

Constructing Quality-Adjusted Price Indices from Revenue and Cost Data[†]

▪ SERGIO AQUINO DESOUSA^{*}

RESUMO

Este artigo demonstra como construir um índice de preços ajustados pela qualidade sem a observação direta de dados ao nível de produto (preços, quantidades e características). A técnica aplicada no artigo permite medir a inflação dos preços com base na variação intertemporal de bem-estar através do uso de dados de receita e despesa comumente disponíveis (pelo menos para o setor industrial) ao nível de empresas. No entanto, tal técnica exige a imposição de hipóteses sobre a evolução da qualidade do bem externo assim como a estrutura da demanda e da oferta. Com dados sobre a indústria colombiana de cerveja e combinando metodologias originalmente desenvolvidas por Katayama, Lu e Tybout (2003), DeSouza (2006a) e Trajtenberg (1990) estimam-se os parâmetros da demanda e constroem-se índices de preços para o período de 1977 a 1990 a partir da mensuração do bem-estar dos consumidores.

PALAVRAS-CHAVE

índice de preços, inovação, modelos de escolha-discreta

ABSTRACT

This paper shows how to construct quality adjusted price indices without direct observation of product-level data (prices, quantities and characteristics). The technique used here allows for a welfare based measurement of price change using commonly available (at least for the manufacturing sector) plant-level data on revenue and cost. However, one has to be explicit about the evolution of the outside good quality and the structure of demand and supply. Using data on the Colombian beer industry and combining the methodologies originally proposed by Katayama, Lu and Tybout (2003), DeSouza (2006a) and Trajtenberg (1990) I am able to uncover the demand parameters and build welfare-based price indices for the 1977-1990 period.

KEY WORDS

price index, innovation, discrete-choice demand models

JEL CLASSIFICATION

E31, L66, O33

[†] I would like to thank James Tybout, Mark Roberts and two anonymous referees for their useful comments. I am also thankful to CAPES and FUNCAP for providing financial support to this work.

^{*} Professor Adjunto do Curso de pós-graduação em Economia - CAEN/UFC e do Departamento de Teoria Econômica da Universidade Federal do Ceará - Av. da Universidade, 2700, 2º andar - Fortaleza - Ceará. Tel/Fax: 55-85-4009-7751. E-mail: sergiodesouza@caen.ufc.br.

(Recebido em julho de 2005. Aceito para publicação em março de 2007).

I. INTRODUCTION

Most statistical institutions report price indexes calculated from a weighted average of prices. A better way to measure price changes, known as the hedonic price approach, consists of regressing observed prices (in logs) on product characteristics, to control for quality variation, and time dummies. These dummies are interpreted as the increase in price net of quality improvements. The hedonic strategy, however, has its own limitations, as it places great demands on the data set. The statistical agency has to define what characteristics are relevant in determining prices. The agency also has to observe (or collect) prices and characteristics of many products offered in the market. With the development of discrete choice models another technique has been proposed (Trajtenberg, 1990). The basic idea consists of uncovering the true price change from the variation in consumer surplus which is calculated after estimating the parameters of a discrete-choice demand system. The strategy goes as follows: *i.* setup a theoretical model of consumer choice¹, *ii.* match the model's prediction about economic variables (e.g., prices and quantities) to their empirical counterparts to obtain the relevant parameters, *iii.* from these parameters, calculate the price variation that would have had the same welfare effect as the innovations that actually took place.²

Steps *i* and *ii* can be undertaken using traditional discrete-choice models found in the Industrial Organization literature (e.g. Berry, 1994). The original contribution of Trajtenberg comes from step *iii*. Using the structural approach outlined in the previous steps, he shows how to construct quality-adjusted price indices for a specific product.³ Berry's approach to estimate demand, and possibly supply, parameters places even greater demands on the data as not only prices but also quantities have to be observed at the product-level.

Detailed data set, as required by the hedonic and berry's approach, may be difficult or costly to obtain in many instances. Confidentiality issues and/or strategic purposes may induce some firms not to release the relevant data. On the other hand, official statistical agencies are usually more successful in the task of gathering information for plant-level surveys of the manufacturing sector. One of the reasons may rely on the fact that, in many cases, firms are asked to report only sales revenue and input

1 Sometimes, modeling producers' choices is useful to undertake step *ii*.

2 A notorious example of such approach can be found in Nevo (2003). He uses discrete-choice demand parameters, estimated from cross section of product level data, to compute welfare-based price indices.

3 The combination of welfare measures and the discrete-choice framework to calculate price indices can also be found in Timmins (2006). His methodology is developed to calculate an index that measures the cost of living, i.e. the cost of consuming a basket of products. He uncovers this "true" cost indirectly from a discrete-choice model of optimal residential location. Therefore, he does not need to specify which products compose this basket. This is an advantage in the context of cost-of-living indices, but it does not allow for the calculation of a product-specific price index, which constitutes the goal of this paper.

expenditures, rather than (more revealing) strategic information on prices and quantities. Such surveys are commonly available for many countries and cover most industries in the manufacturing sector, providing an easily accessible source for applied researchers.

