey related work in databases where monetization is considered as well as study existing pricing models in Economics.

The economic bidding system of distributed database Mariposa [14] operates under a limited-resources assumption, and charges clients based on amount of resources it expects to expend on their query, rather than an externally defined value of the data itself. This assumption leads to database being able to service only a fraction of its clients, and it hence uses an auction system to determine who is given the privilege. We are more interested in heterogeneous data utilities, which can be more easily studied by assuming unbounded computational resources.

The paper on Relational Data Markets [9] proposed a distributed database system in which providers can set arbitrary costs for their data, and clients can collect data from multiple overlapping providers of their choice, in order to fill their needs at minimal cost. This intends to retrieve datasets at an optimal pricing from the buyer's perspective and assumes data providers set an arbitrary cost for their data which is in contrast to our problem of determining an approach for seller to price their data.

The motivation behind pricing theory is to maximize profits on top of recovering cost of production. How to Price Shared Optimizations in the Cloud [8] offers game theoretic approach to recover costs of optimizations in the face of collaborations, where multiple users access the same dataset and one optimization can benefit multiple users. This approach solves an orthogonal problem to ours, since we aim to price data for sellers taking into consideration the various buyer segments and their willingness to pay. In our framework, we assume negligible marginal cost of production and the constant fixed costs which don't disturb the equilibrium solution as long as total profits is greater than fixed costs.

In [12], the introduction of a product is characterized as a multistage game. At the beginning of the first stage, two identical profit-maximizing firms set their product design denoted by a product index and the time to market is assumed to be linear function of product index. The firm with lower product index enters the market first and sets a monopoly price. The second firm enters at a later time with a duopoly price forcing the first firm to re-adjust its price to stay competitive. The paper establishes Nash equilibrium between product indices of the two firms and the monopoly and duopoly prices. Given the equilibrium, the optimal product index and time of entry into market are computed to maximize profits for both the firms. This is a possible extension to our work as it explores competition and effect on prices of firms' products. The concept of product index can be applied in case of competing firms with overlapping datasets such that dataset with higher design would have higher product index over the other.

3. PROPERTIES OF A PRICING MODEL

Here, we describe a set of desirable properties that a "good" pricing scheme for datasets should satisfy. Note that these properties are stated from the perspective of a seller trying to maximize his profits. The price is modeled as a positive increasing function $P(n)$ of the number of transactions($n \geq 0$) to be bought.

1. **Arbitrage Free:** A pricing model should ensure that the price of the union should be less than the prices of the individual units i.e.,
$P(n_1 + n_2) \leq P(n_1) + P(n_2)$
2. **Diminishing Returns:** The price should vary sub-linearly with the number of units sold, i.e. we are looking for a function $P(n)$ such that $\frac{P(n)}{n}$ is a decreasing function.
3. **Consumer Buying Power:** The pricing model should capture and closely follow the buying power of the customers, i.e. should try and be as close to the customers' maximum WTP(defined in 4.1.1).
4. **Competition:** The pricing should take into account the number of competitors selling overlapping records in their datasets and the price of their datasets.

In the next section we study the effect of Property 3 and Property 4 on the pricing function in order to maximize profit. In Section 5 we discuss the problem of ensuring that the function is arbitrage free, which we refer to as a "valid" function.

4. ECONOMIC FRAMEWORK

This section models an economic framework for pricing in datamarkets. We study two commonly used pricing models in datamarkets- unit pricing model and step pricing model, both in absence and presence of competition. Section 4.2 details pricing models which, in light of details about consumer segment, is extended in Section 4.3. Pricing in datamarket is modeled as two stage game where the seller determines the pricing model in the first stage and optimal prices in the second stage. In absence of competition, model and prices are given by comparing *Proposition 1* with *Proposition 2*. Section 4.2 addresses pricing when there is competition.

4.1 Definitions:

4.1.1 Willingness To Pay:

Willingness To Pay(WTP) is defined as the maximum amount a person is willing to pay for a unit quantity of a product. It is expressed as a function on the number of units, thus WTP(n) refers to the willingness to pay for the next unit of the product assuming he has bought n units already. In our case unit quantity is a transaction. WTP has following properties:

(a) $WTP(n) \geq 0$; where $WTP(n)$ is WTP for n^{th} transaction.

(b) WTP[1] is continuous, smooth and non-increasing on n (number of transactions). i.e.$(WTP)' \leq 0$

4.1.2 Cumulative Willingness To Pay:

Cumulative willingness to pay($CWTP(n)$) is defined as maximum willingness to pay for a total of n number of transaction. $CWTP(n) = \sum_0^n WTP(n)$. The following properties follow for CWTP:

(a) $CWTP(n) \geq 0$;

(b) CWTP[1] is continuous, smooth and non-decreasing on n. i.e. $(CWTP)' \geq 0$

(c) CWTP[1] is a concave function on n. i.e. $(CWTP)'' \leq 0$

[1]*n is usually large therefore WTP is loosely refereed as being continuous over n.

4.1.3 Seller

Sellers are assumed to be rational players selling the dataset. Sellers allow customers to query over their dataset and charge them on the number of transactions. A fixed number of records (Example 100 for Azure DataMarket) is counted as one transaction. Each seller of a dataset is fully informed about his/her dataset, the composition of its consumer segment and the distribution of demand across the consumer segment.

4.1.4 Customer

Customers are players interested in buying a dataset from the market with positive willingness to pay. Customers are completely informed and rational. They only buy when their $CWTP(n) \geq P(n)$(price for n transactions). If there exist choices in the market, they choose to buy from the seller with the cheapest price.

4.1.5 DataMarket

A DataMarket brings sellers and customers of a data together. The market is assumed to be a fair market with healthy competition. Each competitor sets prices independently and they don't collaborate to influence the dynamics of the market.

4.2 Assumptions:

We make the following assumptions in developing our pricing model

(a) The pricing model depends only on the number of transactions to access the dataset.

(b) Each transaction is priced independent of the data records returned as the result. The the pricing model is data invariant.

(c) There is a cost associated with the production of the dataset only in the beginning when it is published on the datamarket. There is no cost for the seller for every time it is sold.

(d) The WTP function is the same for all customers. This can be justified in some sense by averaging over all the customers.

