Session Number: Session 7A
Time: Friday, August 29, AM

**Paper prepared for the 30th General Conference of
The International Association for Research in Income and Wealth**

Portoroz, Slovenia, August 24-30, 2008

Non-Market Hedonics for Measurement and Modelling: applying lessons from price statistics to measuring and monitoring government activity

Matthew Powell
Oxford Policy Management Limited and
Office for National Statistics

For additional information please contact:
Matthew Powell (Matthew.Powell@opml.co.uk)
(Matthew.powell@ons.gov.uk)

Oxford Policy Management Limited
6 St Aldates Courtyard
38 St Aldates
Oxford
OX1 1BN
United Kingdom

Non-Market Hedonics for Measurement and Modelling: applying lessons from price statistics to measuring and monitoring government activity [1]

The issue of measuring change in the quantity and quality of non-market output is central to efforts to measure, monitor, and evaluate Government and other Non Market Activities.

This paper attempts to apply the lessons learned in developing hedonic methods for price statistics to the problems of measuring non market output. It starts with an examination of the economic theory of index numbers and its extension to allow for the measurement of quality change using a hedonic approach, before developing the theory further to cover non market production. This is followed by a discussion of how theoretical indices are calculated in practise and a suggestion about how the new theory of non market output could be applied. Finally a demonstration using data for secondary education from [*an Asian developing country*] *is followed by a discussion of the possibilities for practical application of the methods suggested and an outline of open research questions*

Keywords: Non market services, National Accounts, Monitoring and Evaluation, Productivity, Technical change, Price Statistics, Hedonic Quality Adjustment.

Calculating the change in the volume of non-market services is a problem of interest to analysts in several fields. Public services make up a large fraction of overall GDP and the way their output is measured has a substantial impact on national accountant's assessments of economic growth, welfare, and productivity[2]. Of at least equal importance is the need of public sector managers at all levels to ensure that inputs are being used in an appropriate way so that services are being properly managed in the interests of users, providers and the taxpayers who fund them[3]. Without established measures it is difficult to either identify the ideal position for service providers to reach or monitor their progress towards it. Finally the international development community has a keen interest in ensuring that their interventions in areas such as health and education are having the desired effects on the delivery of health and educational services[4]. Monitoring and Evaluation in this context involves a lot more than defining and measuring output but defining and measuring output is an essential component of it.

Despite the practical importance of the topic economists and economic statisticians have found it difficult to accommodate the non-market sector within their traditional measurement methods because these rely so heavily on market prices. This is just one example of the role that prices play in summarising enormous amounts of information about individual preferences and production possibilities. Without prices it is necessary to examine these functions directly. Slightly paradoxically out of all official statisticians it is those charged with measuring market prices who have been most heavily involved in working with preference functions. Since the spread of personal computers in particular they have been increasingly involved in the use of 'hedonic' behavioural equations to estimate quality change. Analysts have not only developed substantial practical experience but made extensive investigations into the theoretical underpinnings of the techniques they use. Even more recently public sector analysts

[1] The author would like to thank… The views and expressed in this paper and the estimates presented are entirely the responsibility of the author and do not represent the opinions of OPML or the Office for National Statistics.

[2] See ONS 2005 and 2007

[3] RSS report

[4] Mackay 2006 for Dev. sate of art

have started to link information on the costs and activities of public sector producers and the characteristics and experiences of their users within the same datasets.

The purpose of this paper is to show how the theory and techniques developed for hedonic quality adjustment of price indices can be applied to the combined datasets of public sector producers and users to improve the measurement, and also the management, of public sector activities. The paper's general strategy is to estimate functions for the costs of producing non market services and use to which households can put them and it can be seen as falling within the "household production" approach. Much of the material will be unfamiliar to at least part of the intended audience so there are quite a lot of preliminaries to go through.

The paper begins with short revision of theoretical economic of index numbers, drawn mainly from Diewert (1981), and the extension by Triplett (1983) to deal with theoretical price indices of characteristics as well as goods. It then shows how the same ideas can be used to define theoretical quantity indices for the non-market sector. The next section deals with the practical estimation of the theoretical indices defined in the first and follows the same pattern; first the general assumptions and techniques that apply to all practical quantity measurement, then the techniques used for hedonic quality adjustment in the market sector, finally the application of the lessons of the first two subsections to the non market sector. The final substantive section of the paper is an attempt to apply the techniques developed in the first two sections to an actual dataset for secondary schools and their pupils collected in 2004 in [an Asian developing country] in order to understand how they might be used in practice. It is important to recognise that this is an attempt to explore and develop a technique as opposed to an actual policy analysis. While the results are interesting no direct policy conclusions can be drawn.