The data set used here falls within the class of data sets discussed in the last paragraph. It reports only plant-level revenue and total cost instead of price and quantities, hampering initial attempts to undertake steps *i* and *ii* according to the standard approach found in Berry's work. To overcome this restriction, Katayama, Lu and Tybout (2003), KLT henceforth, builds on berry's work and devise an econometric methodology to uncover the demand parameters set as well as marginal costs, quality, prices and quantities from plant-level data that report only revenue and cost data. A similar methodology is developed by DeSouza (2006a). However, instead of estimation through econometric techniques, the methodology found in DeSouza (2006a) takes advantage of the extra information provided by aggregate data to calibrate the relevant parameters and variables. Although it may be difficult to obtain detailed data on quantities at the plant level, the same is not true for aggregate variables. For instance, in the beverage sector, the amount of beer, in liters, consumed in a given year is widely available for many countries. Aggregate quantities also carry information on demand parameters and therefore may help in determining their magnitudes.

This paper shows how to construct quality-adjusted price indices when only revenue and cost data are observed. Indeed, by applying the methodologies mentioned above to the Colombian beer market, it is possible to uncover the demand parameters and construct welfare-based price indices. This paper is organized as follows. The second section introduces the demand discrete-choice model, the consumer surplus function and a model of firms' supply. This section also includes the description of the empirical strategies mentioned above. The ensuing section comments on the beer market idiosyncrasies and presents the demand parameters estimates. The fourth section discusses the methodology to uncover the quality adjusted price indices, introduces consumer surplus identification issues, and presents the index numbers. Finally, the last section adds a few remarks and suggestions for future work.

II. THEORETICAL MODEL AND EMPIRICAL STRATEGIES

In this section I shall lay out the model used in this investigation. The demand for a product is aggregated from individual choices over a finite set of differentiated products using the familiar Nested Logit setting. In turn, the supply relation is assumed to be governed by profit maximizing firms playing a Bertrand game (for convenience, time subscripts are deleted in this section).

Demand

Consumers rank products according to their characteristics and prices. There are $N+1$ choices in the market, N inside goods (domestic varieties) and one outside good (imported variety). Consumer i chooses a good j , given price p_j , (unobserved) characteristics ξ_j , and unobserved idiosyncratic preferences. Products are grouped into two nests, indexed by g , which takes the values of 0 or 1. The first nest ($g=0$) contains only the outside good (imported variety in this application) whereas the second ($g=1$) contains the N inside goods (domestic varieties).

Then, for product j belonging to nest g define utility of consumer i as

$$u_{ij} = \delta_j + \zeta_{ig} + (1 - \sigma)\varepsilon_{ij} \quad \text{for } j=0,1,\dots,N \quad (1)$$

where, δ_j is the mean utility, which is given by the sum of a scalar summarizing unobserved product characteristics⁴ (ξ_j) and income disutility ($-\alpha p_j$), i.e., $\delta_j = \xi_j - \alpha p_j$. The first random term (ζ_{ig}) on the right hand side (RHS) of the equation above is a common shock to all products in nest g and its distribution depends on the parameter σ ($0 \leq \sigma < 1$). As σ approaches zero the within correlation of utilities within each nest decreases, and as σ approaches one the within correlation increases. The second random term ε_{ij} is identically and independently distributed extreme value. McFadden (1981) shows that we can integrate out $\zeta_{ig} + (1 - \sigma)\varepsilon_{ij}$ to obtain a closed form solution for the within aggregate market shares (sw_j) for an inside good as follows

$$sw_j = \frac{\exp[(\delta_j - \delta_0) / (1 - \sigma)]}{\sum_{k=1}^N \exp[(\delta_k - \delta_0) / (1 - \sigma)]}, \quad j=1,2,\dots,N$$

Here, δ_0 is the outside good mean utility ($\delta_0 = -\alpha p_0 + \xi_0$). The share of all domestic brands is given by $s_d = D / (D + 1)$, where $D = \left\{ \sum_{k=1}^N \exp[(\delta_k - \delta_0) / (1 - \sigma)] \right\}^{1-\sigma}$. Thus the market share for a domestic variety s_j is given by

$$s_j = \left(\frac{\exp[(\delta_j - \delta_0) / (1 - \sigma)]}{\sum_{k=1}^N \exp[(\delta_k - \delta_0) / (1 - \sigma)]} \right) \cdot \left[\frac{D}{D + 1} \right] \quad (2)$$

4 Usually the mean utility includes a vector of observed characteristics. However, as plant-level data sets rarely report such data, these are excluded from the model.

which is simply the product of sw_j times s_d .