4.3 Pricing:

A huge variety of pricing models exist in the present market. Each seller selects the individual model and its features based on the product and customer segment. Say, a seller S is interested in selling a dataset DS to a customer segment C and chooses a pricing model P. n is a random variable denoting the number of transactions desired by the members of the customer segment C. $P(n)$ gives the price for n transactions over dataset DS. $CWTP(n)$ gives the CWTP of C for n transactions. Across C, n is distributed over [0 N] with distribution $D(n)$.
We define a buying function B as a function of the number of transactions. It denotes the price a customer will pay if he intends to buy n units. Note this may be different from P(n) as he may end up buying less than n items if $CWTP(n) < P(n)$.

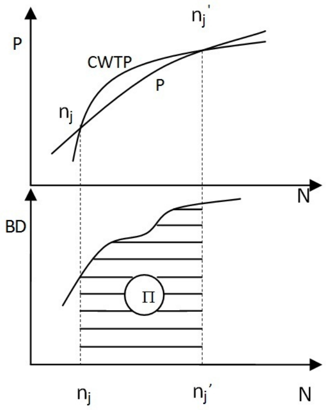

Figure 1: The Figure shows the shaded area under the BD curve(where the consumer CWTP is greater than the price) as the profit

Profit of S is given by the area under the BD(i.e. the product of the $B(n) * D(n)$) curve, as shown in Figure 1.

$$\Pi = \int_0^N B(n)D(n)dn \tag{1}$$

4.3.1 *Pricing model* P*:*

Two commonly observed pricing models used presently to price information goods and data are : (1) Per unit pricing or metered pricing. (2) Batch pricing or Step Pricing. Another pricing model that is often observed is Subscription (fixed cost for unlimited uses) based pricing which we look at as a special case of step pricing with a single step.
We reviewed some of the pricing models presently used to price information goods, cloud services, database services etc(as detailed in Section 2).We see that the models and parameters of the pricing models are not chosen in a proper manner, which keeps the seller's profit suboptimal. In the remaining part of the section, we provide mathematically sound Economics framework based on the above definitions, to choose which pricing scheme a seller should choose, based on his dataset and consumer segment, and how to determine the parameters of the model for profit maximization.

4.3.2 *Unit Pricing Model:*

Unit pricing or pay as you go pricing or metered pricing charges for the exact number of executed transactions. The pricing function varies linearly with the number of transaction ,i.e.

$$P(n) = P_u n; \text{ where } P_u \text{ is the price for } 1^{st} \text{ transaction}$$

As shown in Figure 2, define critical points n_j and n_j' where the function CWTP(n) and P(n) intersect. Thus the intervals (n_j, n_j') define where $CWTP(n) > P(n)$. Note to handle the boundary case, if $CWTP(\epsilon) > P(\epsilon), \epsilon --> 0$ we can define $n_1 = 0$. Similarly if $CWTP(N) > P(N)$, then we set $n_m' = N$, where m is the last interval.

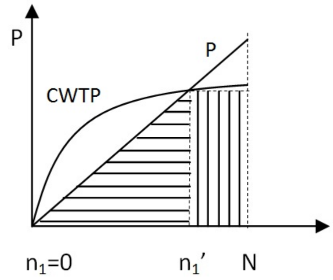

Figure 2: The Figure shows the critical points where the CWTP and price curves intersect and the shaded regions show where the customers buy from a unit pricing model

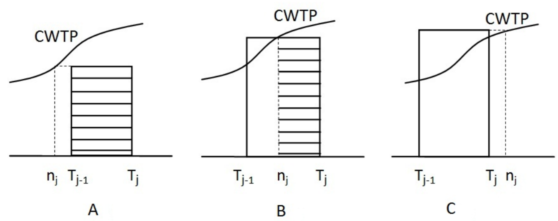

Figure 3: The Figure details the 3 possible scenarios of the location of critical points for a particular step of a step pricing model

The profit from unit pricing is given by

$$\begin{aligned}\Pi_u &= \int_0^N \mathbf{B(n)}\, D(n)dn \\ &= \int_{n_1=0}^{n_1'} P_u \cdot n \cdot D(n)dn + \int_{n_1'}^{N} P_u \cdot n_1' \cdot D(n)dn \end{aligned} \tag{2}$$

4.3.3 *Step Pricing Model:*

Step pricing or batch pricing can be viewed as a subscription based pricing with multiple levels of the form (P_1, T_1), $(P_2, T_2), ..., (P_m, T_m)$ where m is the total number of steps and $P_1 \le P_2 \le P_3 ... \le P_m$ and $T_1 \le T_2 \le T_3 ... \le T_m = N$. This structure implies that the customer can buy upto T_j tuples by paying P_j. The buying function is extended to pricing levels as

$$\begin{aligned} B_j(n) &= P_i \quad \text{where } i = max\big(i \le j | (P_i \le CWTP(T_i)\big) \\ &= 0 \qquad\qquad\qquad\qquad otherwise \end{aligned} \tag{3}$$

Buying function returns the price(say P_i) of the level customer has the willingness to pay for and he buys upto $\min(T_i, n)$ transactions for P_i.

Now, consider a level $j(T_{j-1} \le n < T_j)$. Since $(CWTP)' \ge 0$, each pricing level will have at most one critical point(n_j). In other words, $P_j = CWTP(n)$ will have atmost one solution. Based on if $n_j \le T_{j-1}$ or $T_{j-1} < n_j \le T_j$ or $n_j > Tj$ we analyze three different cases. as shown in Figure 3.

C1: $n_j \le T_{j-1}$
In this case everyone in level j will buy from level j.

Profit from level j

$$\Pi_j = \int_{T_{j-1}}^{T_j} P_j D(n) dn \quad (4)$$

C2: $T_{j-1} < n_j < T_j$
Customers with $n_j \leq n < T_j$ will buy from level j and customers with $T_{j-1} \leq n < n_j$ will buy from a level i given by $max(i \leq j|(P_i \leq CWTP(T_i)))$. Profit from level j

$$\Pi_j = \int_{T_{j-1}}^{n_j} P_i D(n) dn + \int_{n_j}^{T_j} P_j D(n) dn \quad (5)$$

C3: $n_j \geq Tj$
Everyone in level j will buy from a level i given by $max(i|(P_i \leq CWTP(T_i) < P_{i+1}))$. Profit from $j^{th} level$

$$\Pi_j = \int_{T_{j-1}}^{T_j} P_i D(n) dn \quad (6)$$

An intelligent seller will never want to fall in case C3 as then the j^{th} level is never used by the customers and thus has one less step to try and approximate the willingness to pay function. [We show later that profit decreases with decrease in the number of levels thus a seller needs to avoids redundant levels for profit maximization]. Thus S always want $n_j < T_j$ or in other words $P(T_j) > CWTP(T_j)$

It follows directly from the equation(4)and Figure 3, that just by increasing the price of the j^{th} level to $CWTP(T_{j-1})$ the seller can increase his profit. Thus, in the case C1, S places itself sub-optimally if $P_j < CWTP(T_{j-1})$. In other words $n_j \geq T_{j-1}$.