A note on notation. A paper such as this is necessarily quite notation heavy. In the rest of the paper I will follow standard practise by denoting vector variables or functions in bold and scalers in standard facing. Thus $Z(\mathbf{q}) = Z(q_1,q_2,\ldots q_i\ldots\ldots q_n)$. The partial derivative of $Z(\mathbf{q})$ by q_i evaluated for $\mathbf{q} = \mathbf{q}^{\mathbf{0*}}$ (i.e $\partial Z/\partial q_i \;|_{q^{0*}}$) is designated by $Z_{qi}|_{\mathbf{q0*}}$

1 Defining theoretical indices of hedonic quality adjustment for "other individual non market services".

1.1 Theoretical indices of quantity changes with no quality change

The first economic[5] theories of quantity change were derived from the more venerable and voluminous economic theory of price indices. These theories were based on the concept of the cost index. Cost indices are normally defined in terms of utility functions. However the theory for production functions is identical and more suited to the household production approach of this paper. We therefore define the cost index **C()**as the minimum cost to a representative producer of achieving a reference output level[6] $\mathbf{Z(q^r)}$ derived from a reference input quantity bundle $\mathbf{q^r}$**,** when facing a vector of prices **p**. Formally $\mathbf{C(p^t,Z(q^r))}$ **=** minimise $\mathbf{p^t\ q}$ with respect to $\mathbf{q:Z(q)>=Z(q^r)}$**.**

The true price index for period **t** relative to period **0** is defined as

$$(1) \qquad \mathbf{P^{0t}(p^t,p^0|Z(q^r))=C(p^t,Z(q^r))/C(p^0,Z(q^r))}$$

Note that the estimation of hypothetical situations, the minimum cost of attaining the output provided by a reference bundle at a time when the reference bundle was not purchased, is inherent in the economic approach to index numbers.

The first index of quantity change derived from economic theory was the Konus index or the change in value index deflated by the theoretical price index in (1): (Konus 1924),

$$(2) \quad \textbf{Konus} \qquad \mathbf{Q_K^{0t}(q^t,q^0,p^t,p^0,q^r)=(p^tq^t/p^0q^0)*C(p^0,Z(q^r))/C(p^t,Z(q^r))}$$

The alternative Allen index (Allen 1949) replaces the reference output level with a reference set of prices $\mathbf{p^r}$ and calculates the minimum cost of achieving the output available from $\mathbf{q^t}$ at the reference prices divided by the minimum cost of achieving the output available from $\mathbf{q^0}$ when facing the same prices.i.e.

$$(3) \quad \textbf{Allen} \qquad \mathbf{Q_A^{0t}(q^t, q^0 , p^r)=C(p^r,Z(q^t))/C(p^r,Z(q^0))}$$

The theoretical Konus and Allen indices[7] can be related by taking common reference points -the most obvious being the prices and quantities actually observed in period t (a

---

[5] Economic approaches to index calculation were originally, and in many ways still are, secondary to the dominant "axiomatic" tradition which lays down a list of desirable properties and seeks index formulae that meet them.

[6] We make the simplifying assumption that that multiple outputs are aggregated into a single scalar measure

[7] Other important theoretical indices include those defined in Malmquist 1953. They have the potentially attractive feature of being defined without the use of price indices. However, despite extensive academic use of the index for productivity analysis, there is no tradition of using Malmquist indices in National Accounts so measures based on it will not generally be directly comparable with market sector measures except in the homothetic case when they equal the Konus and Allen.

Paache reference) and period 0 (a Laspeyres reference)[8]- and using the natural (for an economist) assumption that the quantities actually chosen in any period are the cost minimising ones. i.e. $\mathbf{C(p^t,Z(q^t)) = p^t\ q^t}$ for all t. If this assumption holds the Paache perspective Konus quantity index equals the Laspeyres perspective Allen quantity index and vice versa. i.e.