Further, taking the log-difference between s_j and s_0 the demand system takes the simple log-linear functional form

$$\ln s_j - \ln s_0 = \alpha.(p_j - p_0) - \sigma \ln s_{wj} + \xi_j - \xi_0 \quad (3)$$

Supply

Each firm f produces a subset F_f of the goods sold in this market and set prices in a Bertrand game to maximize the sum of profits Π_f , which is given by

$$\Pi_f = \sum_{j \in F_f} (p_j - mc_j).M.s_j \quad (4)$$

where M is the total market size and mc_j is the marginal cost of producing brand j . Then, it can be shown that the price p_j of any product j produced by firm f must satisfy the following equation

$$s_j + \sum_{r \in F_f} (p_r - mc_r) \frac{\partial s_r}{\partial p_j} = 0 \quad (5)$$

Note that (5) is flexible enough to accommodate different market structures. The first is the single firm product, in which the firm can only control the price of its unique brand. The second is the multi-product firm, in which the firm internalizes the price decision of all of its brands. The third is a monopoly firm who owns all brands available in the market. In this case, the Bertrand game degenerates to a single-agent decision problem.

Empirical Methods

If prices and quantities were available all the parameters of the model could be estimated through GMM using demand and supply side moments. Indeed, notice that, under some rearrangements, (3) provides a closed form for the demand system that can be solved, up to the model parameters, for the unobservable term $\xi_j - \xi_0$. The unobservable ($\xi_j - \xi_0$) can be combined with appropriate instruments (product characteristics) to estimate the parameters following the GMM approach found in Berry (1994). However, Berry's strategy can not be used here as product characteristics are not observed.

Fortunately, KLT, building on berry's work, show that commonly available information on revenue and total costs (not prices and quantities) can be used to uncover the relevant variables. Their algorithm- henceforth referred to as the transformation algorithm- goes as follows.

Note that firm j 's revenue (R_j) and variable cost⁵ (TC_j) can be written as $R_j = p_j \cdot q_j$, $TC_j = mc_j \cdot q_j$, where q_j represents firm j 's output. Thus, we can write the market share for firm j as $s_j = q_j / (Q + Q_0)$, where Q and Q_0 represent the total output produced by domestic firms and total imported quantity, respectively. Then, these three identities (revenue, cost and market share) together with the F.O.C (5) can be solved for quantity as a function of data ($\mathbf{R}, \mathbf{TC}, Q_0$) and the demand parameters (α, σ), where \mathbf{R} collects the revenue of all plants in the sample and \mathbf{TC} collects the costs of all plants in the sample.

Similarly, we can retrieve $mc_j = mc_j(\alpha, \sigma, \mathbf{R}, \mathbf{TC}, Q, Q_0)$, $p_j = p_j(\alpha, \sigma, \mathbf{R}, \mathbf{TC}, Q, Q_0)$ and $q_j = q_j(\alpha, \sigma, \mathbf{R}, \mathbf{TC}, Q, Q_0)$. Thus, from $\sum_N q_j(\alpha, \sigma, \mathbf{R}, \mathbf{TC}, Q, Q_0) = Q$, we are able to solve for $Q = Q(\alpha, \sigma, \mathbf{R}, \mathbf{TC}, Q_0)$. Then, conditional on the models' parameters, we can retrieve prices, quantities and market shares. Thus, relative quality, defined as $a_{jt} = \xi_{jt} - \xi_{0t}$, can be determined from the demand system (3). To summarize, given $(\alpha, \sigma, \mathbf{R}, \mathbf{TC}, Q_0)$, the KLT transformation algorithm⁶ shows how to obtain firm level prices, marginal costs, relative quality and quantities as well as aggregate output (Q).

With the transformation algorithm in hand it is possible to determine the model parameters using two different (but closely related) approaches. The first one also comes from KLT. They assume a VAR process for the co-movements of a_{jt} and mc_{jt} and prior distributions on the parameters, updated by data according to the Bayesian rule to obtain the posterior distributions, from which inference is made. The second method, developed by DeSouza (2006a), relies heavily on the transformation algorithm laid out above, but does not require the imposition of a VAR. Shortly, it consists of matching the aggregate output implied by the structural model to its observed counterpart in order to calibrate the relevant parameter set.⁷ Appendix B discusses these two methods in more depth.

5 $TC_j = mc_j \cdot q_j$ as long as marginal costs are flat.

6 For more details, see Appendix A.

7 DeSouza (2006b) also uses data on aggregate output, but there it is employed to improve KLT's econometric (Bayesian) estimates, rather than calibrate them.

III. DATA AND PARAMETER ESTIMATES

The data set consists of an unbalanced panel of plants in the beer industry, with more than 10 employees, covering the period from 1977 to 1990. Originally, the data were gathered by the Colombia's *Departamento Nacional de Estadística* (DANE) and have been cleaned as described in Roberts (1996). The revenue series are constructed as the total sale revenue divided by a general wholesale price deflator. The total variable costs are defined as the sum of payments to labor, intermediate input purchases and energy purchases. Since some of the cost is incurred in the export activity we have to scale it by the ratio of total domestic sales to total sales and deflate the result by the same wholesale price deflator mentioned before. Note that prices and quantities are not directly observed. Therefore, the methodologies described in the second section have to be used.

From an additional source (the United Nations database) I obtain the quantity of beer (in hectoliters) produced in the country for the same sample period. Ideally, we would want to have data on the quantity of beer consumed in the country. However, the data in hand is not so restrictive since there is very little export activity in this sector.