Above arguments impose the constraint that $T_{j-1} \leq n_j < T_j$ on the critical point. In other words the j^{th} critical point should always fall within the j^{th} level window. Also for the customers looking for n transactions with $T_{j-1} \leq n < T_j$, but having $CWTP(n) < P_j$, will look to buy T_{j-1} transactions for P_{j-1},i.e. $i = j - 1$ in equation(5). Thus the buying function for the j^{th} level simplifies as

$$\begin{aligned} B_j(n) &= P_j && \text{if } CWTP(n) \geq P_j \\ &= P_{j-1} && \text{if } CWTP(n) < P_j \end{aligned} \quad (7)$$

and the three cases can be easily tied together as
$T_{j-1} \leq n_j < T_j$

$$\begin{aligned} \Pi_j &= \int_{n_{j-1}}^{n_j} B(n) D(n) dn \\ &= \int_{T_{j-1}}^{n_j} P_{j-1} D(n) dn + \int_{n_j}^{T_j} P_j D(n) dn \end{aligned} \quad (8)$$

Total profit from all the levels is given by:

$$\Pi_s = \sum_{j=0}^{m} (\Pi_j) \quad (9)$$

4.4 Consumer Segment C:

This section extends previously discussed models in light of properties of consumer segment C and determines the optimal prices in each case. Based on the willingness to pay customer segment can be viewed to be comprised of two types of customers: c-type and d-type. c-type customers have constant marginal WTP up to a fixed number of transactions.

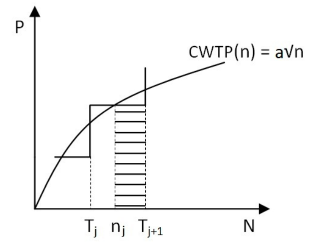

Figure 4: The figure shows the CWTP for a d-type customer under step pricing model

In other words, for c-type customers $CWTP_c$ is linear. d-type customers have declining marginal WTP i.e a sublinear $CWTP_d$. For example a financial analysis firm analyzing a particular dynamic dataset as its service, keeps continuously querying on it and has a constant WTP upto a fixed number of transactions(i.e. linear $CWTP_c$). On other hand a financial consultancy that works across different datasets has a declining marginal WTP for one particular dataset i.e sublinear $CWTP_d$.

Let k fraction of C be of d-type and remaining $(1 - k)$ be of c-type. Across this consumer segment C let n(desired number of transactions) be uniformly distributed over [0, 1],i.e. we normalize N to 1. To simplify calculations, for the remaining part of the section we will assume $D(n)$ to be uniformly distributed from [0, 1].

4.4.1 Unit pricing:

We look at the problem of maximizing the profit under 3 cases - when S can expect demand from both c-type and d-type buyers or from only c-type or from only d-type. The $CWTP(n)$ for each c-type customer is $w \cdot n$ where w could be interpreted as the average WTP for each transaction and n the desired number of transactions. For d type customers we assign a sublinear function to $CWTP_d(n)$ for further analysis, in particular we take $CWTP_d(n) = a\sqrt{n}$, as shown in Figure 4. Analysis is fairly general incase S finds another sublinear function as CWTP for its d-type customers.

Demand only from c-type: Thus no d-type customer buys any units, i.e. $P_u n > CWTP_d(n) \forall n \in [0, N = 1]$. Note for our particular choice of function $CWTP_d(n) = a\sqrt{(n)}$, this case cannot occur as $CWTP_d(\epsilon) > a\sqrt{(\epsilon)}$, as $\epsilon --> 0$. But we include this case for completion in case the $CWTP_d$ has some other form.

$$\Pi_u = \begin{cases} \int_0^{N=1} (1-k) P_u n dn = P_u \frac{(1-k)}{2} & \text{when } P_u \leq w \\ 0 & \text{when } P_u > w \end{cases}$$

Profit is maximum for $P_u^* = w$ and $\Pi_u^* = w\frac{(1-k)}{2}$

Demand only from d-type: Thus no c-type customers buy any units, i.e. $P_u > w$

Thus $\Pi_u = k \cdot \int_0^{n_1'} P_u \cdot n \cdot dn + k \cdot \int_{n_1'}^{1} P_u \cdot n_1' \cdot dn$

Here, $n_1 = 0, n_1' = \min(1, \frac{a^2}{P2})$.
Thus

$$\Pi_u = \begin{cases} \int_0^{\frac{a^2}{P_u^2}} kP_u n dn + \int_{\frac{a^2}{P_u^2}}^1 kP_u(\frac{a^2}{P_u^2})dn & \text{when } P_u > a \\ \int_0^1 kP_u n dn = \frac{kP_u}{2} & \text{when } P_u \le a \end{cases}$$

Solving $\frac{\partial \Pi_u}{\partial P_u} = 0$ gives the maximum profit(Π_u^*) and the optimal unit price(P_u^*). Note this is the maximum profit for the case $P_u > w$.

Demand from both c-type and d-type:

Thus for this case $P_u \le w$.

$$\Pi_u = \begin{cases} \int_0^{\frac{a^2}{P_u^2}} kP_u n dn + \int_{\frac{a^2}{P_u^2}}^1 kP_u(\frac{a^2}{P_u^2})dn + \int_0^1 (1-k)P_u n dn; \; P_u > a \\ \int_0^1 kP_u n dn + \int_0^1 (1-k)P_u n dn; P_u \le a \end{cases}$$

Solving $\frac{\partial \Pi_u}{\partial P_u} = 0$ gives the maximum profit(Π_u^*) and the optimal unit price(P_u^*).