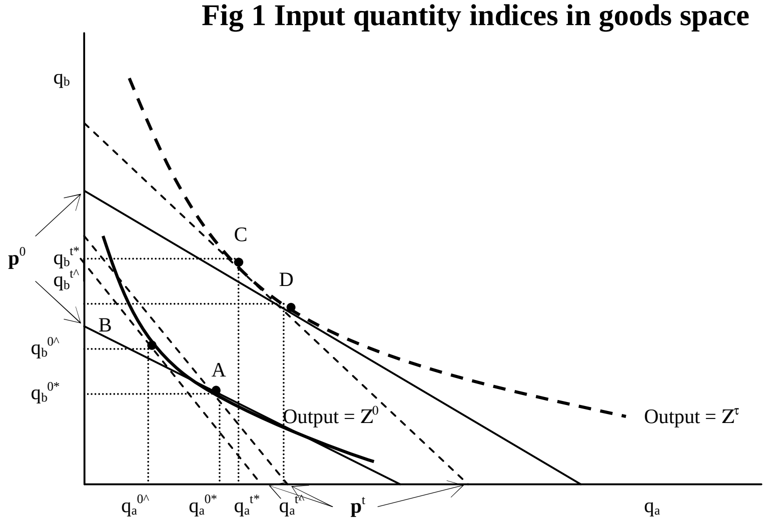
$$\text{(4)}\qquad \mathbf{Q_{K,Pche}^{0t}(q^t,q^0,p^t,p^0,q^t)=}$$

$$\mathbf{(p^tq^t/p^0q^0)*C(p^0,Z(q^t))/C(p^t,Z(q^t))=(p^tq^t/p^0q^0)*C(p^0,Z(q^t))/p^tq^t=}$$

$$\mathbf{C(p^0,Z(q^t))/p^0q^0=C(p^0,Z(q^t))/C(p^0,Z(q^0))= Q_{A,Lasp}^{0t}(q^t,q^0,p^0)}$$

It is obvious that where preferences are homothetic so that the Laspeyres and Paache points of view give identical theoretical indices the Konus and Allen indices are identical. Diewert (Diewert 1981) shows in that case they are also equal to the actual ratio of outputs in the two periods. For the rest of this paper we shall use the term "Quantity Index" to referring to both the Paache perspective Konus index $Q_{K,Pche}^{0t}\mathbf{(q^t,q^0,p^t,p^0,q^t)}$, and the Laspeyres perspective Allen index, $Q_{A,Lasp}^{0t}\mathbf{(q^t,q^0,p^0)}$, and indicate simply as $Q^{0t}(q^t,q^0,p^0)$ except where the two are unequal. Figure 1 shows theoretical indices of input quantities with two inputs, q_a and q_b.[9]

Fig 1 Input quantity indices in goods space



The curve Z^0 shows the combination of inputs that will produce an output $Z(\mathbf{q}^0)$ while Z^t shows the combination of inputs that will produce an output $Z(\mathbf{q}^t)$.P^0 shows the price of q_b in terms of q_a at time 0 and P^t the price at time t.

Point A or (q_a^{0*},q_b^{0*}) is the minimum cost combination of inputs for producing $Z(\mathbf{q}^0)$ at prices P^0 and point C or (q_a^{t*},q_b^{t*}) is the minimum cost combination of inputs for producing $Z(\mathbf{q}^t)$ at prices P^t. The cost of A equals $p_a^0q_a^{0*}+p_b^0q_b^{0*}$ and the cost of C equals $p_a^tq_a^{t*}+p_b^tq_b^{t*}$. Note that Z^0 is the maximum output that can be produced for cost of $p_a^0q_a^{0*}+p_b^0q_b^{0*}$ at prices P^0 and Z^t is the maximum output that can be produced for cost of $p_a^tq_a^{t*}+p_b^tq_b^{t*}$ at prices P^t.

---

[8] Many analysts recommend using "superlative " indices that combine different perspectives. I take no view on the issue in this paper. For our present purposes it would complicate the analysis without changing the argument.

[9] Extension of the argument to many inputs is a standard feature of production theory. See for example Varian 1992.

Point B or ($q_a^{0\wedge}$,$q_b^{0\wedge}$) is the minimum cost combination of inputs for producing Z(**q**^{0}) at prices P^t and costs $p_a^t q_a^{0\wedge}+p_b^t q_b^{0\wedge}$, and point D or ($q_a^{t\wedge}$,$q_b^{t\wedge}$) is the minimum cost combination of inputs for producing Z(**q**^{t}) at prices P^0 and costs $p_a^0 q_a^{t\wedge}+p_b^0 q_b^{t\wedge}$.

The theoretical quantity index is given by the change in total value between point A and point C divided by the price change equal to the change in total value between point D and point C. This is the same as the cost of point D (i.e.**C(p^0,Z(q^t))** in the notation of **(2)**)divided by the cost of point A (i.e.**p^{0}q^0**). The value of this theoretical index in our example is $p_a^0 q_a^{t\wedge}+p_b^0 q_b^{t\wedge}$ divided by $p_a^0 q_a^{0*}+p_b^0 q_b^{0*}$.