I also use auxiliary data to uncover the price of the imported good (p_{0t})⁸ as well as its imported quantity in hectoliters (q_{0t}). In a separate publication DANE also reports the net weight (in kilos) and the monetary value of imports (in pesos). Assuming that beer has the same density as water (1kg per liter), it is easy to transform the net weight in kilos to volume of imported beer in hectoliters (q_{0t}). Then, p_{0t} follows from the ratio of the peso value of imports to q_{0t} .

While it is common for data sets to report plant-level revenue and cost data, they do not usually contain information on plant ownership. This is important for estimation since each ownership arrangement implies a different supply function and therefore, given $(\alpha, \sigma, R_{jt}, TC_{jt}, Q_{0t})$, different values for the unobserved variables (price, quantities, marginal cost and qualities). In the Colombian beer sector, however, ownership is not an issue since one Company (Bavaria S.A) controlled the non-imported beer market during the sample period studied here.

Indeed, after an aggressive horizontal merger strategy, Bavaria became the dominant firm in the beer production by acquiring its rivals (*Cerveceria Aguila*, *Cerveceria Union* and *Cerveceria Andina* and other smaller producers) in the beer business by the early seventies. Its dominance went unchallenged until 1995 when *Cerveceria Leona* entered the beer market as retaliation for the Bavaria entry in the soft drinks business, which

8 This is a composite good that bundles together all the different imported varieties.

was controlled by *Leona's* parent company. Since the data sample period ranges from 1977 to 1990 all the estimations presented below assume that a single firm owns all plants.

The parameters estimates are displayed below in table 1. Three different models are considered: the Nested Logit model ($0 < \sigma < 1$) and the Simple Logit model ($\sigma = 0$), both based on Bayesian techniques (NLB and SLB henceforth) and the Simple Logit model, implemented according to the calibration approach found in DeSouza (2006a)-this method will be referred to as CAL from now on).⁹

TABLE 1 – PARAMETERS ESTIMATES*

	SLB	NLB	CAL
α	3.115 (0.215)	2.738 (0.193)	3.046
σ		0.941 (0.010)	

* Standard deviation in parentheses.

When compared to the simple logit models, the NLB yields lower estimates for price disutility α and high estimates for the within value σ , which means that the set of the inside goods (domestic goods) is highly differentiated from the outside good (imported variety). Notice also that although α is allowed to vary over time in the CAL setting, its mean (3.04) is about the same magnitude as the price coefficients estimates for both Bayesian models.

IV. QUALITY ADJUSTED PRICE INDEX

The most popular method to obtain quality-adjusted price changes is given by the hedonic equation. Price is the dependent variable and product characteristics and time dummies are the RHS variables. The price change that is not explained by improvements in product characteristics is the true price variation. One problem with this approach consists of choosing the set of characteristics that are relevant to determine prices. There is always a high degree of discretion in setting up the hedonic regression. Different characteristics bundles may affect the index calculation. For instance, if gas mileage is omitted in the hedonic price regression for new cars, and a given car improves its fuel efficiency, the index would overstate the true change in prices, for this variable would be captured by the regression error term and not by

⁹ See Appendix B to check on the details of the estimation procedures that generate the results presented in table 1. DeSouza (2006a and 2006b) are also useful references.

the regressors. In some data sets, like the one used in this paper, not even product characteristics are observed.

Another methodology was devised by Trajtenberg (1990). The basic idea consists of uncovering the true price change from the variation in consumer surplus which is calculated from the parameters of a discrete-choice demand system. Below, I lay out his methodology in more detail.

Trajtenberg Method

Once the demand parameters of the discrete choice model are estimated, the consumer surplus for the Nested Logit model can be calculated from the following equation (McFadden, 1981)

$$CS_t(\vec{P}_t, \vec{\xi}_t) = \frac{1}{\alpha} \ln \left(\sum_{k=0}^{N_t} \exp[(\xi_{kt} - \alpha p_{kt}) / (1 - \sigma)] \right)^{1-\sigma} \quad (6)$$

where \vec{P}_t denotes the vector that comprises all prices p_{jt} at time t and $\vec{\xi}_t$ the vector that contains all the ξ_{jt} 's at time t . The consumer surplus for the simple logit model is calculated by simply setting $\sigma = 0$. The intertemporal variation in consumer surplus is then given by $\Delta CS_t = CS_t(\vec{P}_t; \vec{\xi}_t) - CS_{t-1}(\vec{P}_{t-1}; \vec{\xi}_{t-1})$.