***Proposition 1**: Maximum profit and optimum price for the unit pricing model is given by Π_u^* and P_u^*. For c-type customers, unit pricing achieves maximum profit by setting $P_u = w$.*

4.4.2 Step Pricing:

Demand only from c-type

i.e. no demand form d-type or $(P_j > a\sqrt{n})\forall j \in [1, 2.., m]$ The j^{th} critical point is given by solving $P_j = wn_j$ and P_j's are constrained as $(T_{j-1} \le \frac{P_j}{w} < T_j)$.

$$\Pi_j = \int_{T_{j-1}}^{\frac{P_j}{w}} (1-k)P_{j-1} dn + \int_{\frac{P_j}{w}}^{T_j} (1-k)P_j dn$$
$$= (1-k)\{P_{j-1}(\frac{P_j}{w} - T_{j-1}) + P_j(T_j - \frac{P_j}{w})\} \tag{10}$$

$$\Pi_s = \sum_{j=1}^{m} \Pi_j = \sum_{j=1}^{m} \{(1-k)\{P_{j-1}(\frac{P_j}{w} - T_{j-1}) + P_j(T_j - \frac{P_j}{w})\}\}$$
$$= \frac{(1-k)}{w}\{\sum_{j=1}^{m-1} P_j P_{j+1} - \sum_{j=1}^{m} P_j^2 + P_m w\} \tag{11}$$

Its interesting to observe that the profit from a site pricing is independent of the step size(T_j's) as long as $(wT_{j-1} \le P_j < wT_j)$ holds. Optimal P_j^* are obtained by solving m equations($\frac{\partial \Pi_s}{\partial P_j} = 0$) in m variables(P_j).

$$P_j^* = \frac{wj}{m+1}$$
$$\Pi_s^* = \frac{(1-k)wm}{2(m+1)} \tag{12}$$

Optimal profit increases with number of steps but number of steps are restricted to be small for obvious reasons.

Demand only from d-type:-

i.e. no demand form c-type or $(P_j > wn_j)\forall j \in [1, 2.., m]$ The j^{th} critical point is given by solving $P_j = a\sqrt{n}$ and P_j's are constrained as $(T_{j-1} \le \frac{P_j^2}{a^2} < T_j)$ (from **??**8).

$$\Pi_j = \int_{T_{j-1}}^{\frac{P_j^2}{a^2}} kP_{j-1} dn + \int_{\frac{P_j^2}{a^2}}^{T_j} kP_j dn$$
$$= kP_{j-1}(\frac{P_j^2}{a^2} - n_{j-1}) + kP_j(T_j - \frac{P_j^2}{a^2}) \tag{13}$$

$$\Pi_s = \sum_{j=1}^{m} \Pi_j = \sum_{j=1}^{m} k\left\{P_{j-1}(\frac{P_j^2}{a^2} - n_{j-1}) + P_j(T_j - \frac{P_j^2}{a^2})\right\}$$
$$= \frac{k}{a^2}\left\{\sum_{j=1}^{m-1} P_j P_{j+1}^2 - \sum_{j=1}^{m} P_j^3 + P_m a^2\right\} \tag{14}$$

Its interesting to observe that the profit from a site pricing from d-type customers is also independent of the step size(T_j's) as long as $(T_{j-1} \le \frac{P_j^2}{a^2} < T_j)$ holds. Optimal price levels P_j^*s and maximum profit Π_j^* is obtained by solving m equations($\frac{\partial \Pi_s}{\partial P_j} = 0$) in m variables(P_j).

for $m = 2$

$$P_1 = \frac{a}{\sqrt{9 - 2\sqrt{3}}} \quad P_2 = \frac{\sqrt{3}a}{\sqrt{9 - 2\sqrt{3}}}$$

for $m = 3$ ($\alpha = \sqrt{9 - 2\sqrt{3}}$)

$$P_1 = \frac{a}{\sqrt{\alpha(3\alpha - 2\sqrt{3})}}; P_2 = \frac{\sqrt{3}a}{\sqrt{\alpha(3\alpha - 2\sqrt{3})}}; P_3 = \frac{a\sqrt{\alpha}}{\sqrt{\alpha(3\alpha - 2\sqrt{3})}}$$

Demand from both c-type and d-type:-

i.e.$\exists j|(p_j < wn_j)$ for some demand from c-type and $\exists j|(p_j < k\sqrt{n}_j)$for some demand from d-type.

$$\Pi_s = \frac{(1-k)}{w}\left\{\sum_{j=1}^{m-1} P_j P_{j+1} - \sum_{j=1}^{m} P_j^2 + P_m w\right\} +$$
$$\frac{k}{a^2}\left\{\sum_{j=1}^{m-1} P_j P_{j+1}^2 - \sum_{j=1}^{m} P_j^3 + P_m a^2\right\} \tag{15}$$

Optimal price levels P_j^*s and maximum profit Π_j^* is obtained by solving m equations($\frac{\partial \Pi_s}{\partial P_j} = 0$) in m variables(P_j).

***Proposition 2**:Optimal profit for a firm selling to c- and d-type customers using the step-pricing scheme is independent of the step size(T_j's) as long as price of each level obeys the constraint of profit maximization $\left((T_{j-1} \le \frac{P_j}{w} < T_j) \text{ for c-type and } (T_{j-1} \le \frac{P_j^2}{a^2} < T_j) \text{ for d-type} \right)$. The maximum profit Π_s^* is obtained by setting price levels as P_j^*.*

4.5 Duopoly Competition:

In this section, we determine the optimal pricing strategies in light of competition. Two firms selling datasets D1 and D2 are considered competitors when D1 and D2 have overlapping views and thus overlapping target consumer segment. We model duopoly as two stage game where the sellers determine the pricing scheme in the first stage and their optimal prices in the second stage. This formulation is appropriate as decision about the pricing scheme require substantial attention from senior executives, lengthy experimental trials, advertising strategies, marketing strategies, accounting and other infrastructure considerations. On other hand, tweaking pricing parameters of a pricing model, in response to market dynamics, is relatively easier.

In the first stage, each seller selects from three options among pricing schemes: (i) unit-pricing scheme,(ii) site-pricing scheme and (iii) both pricing schemes.

Lemma 1*: When each firm offers one and the same pricing scheme, then the Bertrand equilibrium is obtained thus equilibrium prices and profits are equal to zero.*

When both firm offers the same pricing scheme in the first stage then in the second stage the pricing game reduces to a Bertrand game. Buyers purchase from the firm that offers the lower price. Firms keep on undercutting its competitors price to capture consumer segment and at equilibrium, no firm can have positive price as it gives undue advantage to its competitor. Thus Bertrand equilibrium is obtained. In case of step pricing the subgame(in second stage) at each level reduces to a Bertrand game. Any positive level price cant be sustained in equilibrium as the firms have incentive to cut each others price by a small amount. Hence Bertrand equilibrium is obtained.