Note that both approaches to defining the theoretical index of quantity change are crucially dependent on movement along the current period constant output curve Z^t and that at the cost minimisation points A and C the technical rate of substitution between q_a and q_b is equal to the ratio of p_a to p_b. Formally;

(5) **C(**p_a,p_b,Z^r) is solved by choosing q_a and q_b to minimise $p_a q_a+p_b q_b$ subject to $Z(q_a,q_b)=Z^r$

the Lagrangian is $\mathcal{L}_{cost} = p_a q_a+p_b q_b+\lambda_{cost}(Z(q_a,q_b)-Z^r)$

the first order conditions are

$p_a+\lambda_{cost} Z_{qa}|_{\mathbf{q*}}=0$, $p_b+\lambda_{cost} Z_{qb}|_{\mathbf{q*}}=0$, and $Z(q_a,q_b)=Z^r$

so $p_b/p_a=Z_{qb}|_{\mathbf{q*}}/Z_{qa}|_{\mathbf{q*}}$

N.B. choosing q_a,q_b to maximise $Z(q_a,q_b)$ subject to $p_a q_a+p_b q_b$ equalling a fixed cost C^r gives us the same results from the first order conditions

$Z_{qa}|_{\mathbf{q*}}+\lambda_{output} p_a=0$, $Z_{qb}|_{\mathbf{q*}}+\lambda_{output} p_a=0$, and $p_a q_a+p_b q_b=C^r$

If we set p_a^0 as 1 we can calculate the theoretical index for input quantities using the ratio of marginal physical products evaluated at $q^{0*}{}_a$,$q^{0*}{}_b$

(6)
$$\mathbf{Q^{0t}(q^t,q^0,p^0)=C(p^0,Z(q^t))/\ C(p^0,Z(q^0))=}$$

$$[q_a^{t\wedge}+q_b^{t\wedge}(Z_{qb}|_{\mathbf{q0*}}/Z_{qa}|_{\mathbf{q0*}})]/[q_a^{0*}+q_b^{0*}(Z_{qb}|_{\mathbf{q0*}}/Z_{qa}|_{\mathbf{q0*}})]$$

where $q_a^{t\wedge}$ and $q_b^{t\wedge}$ are defined by solving the pair of equations

(7)
$$Z_{qb}|_{\mathbf{q0\wedge}}/Z_{qa}|_{\mathbf{q0\wedge}}=Z_{qb}|_{\mathbf{q0*}}/Z_{qa}|_{\mathbf{q0*}} \text{ and } Z(q^{t\wedge}{}_a,q^{t\wedge}{}_b)=Z^t$$

1.2 Theoretical indices of input quantity with hedonic quality adjustment.

The methodology discussed so far has assumed that the products that users are deriving welfare from in the current period are identical to those they were consuming in the base period. Relaxing this assumption produces a variation on the cost function in (1) with the cost function **C()** representing the minimum cost of achieving a reference output level from a reference quantity bundle **q,** conditional on facing a vector of prices **p** for a set of available inputs **I^t**. available at time t so that the theoretical Laspeyres perspective input price index becomes

(1a)
$$\mathbf{P^{0t}(p^t,p^0|I^t,I^0)=C(p^t,Z(q^0)|I^t)/C(p^0,Z(q^0)|I^0)}$$

The corresponding specification for an input quantity index is

(3a) $$\mathbf{Q^{0t}(q^t,q^0,p^t,p^0,q^t)|I^0,I^0)=(p^tq^t/p^0q^0)*C(p^0,Z(q^t))|I^t)/C(p^t,Z(q^t)\ )|I^t)}$$

We can still assume that that the cost minimising expenditure is equal to the actual expenditure $\mathbf{C(p^t,Z(q^t)|I^t)=p^tq^t}$ for all t. However the numerator in the second term on the right of (3a),$\mathbf{C(p^0,Z(q^t)|I^t)}$ now represents the minimum cost of achieving the output observed at time t given the available input set I^t valued at the input prices prevailing at time 0. The obvious difficulty is that some of the inputs in the vector $\mathbf{q^t}$ may not have been available at time 0.