McFadden (1981) also shows that ΔCS_t can be written as a difference between two expenditure functions ($\Delta CS_t = e(\vec{P}_{t-1}, \vec{\xi}_{t-1}) - e(\vec{P}_t, \vec{\xi}_t)$, where $e(\cdot)$ is the expenditure function). Then, the price index obtains after solving for ρ_t from the following equality¹⁰

$$\Delta CS_t = e(\vec{P}_{t-1}, \vec{\xi}_{t-1}) - e((1 - \rho_t)\vec{P}_t, \vec{\xi}_{t-1}) \quad (7)$$

The variable ρ_t , known as the welfare equivalent price reduction, measures the hypothetical average price reduction that would have had the same welfare effect as the innovations (or quality improvements) that actually took place. In other words, consumers would have been equally well off if they had been offered the old set of products at prices lower by a factor of ρ_t . This factor follows the same qualitative movements as ΔCS_t . For example, suppose nominal prices remain the same, then the larger is the increase in quality the larger is the price reduction ρ_t . Consumers would

¹⁰ Using the expenditure function to uncover price indices was originally proposed by Trajtenberg (1990).

demand a lower price to compensate for a worse set of products. For obvious reasons consumer surplus would also increase with higher quality.

Solving (7) involves a highly nonlinear equation. A simple solution was derived by Trajtenberg (1990). After assuming mild restrictions¹¹ on the average price change he shows that

$$\Delta CS_t = \rho_t \bar{p}_{t-1} \quad (8)$$

where \bar{p}_t is the unweighted average price across firms at period t . The price index (I_t) is then constructed by the simple formula $\frac{I_t}{I_{t-1}} = (1 - \rho_t)$ with $I_0=100$.

The unobservable ξ_j has an interesting interpretation that is particularly important in the context of price indices. It can be viewed as the unobserved quality of product j . Thus, unlike the hedonic price index, the welfare based index captures changes in omitted attributes. Also, since consumer surplus is a function of prices and attributes of all the goods available to consumers at a certain period t , the temporal variation of surplus provides a natural way to control for the entry of new products. On the other hand, the major limitation for the practical use of this method by statistical agencies hinges on the data requirement. It demands data on prices for virtually all firms active in the market, otherwise demand and supply could not be estimated properly. Obviously, gathering such a comprehensive data set is very costly making it difficult to consider such technique as a practical substitute for the hedonic price index. Fortunately, as show in the section II, KLT and DeSouza (2006a) offer methods to uncover the demand parameter set and pin down price and (relative) quality using commonly available firm level data on revenue and cost for firms in the manufacturing sector. But, before, constructing the price index it is important to discuss identification issues regarding the consumer surplus variation.

Identification Issues

The demand system only permits the identification of the relative movements in quality of the inside goods with respect to the outside good (recall that $a_j \equiv \xi_j - \xi_0$). Then, unless we postulate an assumption on the evolution of the outside good, it is

11 The price of each brand can always be written as $p_{it} = \bar{p}_t + \Delta p_{it}$, where \bar{p}_t is the unweighted average across brands. Assume now that the intertemporal price variation takes the form $p_{it} = (1 - \rho_t) \bar{p}_{t-1} + \Delta p_{it-1}$, i.e, the distribution moves leftward by a factor $(1 - \rho_t)$ but the variance remains the same. Then, (8) follows easily (see Trajtenberg, 1990, page 33, for the proof)

not possible to identify the intertemporal variation in consumer surplus (ΔS_t). To clarify this point note that, from (6), ΔCS_t can be rewritten as

$$\Delta CS_t = \frac{1}{\alpha} \left[\Delta \xi_{0t} - \alpha \Delta p_{0t} + (1 - \sigma) \ln \left(\frac{\sum_0^{N_t} \exp \left\{ \frac{a_{kt} - \alpha(p_{kt} - p_{0t})}{1 - \sigma} \right\}}{\sum_0^{N_{t-1}} \exp \left\{ \frac{a_{kt-1} - \alpha(p_{kt-1} - p_{0t-1})}{1 - \sigma} \right\}} \right) \right] \quad (9)$$

Once the demand parameters are determined from the techniques describe in section II, all terms but $\Delta \xi_{0t}$ can be identified. The lack of identification of $\Delta \xi_{0t}$ has quite different consequences for welfare variation.

For example, relative quality could go up due to different reasons. First, the inside good could become more attractive. Alternatively, the outside good attractiveness could fall instead. In the former case, welfare would increase while in the latter it would decrease. Thus, we have to be explicit about the assumptions on the evolution of the outside good. Below, I make an assumption that help to identify the absolute quality movements.¹²

Assumption: the quality of the outside good does not change over time, i.e, $\Delta \xi_{0t} = 0$ for all t .

This assumes away the identification problem by imposing that all the variation in relative quality is due to the variation in product appeal of the inside goods. The extent to which this assumption is appropriate will depend on the characteristics of each market. In some instances, especially in developing countries, the imported variety exhibits much more innovation than its domestic counterpart.

Calculating the Quality-adjusted price index

At this point, it is useful to summarize in a few steps how the methodologies presented in this paper can be combined to determine the quality-adjusted price index.

¹² I tried the following alternative assumption: **the quality of the outside good changes over time and there is a domestic variety whose quality does not change.** The evolution of $\Delta \xi_{0t}$ is determined as follows. First, pick the smallest plant ranked by the number of employees. The corresponding firm, indexed here as firm k , is then selected to be the one whose quality is not varying through time which makes it possible to pin down $\Delta \xi_0$ as being equal to $-\Delta a_{kt}$. However, this assumption led to an unreasonable pattern for the price index.