Lemma 2*: When one firm offers both pricing schemes, then irrespective of the choice of pricing scheme of the other firm, the Bertrand equilibrium is obtained.*

Let firm A offers both unit and step pricing scheme and firm B offers only unit pricing. In pricing subgame(in second stage) where firm A competed with firm B over unit pricing scheme, leads to a bertrand game in unit pricing. Both the firms are forced to reduce unit prices to zero(Lemma 1)and no customer buys from step pricing scheme with positive price levels. This leads to Bertrand game. The same argument follows in case where firm B offers step pricing and firm A offers both pricing scheme.

Proposition3*: Pareto-dominant pure strategy equilibrium in the duopoly game: One seller offers the unit-pricing scheme and the other seller offers the site-pricing scheme (prices in Duopoly equilibrium)*

The following table illustrates Proposition 3 in detail, determining the profit of each firm in nine potential scenarios.

Adopting Pricing Schemes (PS) for Profit			
Firm B/Firm A	Unit PS	Site PS	Both PS
Unit PS	$(0,0)$	(Π_u, Π_s)	$(0,0)$
Unit PS	(Π_s, Π_u)	$(0,0)$	$(0,0)$
Unit PS	$(0,0)$	$(0,0)$	$(0,0)$

In summary, we derive optimal price(P^*) and optimal profit(Π^*) for a data seller entering the datamarket, for both unit pricing model and step pricing model. In absence of competition, the dataseller should compare the profit from both the models first and then select that pricing model which yields maximum profit as per the composition of its target consumer segment. We show that for the unit pricing scheme, setting your per unit price equal to the willingness to pay for c-type customers maximizes the profit. We also show that for step pricing scheme, optimal profit is independent of the step sizes as long as price of each level obeys the constraint of profit maximization. For step pricing, increasing the number of steps increase the profit. Under duopoly competition, we provide guidelines on how a firm should choose its pricing model.

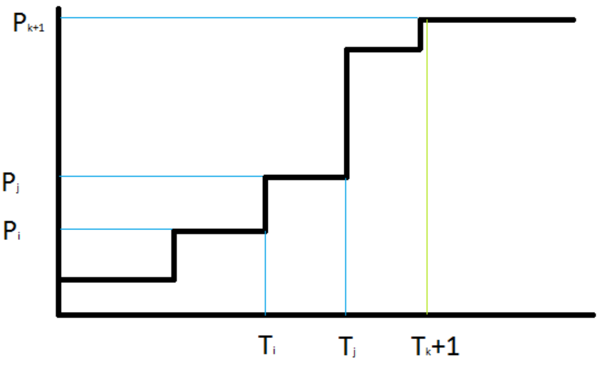

Figure 5: The Figure shows the presence of arbitrage, when $T_k + 1$ can be generated by using lower levels, i and j, s.t. $P_i + P_j < P_{k+1}$ and $T_i + T_j > T_k + 1$

In case, the data seller has already entered the market, he may not be in a position to redefine his entire pricing model as per the above framework. In the following section we provide guidelines for him to figure out if his present pricing is Arbitrage free. If not, we help him to come up with a pricing model very close to his present model and free from Arbitrage.

5. VALID PRICING STRUCTURE

We find that most price structures currently in Azure DataMarket are in the form of subscription pricing with multiple levels, i.e. of the form $(P_1, T_1), (P_2, T_2), ..., (P_n, T_n)$ where $P_1 \le P_2 \le P_3.. \le P_n$ and $T_1 \le T_2 \le T_3.. \le T_n$. This structure implies that the customer can buy upto T_k transactions by paying P_k. We find that these step functions do obey the diminishing returns property in a weak way. In particular $\frac{P_1}{T_1} \ge \frac{P_2}{T_2} \ge \frac{P_3}{T_3} ... \ge \frac{P_n}{T_n}$,
i.e. the price per transaction comes down as the number of transactions bought increases. Note this is ensured only at the end points of the structure, i.e. if the customer wants to buy exactly T_i transactions. But, due to the presence of discrete steps the function is no longer concave and is susceptible to arbitrage, i.e. the customer looking for T transactions with $T_j < t < T_{j+1}$, can in certain cases actually fulfill his requirement by paying less than P_{j+1}. Formally if $O(t)$ is the least price to get t transactions and $I(t)$ be the price I by the price structure,i.e. $I(t) = P_{j+1}$ if $T_j < t < T_{j+1}$, then $O(t) < I(t)$, as shown in Figure 5

Note that the inherent property of a step function that makes it non-concave would imply that a customer would be able to arbitrage this structure if he were allowed to buy a fractional number of such contracts. However the natural restriction on buying an integral number of contracts leads to a set of conditions that can ensure that even though the structure isn't concave in a strict sense, it is arbitrage-free. Note one application where such an analysis would be useful is in determining the total sales given the demand function $D(t)$, i.e. the number of customers requiring t transactions for each t. If such a demand function were known(or estimated), the profit would be expected to be the integral of the product of $D(t)$ and the price buying $I(t)$ i.e.

$\Pi = \int D(t) \cdot I(t) dt$.

However as this structure is susceptible to arbitrage the actual profit would only be

$\Pi = \int D(t) \cdot O(t) dt$,

assuming that the customers would take advantage of arbitrage, if possible. Thus ensuring that

$O(t) = I(t) \forall t$

allows the seller to make an accurate estimate of the sales. We call such a structure valid.

5.1 Conditions for Validity

Here we derive sufficient and necessary conditions on a pricing structure to ensure that $O(t) = I(t)\ \forall t$, and give an $O(n^2)$ algorithm to check for the validity of any arbitrary function.

Claim: Pricing structure is valid if $P_{k+1} \leq \min_{1 \leq j \leq k} P_j + I(T_k + \epsilon - T_j) \forall k = \{1, 2, ...n-1\}$, $\epsilon - -> 0$

Proof: Let the conditions be true. We will use induction on the number of steps to show this.

Base: $k = 1$, There is no way to arbitrage the first level, as $P_1 \leq P_k \forall k$

Induction Hypothesis: Assume that the price structure is valid upto the k^{th} level,i.e. for $O(t) = I(t) t \leq T_k$

Induction Step: We need to show that the pricing structure is valid upto the $k + 1^{th}$ level, i.e. $O(t) = I(t)$, for $T_k < t \leq T_{k+1}$.