- 1- Calculate the demand parameters using either the econometric (Bayesian) method or the calibration (CAL) approach. Note: The CAL model is simple to estimate and is perhaps better suited for agencies and students who are reluctant to incur the relatively high computational cost of the Bayesian Monte Carlo technique.
- 2- Given the demand parameters, recover the p_j 's and a_j 's from the transformation algorithm and determine the welfare equivalent price reduction ρ_t . It follows, from the assumption defined above and equations (8) and (9), that ρ_t can be computed as

$$\rho_t = \frac{1}{\alpha \bar{p}_{t-1}} \left[-\alpha \Delta p_{0t} + (1 - \sigma) \ln \left(\frac{\sum_0^{N_t} \exp \left\{ \frac{a_{kt} - \alpha(p_{kt} - p_{0t})}{1 - \sigma} \right\}}{\sum_0^{N_t} \exp \left\{ \frac{a_{kt-1} - \alpha(p_{kt-1} - p_{0t-1})}{1 - \sigma} \right\}} \right) \right] \quad (10)$$

Notice that for the simple logit model ($\sigma = 0$)

- 3- Compute the index from the simple formula $\frac{I_t}{I_{t-1}} = (1 - \rho_t)$ with $I_0 = 100$.

The consumer surplus variation implies (according to the SLB model)¹³ an equivalent price reduction of only 4.29 % (see Table 2) throughout the sample period. It means that at the end of 1990 consumers demand a reduction in prices of 4.29%¹⁴ to go back to 1977 set of products.

TABLE 2 – WELFARE EQUIVALENT PRICE REDUCTION AND THE PRICE INDEX (SLB MODEL)

Year	ρ (%)	Index
1977	---	100
1978	0.71	99.29
1979	-2.07	101.35
1980	-0.18	101.54
1981	4.21	97.26
1982	6.03	91.39
1983	-2.07	93.29
1984	6.9	86.85
1985	2.25	84.89
1986	0.61	84.37
1987	-2.48	86.46
1988	2.77	84.06
1989	-17.55	98.82
1990	3.14	95.71

¹³ The other models present similar results, which are presented in Tables 4.8 and 4.9.

¹⁴ To obtain this number just subtract the 1977 from the 1990 indices.

Using the NLB estimates to construct the price indices (Table 3) I find results of the same order of magnitude for the welfare equivalent price reduction throughout the sample period (2.37%). In turn, Table 4 shows that the CAL model diverges from both NLB and SLB. In this case, consumer demand a price increase to return to the 1977 settings, meaning that welfare must have decreased on the 1977-1990 period. The lack of robustness may be due to the fact that, unlike the Bayesian methods, α_t can be determined for each time period and therefore welfare variation also accounts for intertemporal differences in this parameter.

TABLE 3 – WELFARE EQUIVALENT PRICE REDUCTION THE PRICE INDEX (NLB MODEL)

Year	$\rho_t(\%)$	Index
1977	---	100.00
1978	0.42	99.57
1979	-1.72	101.29
1980	-0.2	101.50
1981	3.48	97.97
1982	5.79	92.29
1983	-1.9	94.04
1984	5.24	89.11
1985	1.34	87.91
1986	1.26	86.80
1987	-2.15	88.67
1988	1.23	87.57
1989	-14.46	100.24
1990	2.6	97.63

*TABLE 4 – WELFARE EQUIVALENT PRICE REDUCTION THE PRICE INDEX
(CAL MODEL)*

Year	$\rho_t(\%)$	Index
1977	---	100
1978	4.43	95.56
1979	0.06	95.50
1980	-1.11	96.56
1981	-1.49	98.00
1982	-0.6	98.60
1983	-4.75	103.29
1984	1.64	101.59
1985	0.83	100.74
1986	-2.17	102.93
1987	-6.53	109.66
1988	13.78	94.54
1989	-13.15	106.98
1990	-0.63	107.65

Whether the assumption of time invariant quality of the outside good is suitable is largely an empirical matter and depends on the idiosyncrasies of each market. This assumption has less appeal for markets where the imported good is much more likely to exhibit significant temporal variation. For example, in most developing countries, imported vehicles exhibit much more quality improvements than the ones produced domestically.

VI. FINAL REMARKS

The main contribution of this paper is to show how to construct quality adjusted price indices without direct observation of product-level data (prices, quantities and characteristics). The technique used here allows for a welfare based measurement of price change using commonly available (at least for the manufacturing sector) data on revenue and cost. However, the researcher has to be explicit about the structure of demand, supply and has to bear the cost of a high computational burden (although this higher cost can be minimized with the calibration methodology). This paper also stresses the need to impose assumptions on the evolution of relative quality to identify the consumer surplus variation.

This work could be reproduced to build price indices for specific manufactured products in many other countries. Data sets that share the same limitations as the Colombian survey, i.e. that contain information only on revenues and costs not prices or quantities, are available for Chile, Mexico and Turkey, for example. In Brazil, the main information source is provided by the *Pesquisa Industrial Anual*, a survey of Brazilian firms and plants conducted annually by the Brazilian census bureau *IBGE* (see Muendler, 2003, for a detailed description of the data set).