We first show that $O(t) = I(t)$ for $t = T_k + \epsilon$, by contradiction.

Suppose $O(t) < I(t)$ for $t = T_k + \epsilon$. Then $O(t)$ must be using atleast one of the subscription levels $\leq k$. Let it use level j. Then P_{k+1}

$= I(t)$

$> O(t)$,(asumption)

$= P_j + O(t - T_j)$,(Optimality)

$\geq \min_{1 \leq j \leq k} P_j + O(t - T_j)$,

$\geq \min_{1 \leq j \leq k} P_j + I(t - T_j)$,(IH)

which contradicts our condition. Thus we get that $I(t) \leq O(t)$ for $t = T_k + \epsilon$.

We now use this to show that the validity in the rest of the interval follows. Consider any $T_k < t' \leq T_{k+1}$, Note $t' > t$. Thus $O(t') \geq O(t) = I(t) = P_{k+1} = I(t')$. Thus the structure is valid for all t', $T_k < t' \leq T_{k+1}$, which completes the proof. This proves that these conditions are sufficient to ensure validity of the pricing structure. The necessity part is obvious-clearly if any of the conditions doesn't hold, then that gives us a way to arbitrage the structure. Thus we can derive an algorithm to check for validity of the structure by just checking these conditions. To check the k^{th} level we need to $O(k)$ time. Thus to check for all n levels we require $O(n^2)$ time.

Proposition 4: A multi-step pricing function (P,T) is valid iff $P_{k+1} \leq \min_{1 \leq j \leq k} P_j + I(T_k + \epsilon - T_j) \forall k = \{1, 2, ...n-1\}$, $\epsilon - -> 0$

5.2 Dual Conditions for Validity

By ensuring that the price cannot be arbitraged for each $t = T_k + \epsilon$, we derive these conditions that are linear in the pricing levels P. We can derive similar conditions by looking at the dual problem,i.e. given a price $p_k - \epsilon$, can we get more

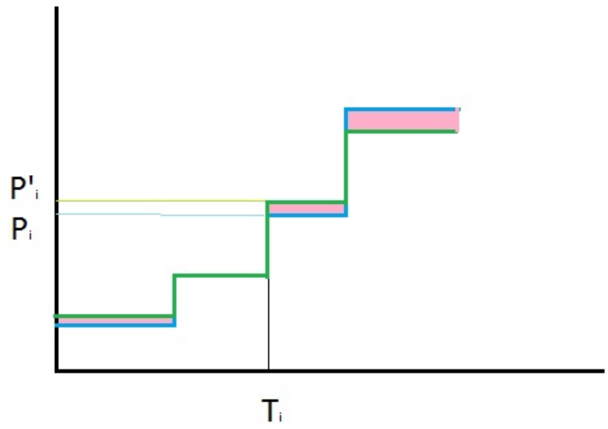

Figure 6: The Figure shows the Original Pricing Structure (P,T) in blue and the new Pricing Structure (P',T) in green. The distance between the 2 functions is defined as the integral of the difference of the functions, i.e. the sum of the areas indicated in pink

that T_k transactions. Following a similar analysis we get the following conditions. Let $IT(p) = T_j$ for $P_j \leq p < P_{j+1}$, i.e. the number of transactions that the structure implies for a given price p. We get the following sufficient and necessary conditions on the structure to ensure validity

$T_k \geq \max_{1 \leq j \leq k-1} T_j + IT(p_k - p_j - \epsilon)\ \forall k = \{2, 3...n\}, \epsilon - - > 0$.

As the intuition is similar to the previous result, we push back the proof to the appendix(A.1).

Proposition 5: A multi-step pricing function (P,T) is valid iff $T_k \geq \max_{1 \leq j \leq k-1} T_j + IT(p_k - p_j - \epsilon)\ \forall k = \{2, 3...n\}, \epsilon - - > 0$.

5.3 Optimal Pricing Structure

In the previous section we derived conditions that must hold for a valid pricing structure. Here, given an invalid structure denoted by (P, T) where $P = \{P_1, P_2...P_n\}$ and $T = \{T_1, T_2...T_n\}$, we look for structures (P', T) or (P, T') which are valid and close to the given structure under a suitable distance measure.

Let us first look at the case when we solve the prices keeping the transaction levels constant, i.e. we vary P keeping T constant. We look to minimize the objective function $\sum_i |P'_i - P_i| \cdot |T_i - T_{i-1}|$. Note this minimizes the sum of the areas of the differences of the rectangles as shown in Figure 6. This is exactly the L1 norm of the difference of two function f and g, which is $\int |f - g| dx$. Thus we get the following problem. For convenience we take $T_0 = P_0 = 0$.

$\min \sum_i |P'_i - P_i| \cdot (T_i - T_{i-1})$

s.t. (P', T) is valid.

The linearity of the objective function and the conditions reduce this to the following linear program in P'.

$\min \sum_i |P'_i - P_i| \cdot (T_i - T_{i-1})$

s.t. $P'_{k+1} \leq P'_j + I(T_k + \epsilon - T_j)\ 1 \leq j \leq k-1, k = \{1, 2...n-1\}$

$P'_1 \leq P'_2 \leq ...P'_n$

Note that the function I is only dependent on T and returns a price for a level, so some P'_i. Note the absolute value in the objective function can be removed, by using the following

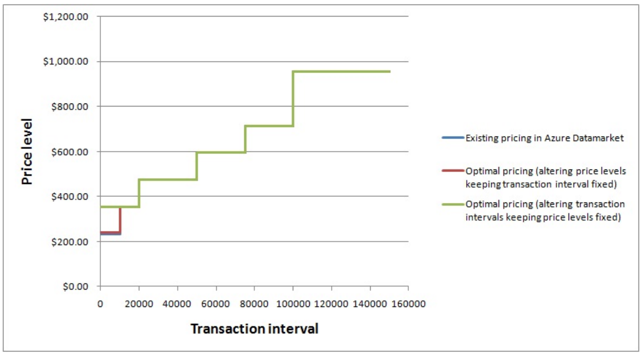

Figure 7: The Optimal Valid Pricing Structures(varying price level or transaction levels) and the original Pricing Structure for Alteryx LLC - Geography Search Service with Geocoding

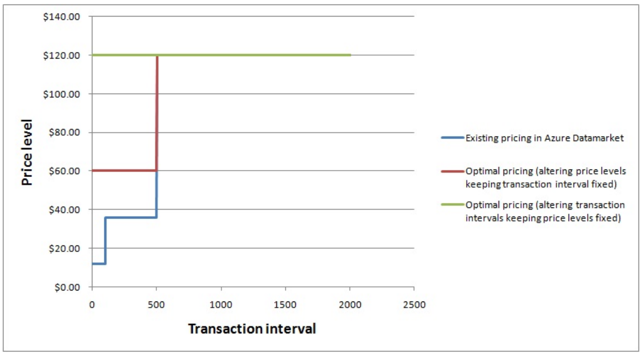

Figure 8: The Optimal Valid Pricing Structures(varying price level or transaction levels) and the original Pricing Structure for AWS Convergence Technologies Inc. (Weather Bug) - Historical Observations

common substitution, where $|x - k|$ in the objective function can be replaced by a pseudo variable t and adding the constraints $t < x - k$ and $t < k - x$

Following a similar derivation, we can express the dual problem of getting a step function (P, T') as a linear program in T'. The program is given by

$$\min \sum_i |T'_i - T_i| \cdot (P_i - P_{i-1})$$
$$\text{s.t. } T'_k \geq T'_j + IT(P_k - P_j - \epsilon)\ 1 \leq j \leq k-1 k = \{2, 3...n\}$$
$$T'_1 \leq T'_2 \leq ...T'_n$$

Note that the function IT is only dependent on P and returns transactions for a price, so some T'_i.

5.4 Results on Azure Datasets

We show the results of running the linear program on two datasets from Azure DataMarket in Figure 7 and Figure 8. We use MATLAB to perform the linear program. Appendix A.3 shows the results of the algorithms on some of the priced datasets present in Azure DataMarket. In the first Arbitrage-free pricing, the price level is adjusted keeping the transaction limit as fixed. In many datasets such as Business Lookup, and Weather Imagery the lower levels are clubbed together as a single step. Small price adjustments are made in the lower steps for datasets such as MLB year-to-date and Demographics dataset. In the second Arbitrage-free pricing, the transaction limit is adjusted keeping the price level as fixed. In many datasets, some of the transaction intervals is zero such as Business Lookup, and Weather Imagery. This is the corollary to the previous arbitrage-free pricing in that these steps are dropped to remove arbitrage.

6. CONCLUSIONS AND FUTURE WORK

After studying the current pricing models of information goods and those in place for datamarkets, the paper brings out the issues with various pricing models. Pricing on the number of units rather than on content-time is desirable for pricing datasets. Adopting the pricing model in Windows Azure Marketplace DataMarket, the paper considers ways to price the data in a more organized way. The paper begins by reflecting on the issues with the pricing model in terms of the disconnect with the target customer market segments and their buying power; the need to account for competing data sellers who provide similar datasets with either overlapping records or other value additions; and the need to avoid arbitrage situations. The paper sets up the Economics framework based on CWTP and computes the optimal prices for both unit pricing model and step pricing model. This could be leveraged by new datasets yet to be published on datamarkets. For existing datasets, the necessary and sufficient conditions for a valid pricing scheme are determined and an $O(n^2)$ algorithm is formulated to validate multi-step pricing models. In case the pricing models has arbitrage, a linear program id formulated to compute the closest valid step pricing under L1 norm. Future research can develop properties to identify competing datasets in the data market and provide mechanism to determine optimal design level, time to market and optimal prices for profit maximization. Differential pricing could be used for business or research data or certain important attributes/tuples that may be flagged. Future work could also concentrate on removing the restrictions on keeping either the tuple level or price levels constant.

7. REFERENCES

[1] Amazon relational database services (rds), ref http://www.amazon.com/rds/.

[2] Amazon simple db, http://www.amazon.com/simpledb/.

[3] Google app engine datastore, http://code.google.com/appengine/docs/datastore.

[4] Microsoft azure data market, https://datamarket.azure.com/, microsoft codename dallas whitepaper.

[5] Microsoft sql azure, http://www.microsoft.com/windowsazure/sqlazure.

[6] Pricing terms in windows azure marketplace datamarket https://datamarket.azure.com/help/b4789752-4aa9-4a57-ad05-b245be5ea336.

[7] Windows azure storage services rest api, ref http://msdn.microsoft.com/en-us/library/dd179355.aspx.

[8] M. Balazinska, D. Suciu, and P. Upadhyaya. How to price shared optimizations in the cloud.

[9] B. Birnbaum and A. Jaffe. Relational data markets. In *In Techinical Report, CSE544 Class Project*, 2007.
[10] V. Choudhary. Use of pricing schemes for differentiating information goods. In *INFORMS*, pages 1–15, 2009.
[11] S. Jain and P. K. Kannan. Pricing of information products on online servers: Issues, models, and analysis. In *Managament Science*, pages 1123–1142, 2002.
[12] T. Klastorin and W. Tsai. New product introduction. manufacturing & service operations management. In *INFORMS*, pages 302–320, 2004.
[13] C. J. Savage and A. J. Vickers. Empirical study of data sharing by authors publishing in plos journals. In *PLoS ONE*.
[14] M. Stonebraker, P. M. Aoki, W. Litwin, A. Pfeffer, A. Sah, J. Sidell, C. Staelin, and A. Yu. Mariposa: A wide-area distributed database system. In *VLDB Journal: Very Large Data Bases*, pages 48–63, 1996.

8. APPENDIX

A.1 Proof of the Dual Conditions For Validity

Let OT(p) be the maximum number of transactions that can be derived for a given price p.

Claim: Pricing structure is valid if $T_k \geq max_{1\leq j\leq k-1} T_j + IT(p_k - p_j - \epsilon)\ \forall k = \{2, 3...n\}, \epsilon - - > 0$

Proof: Let the conditions be true. We will use induction on the number of steps to show this.

Base: $k = 1$, There is no way to arbitrage the first level, as $T_1 \leq T_k \forall k$

Induction Hypothesis: Assume that the price structure is valid upto the k^{th} level,i.e. for $OT(p) = IT(p) p < P_k$

Induction Step: We need to show that the pricing structure is valid upto the $k + 1^{th}$ level, i.e. $OT(p) = IT(p)$, for $P_k \leq t < P_{k+1}$.

We first show that $OT(p) = IT(p)$ for $p = p_{k+1} - \epsilon$, by contradiction.