One relevant issue neglected in this paper is the turnover effect. Introducing non-price strategies in the model may yield different estimates for the relevant parameters and consequently for the welfare effect. Incorporating entry and exit decisions in the theoretical model is quite a challenging task though. Lu and Tybout (2000), using the framework developed in Pakes and Maguire (1994), study the turnover effect caused by higher import competition by modeling entry and exit in a dynamic framework. The complexity of their model and the associate computational burden are such that econometric estimation of all the parameters is not feasible. However, this is certainly the path to be followed for future empirical work.

APPENDIX A – (THE TRANSFORMATION ALGORITHM). UNCOVERING RELEVANT QUANTITIES FROM REVENUE AND COST DATA WITH MULTI-PLANT OWNERSHIP

This appendix paraphrases DeSouza (2006b) and is included here for the sake of clarity. It shows how to uncover relevant plant-level quantities from revenue and cost data up to some parameters of the model. Note that Equation (5) in the text can be rewritten as

$$1 + \sum_{r \in F_f} (p_r - mc_r) \frac{\partial s_r}{\partial p_j} / s_j = 0 \quad (\text{A1. 1})$$

Further, after some algebraic manipulations it can be shown that the following equalities hold for the cross and own price derivatives

$$\frac{\partial s_r}{\partial p_j} / s_j = -\frac{\alpha}{1-\sigma} \frac{s_r}{s_j} [-(1-\sigma)s_j - \sigma s_w]$$

$$\frac{\partial s_j}{\partial p_j} / s_j = -\frac{\alpha}{1-\sigma} [1 - (1-\sigma)s_j - \sigma s_w]$$

Note that $s_r/s_j = q_r/q_j$, $R_j = p_j q_j$, $TC_j = mc_j q_j$, and $s_j = q_j/(Q+Q_0)$ where R_j , TC_j , Q and Q_0 are revenue, total variable cost, total output produced by domestic firms and total imported quantity respectively. Hence, substituting these equations into the pricing rule and solving for quantity of plant j belonging to firm f ($j \in F_f$) we obtain

$$q_j = \left(\frac{1-\sigma}{\alpha \cdot (R_j - TC_j)} + \left(\left(\frac{(1-\sigma)}{Q+Q_0} + \frac{\sigma}{Q} \right) \cdot \sum_{r \in F_f} \frac{(R_r - TC_r)}{(R_j - TC_j)} \right) \right)^{-1} \quad (\text{A1. 2})$$

Aggregating over the q_j 's results in

$$Q = \sum_{f=1,2,\dots,N} \sum_{j \in F_f} \left(\frac{1-\sigma}{\alpha \cdot (R_j - TC_j)} + \left(\left(\frac{(1-\sigma)}{Q+Q_0} + \frac{\sigma}{Q} \right) \cdot \sum_{r \in F_f} \frac{(R_r - TC_r)}{(R_j - TC_j)} \right) \right)^{-1}$$

where N is the total number of firms. This non-linear equation can be solved numerically for Q given $(\alpha, \sigma, R_j, TC_j, Q_0)$. Then, given the same parameters and data, q_j is

determined from (A1. 2), whereas p_j , mc_j , s_j and s_w follow trivially from $p_j = R_j/q_j$, $mc_j = TC_j/q_j$, $s_j = q_j/(Q+Q_0)$ and $s_w = q_j/Q$ respectively.

Finally, the log-linearized version of the demand system (3) can be solved for relative quality a_j

$$a_{jt} \equiv \xi_j - \xi_0 = \alpha \cdot (p_j - p_0) + \sigma \ln s_{wj} + \ln s_j - \ln s_0$$

APPENDIX B – EMPIRICAL METHODS

Econometric Method

Dynamics is introduced into the model through the assumption that relative quality and marginal cost follow an exogenous¹⁵ VAR process given by

$$a_{jt} = b_{01} + \phi a_{jt-1} + \phi^c mc_{jt-1} + \beta t + \varepsilon_{jt}^a \quad (11)$$

$$mc_{jt} = b_{02} + \lambda mc_{jt-1} + \lambda^a a_{jt-1} + \varphi t + \varepsilon_{jt}^c \quad (12)$$

From the demand system, the price setting game and the VAR we are able to uncover demand and supply side “errors”, represented respectively by ε_{jt}^c and ε_{jt}^a . At this point the model seems very close to Berry’s methodology where similar error terms are combined with exogenous product characteristics to form the identifying moment conditions. Here, however, these data are not available. Note that not even prices or quantities are observed; they are themselves functions of data and demand parameters to be estimated from the model. The model is therefore not identified such that traditional econometric technique like GMM and ML do not apply.

To identify it more structure has to be imposed on the parameters. This is achieved by assuming a prior knowledge of the parameters distribution and using the data set and the structure imposed by the model to update this prior according to the Bayes rule

¹⁵ It is also assumed that firms observe their marginal costs and relative quality before they set prices. The exogeneity of the joint evolution of marginal costs and quality is an important assumption since it keeps the model consistent with the assumption that firms maximize static profits (5). Otherwise, if firms could influence marginal costs and quality of their products, we would have to set up a dynamic model of profit maximization following Pakes and McGuire (1994) framework.

$$p(\theta; D) \propto L(D; \theta) \cdot p(\theta)$$

The LHS is the posterior probability updated by the data D , and the RHS is the product of the likelihood function of the data times the prior distribution of the parameters. However, only in special cases, the posterior distribution has a closed form from which we can make inference either by sampling from it or by consulting easily available tables. Fortunately, Monte Carlo techniques have been developed to deal with such problem. Shortly, it relies on ergodic theory to guarantee that a computable statistic converges to the true posterior distribution. Then, once convergence has been attained, we can sample from this statistic and make inference.¹⁶

Calibration Strategy

Aggregate quantities also carry information on demand parameters and therefore may help to pin down the demand parameters. Assume now that the amount of beer, in hectoliters, consumed in a given year is observed.

Note that the model implies an aggregate quantity given some parameters and data. We can then ask the model to match this observed quantity according to

$$Q_t(\alpha_t, R_{jt}, TC_{jt}, Q_0) = Q_t^{obs} \quad (13)$$

All data in this equation are observed which implies that α_t can be determined for each time period t . Nevertheless, the equation above embodies a few assumptions on the nature of the aggregate measure and on the consumer preferences. Note that it requires that the quantity implied by the model matches exactly its empirical counterpart and that the discrete choice model is the simple logit form instead of its nested version ($\sigma=0$). These assumptions do not come without a cost. First, the simple logit model places very restrictive assumptions on consumer preferences such that cross-price elasticities have somewhat undesirable properties.¹⁷ Second, it requires a very reliable source for the total quantity since measurement error is not allowed.

¹⁶ See KLT and DeSouza (2006b) for additional details about the Bayesian methodology.

¹⁷ Berry, Levinsohn and Pakes (1995) discuss this subject in more depth. Note also that the model could be appended to accommodate more sophisticated setups, e.g. the Nested Logit, and consequently, more plausible cross-price effects. However, observing more variables would be necessary to devise other calibrating equations and uncover the extra parameters introduced by these new setups. Thus, due to the limitation commonly found in plant-level data sets the logit model restrictive assumptions cannot be relaxed in this calibration framework.

Given these drawbacks, a legitimate question may arise. Why bother making these restrictions on the model? The reason is that the Bayesian technique proved to be a little cumbersome to implement computationally and many government agencies and students are somewhat reluctant to adopt the Bayesian approach, especially one that requires Monte Carlo simulation. This technique also requires the imposition of priors on the parameters distribution, which even some leading researches are reluctant to do. Also, the calibration requires only data on a cross-section of firms in a given year while the identification of the Bayesian model demands a panel with at least three years of observation since it requires a VAR estimation.

REFERENCES

- Berry, S. Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics*, 25 (2), p. 242-262, 1994.
- Berry, S.; Levinsohn, J.; Pakes, A. Automobile Prices in Market Equilibrium. *Econometrica*, 63(4), p. 841-890, 1995.
- DeSouza, S. Studying Differentiated Product Industries using Plant-Level Data. *Economics Bulletin*, v. 12, n. 3, p.1-11, 2006a.
- _____. Combining Aggregate and Plant-Level Data to estimate a Discrete-Choice Demand Model. *Brazilian Review of Econometrics*, v. 26, n.2, p. 213-234, 2006b.
- Katayama, H.; Lu, S.; Tybout, J. Why Plant-Level Productivity Studies Are Often misleading, and an Alternative Approach to Inference. *NBER Working Paper* N°9617, 2003.
- Lu, S.; Tybout, J. *Import-Competition and Industrial Evolution: A Computational Experiment*. (mimeo). Penn State University, 2000.
- McFadden, D. Econometric Models of Probabilistic Choice. In: Manski, C; McFadden, D. (Eds). *Structural Analysis of Discrete Data*, 1981.
- Muendler, M. *The Database Pesquisa Industrial Annual 1986-2001: a Detective's Report*. (mimeo). UC, San Diego, 2003.
- Nevo, A. New Products, Quality Changes and Welfare Measures Computed from Estimated Demand Systems. *The Review of Economics and Statistics*, 85(2), p.266-275, 2003.
- Pakes, A.; McGuire, P. Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model. *RAND Journal of Economics*, 25(4), p. 555-589, 1994.
- Roberts, M. Colombia, 1977-85: Producer Turnover, Margins, and Trade Exposure. In: Roberts, M.; Tybout, J. (Eds.). *Industrial Evolution in Developing Countries*. Oxford University Press, Oxford and New York, 1996.

- Timmins, C. Estimating spatial differences in the Brazilian cost of living with household location choices. *Journal of Development Economics*, 80, p. 59-83, 2006.
- Trajtenberg, M. *Economic Analysis of Product Innovation: The Case of CT Scanners*. Harvard University Press, Cambridge, MA, 1990.