Suppose $OT(p) > IT(p)$ for $p = p_k - \epsilon$. Then $OT(t)$ must be using atleat one of the subscription levels $\leq k$. Let it use level j. Then T_k

$= IT(p)$
$< OT(p)$,(asumption)
$= T_j + OT(p - P_j)$,(Optimality)
$\leq max_{1\leq j\leq k} P_j + OT(p - P_j)$,
$\leq max_{1\leq j\leq k} P_j + IT(p - P_j)$,(IH)

which contradicts our condition. Thus we get that $IT(p) \leq OT(p)$ for $p = P_{k+1} - \epsilon$.

We now use this to show that the validity in the rest of the interval follows. Consider any $P_k < p' \leq P_{k+1}$, Note $p' < p$. Thus $OT(p') \leq OT(p) = IT(p) = T_k = IT(p')$. Thus the structure is valid for all P', $p_k < p' \leq P_{k+1}$, which completes the proof.

This proves that these conditions are sufficient to ensure validity of the pricing structure. The necessity part is obvious-clearly if any of the conditions doesn't hold, then that gives us a way to arbitrage the structure. Thus we can derive an algorithm to check for validity of the structure by just checking these conditions. To check the k^{th} level we need to $O(k)$ time. Thus to check for all n levels we require $O(n^2)$ time.

A.2 Some Original Pricing Structures in Azure DataMarket

Datasets from Windows Azure Marketplace Datamarket

Company	Dataset	Transaction limit	Price level
Wolfrom Alpha LLC	Facts	5000	$0.00
Alteryx LLC	Consumer Expenditure Data	Unlimited 150 100 50 25 10	$238.80 $178.80 $154.80 $118.80 $94.80 $58.80
CDYNE Corporation	National Death Index	1000000	$396.00
Dun & Bradstreet	Business Lookup	10000 5000 2000 1000	$1,800.00 $900.00 $360.00 $180.00
Dun & Bradstreet	Corporate Linkage	1000	$4,920.00
Dun & Bradstreet	Enterprise Risk Management	64	$4,920.00
StrikeIron	Sales and Use Tax Rates Complete	50000 10000 1000	$3,300.00 $900.00 $150.00
Wolters Kluwer	CCH CorpSystem Sales Tax Rates	50000	$333.00
PracticeFusion	Medical Research Data	Unlimited	$0.00
United Nations	UNAIDS	Unlimited	$0.00
United Nations	WHO Data	Unlimited	$0.00
Alteryx LLC	Geography Search Service with Geocoding	Unlimited 250000 100000 75000 50000 20000 10000	$1,194.00 $954.00 $714.00 $594.00 $474.00 $354.00 $234.00
StrikeIron	US Address Verification	100000 25000 5000	$1,368.00 $504.00 $150.00
GTW Holdings	Singapore Points of Interest	Unlimited	$0.00
United Nations	World Tourism Organization Statistics - Database & Yearbook	Unlimited	$0.00
Zillow Inc	Home Valuation	30000	$0.00
Zillow Inc	Mortgage Information	30000	$0.00
STATS LLC	MLB game-by-game	Unlimited	$9.54
STATS LLC	MLB live scores	Unlimited	$21.54
STATS LLC	MLB year-to-date	500 100 50	$120.00 $24.00 $12.00
Govt. of USA	Data.gov	Unlimited	$0.00
ESRI	2010 Key US Demographics by ZIP code, place, county	150 100 50 25 10	$49.95 $39.95 $24.95 $19.95 $9.95
Weather Central LLC	Super MicroCast Forecast Data	1000000 100000 10000 2500	$2,400.00 $600.00 $120.00 $0.00
Weather Central LLC	Weather Imagery	1000000 100000 10000 2500	$2,400.00 $600.00 $120.00 $0.00
AWS Convergence Technologies	Historical Observations	2000 500 100	$120.00 $36.00 $12.00

A.3 Arbitrage-free Pricing For Invalid Datasets from Azure DataMarket

Datasets from Windows Azure Marketplace Datamarket					
Company	Dataset	Arbitrage-free Pricing 1		Arbitrage-free Pricing 2	
		Keeping Transaction limit fixed	Corrected Price level	Corrected Transaction limit	Keeping Price level fixed
Alteryx LLC	Consumer Expenditure Data	Unlimited 150 100 50 25 10	$238.80 $178.80 $154.80 $118.80 $94.80 $60	Unlimited 150 100 50 25 0	$238.80 $178.80 $154.80 $118.80 $94.80 $58
Dun & Bradstreet	Business Lookup	10000 5000 2000 1000	$1,800.00 $900.00 $900.00 $900.00	10000 5000 0 0	$1,800.00 $900.00 $360.00 $180.00
StrikeIron	Sales and Use Tax Rates Complete	50000 10000 1000	$3,300.00 $1,650.00 $1,650.00	50000 0 0	$3,300.00 $900.00 $150.00
Alteryx LLC	Geography Search Service with Geocoding	Unlimited 250000 100000 75000 50000 20000 10000	$1,194.00 $954.00 $714.00 $594.00 $474.00 $354.00 $240.00	Unlimited 250000 100000 75000 50000 20000 0	$1,194.00 $954.00 $714.00 $594.00 $474.00 $354.00 $234.00
StrikeIron	US Address Verification	100000 25000 5000	$1,368.00 $684.00 $684.00	100000 0 0	$1,368.00 $504.00 $150.00
STATS LLC	MLB year-to-date	500 100 50	$120.00 $62.88 $57.12	500 0 0	$120.00 $24.00 $12.00
ESRI	2010 Key US Demographics by ZIP code, place, county	150 100 50 25 10	$49.95 $39.95 $24.95 $19.95 $15	150 100 50 25 0	$49.95 $39.95 $24.95 $19.95 $9.95
Weather Central LLC	Super MicroCast Forecast Data	1000000 100000 10000 2500	$2,400.00 $1,200.00 $1,200.00 $1,200.00	1000000 0 0 0	$2,400.00 $600.00 $120.00 $0.00
Weather Central LLC	Weather Imagery	1000000 100000 10000 2500	$2,400.00 $1,200.00 $1,200.00 $1,200.00	1000000 0 0 0	$2,400.00 $600.00 $120.00 $0.00
AWS Convergence Technologies	Historical Observations	2000 500 100	$120.00 $60.00 $60.00	2000 0 0	$120.00 $36.00 $12.00

*The Unlimited transaction value was approximated as 10000000 during evaluation.