The *hedonic* method is an attempt to avoid this problem by treating new products as different *qualities* of existing products and defining the quality of each product as a short hand reference to the *quantities* of certain characteristics it embodies[10] that are available in both periods. Note that this implies that the only relevant characteristics in an input cost index are those that affect the cost of production. Consider for instance the replacement of one model of lathe by another where the lathes have only two characteristics affecting output, speed and cutting angle say, and for which the relations between the characteristics are the same as those between the products example in fig 1. The function Z() is now defined over quantities of these characteristics (say $x_1,..,x_n$) as opposed to quantities of products $q_1,..,q_n$**.** The quantity index now becomes

(3b) $$\mathbf{Q^{0t}(x^t,x^0,p^t,p^0,x^t))=(p^tx^t/p^0x^0)*(1/(C(p^t,Z(x^t))|I^t)/C(p^0,Z(x^t))|I^t))=}$$

$$\mathbf{C(p^0,Z(x^t))|I^t)/p^0x^0 \text{ or } C(p^0,Z(x^t))|I^t)/C(p^0,Z(x^0))|I^0)}$$

Where $\mathbf{C(p^t,Z(x^t))|I^t)/C(p^0,Z(x^t))|I^t)}$ is the change in the minimum cost of achieving an output level $\mathbf{Z(x^t)}$ due to the availability of the new lathes. Because all the characteristics available in the second period were available in the first $\mathbf{C(p^0,Z(x^t))|I^t)}$ **is** now a well defined concept. Changes in the relative costs of acquiring characteristics will cause firms to substitute among them in just the same way that changes in the relative costs of acquiring goods does. Note that the existence of a lot of other homogenous factors of production besides lathes that do not embody separate quality characteristics does not affect out argument as they can each be considered as inputs with a single characteristic. It might however change the *value* of $\mathbf{C(p^t,Z(x^t))|I^t)}$ if the introduction of a new lathe causes firms to substitute between lathes and other factors of production.

[10] We are implicitly assuming that new products are simply re-packagings of old characteristics. Truly new characteristics give us the intractable new products problem. The standard proposal for dealing with this is to calculate the demand reservation price for the periods in which the new product didn't exist. The comparison between the reservation price and the actual price then gives a lower bound on the welfare increase an individual purchaser receives from the invention of the product. Although the absence of an expenditure weight in the base period makes it impossible to calculate a traditional Laspeyres index allowing for the appearance of the new good an assumption that utility functions and their aggregation is unchanged from period to period would allow the theoretical calculation of the effect. Balk (Balk 1999) provides a particularly tractable example with two stage CES welfare functions. This approach is never followed in practise. Instead statisticians strive to cover new products as soon as there is significant expenditure and capture the impact of the diffusion throughout the market. Some analysts have claimed that truly new characteristics are actually quite rare. Burnstein for instance (1961) points out that the introduction of television in the US was really a new distribution channel for baseball and vaudeville.

As in the previous case of a change to the price of input goods, the replacement of the lathes used at time 0 by the ones used at time t implies as shift in the level of production from Z^0 to Z^t .If the level of production using the new machine does not change from that using the old then they are the same quality *even if the amount paid for them differs*.

Because producers can only purchase characteristics packaged together in goods the cost function over input characteristics is far more complicated than the cost function over input products but this does not imply that an input characteristics *production* function is particularly complicated. It would be perfectly possible to calculating the theoretical index for the quality difference of the new lathes using characteristics versions of (6) and (7). In practise of course actual output differences between any two periods are likely to be caused by a whole range of factors. Isolating the input quality change requires us to take an observed reference output and estimate the comparison output that would have been produced with all other inputs held constant but using the same quantity of "lathes" of a different type.

1.3 Theoretical indices of input quantities for "other individual non market services" with hedonic quality adjustment

We are now ready to consider a hedonic approach to quality adjustment for non market services. Begin by assuming that these services are provided by ι service delivery units or SDUs that each produce a quantity of services Π_ι that have a number of characteristics $x_1,..,x_n$. The cost of production depends on these characteristics. There are 1...η ..H households that use these characteristics together with household specific characteristics, Ξ^η to produce a homogenous scalar *outcome*[11] Θ_η. The relations are expressed in a characteristics equation (8) and an outcome equation (9).

(8) $$C\ (\Pi_\iota)= C\ (x_{1\iota},.,x_{n\iota})$$

(9) $$\Theta\eta = \Theta\eta\ (x_{1\iota},.,x_{n\iota},\Xi^\eta{}_{1\iota},.\Xi^\eta{}_{m\iota})$$

We might say for example that Π_ι represents hours of teaching received by pupils at a school to which households send their children in order to help them to attain a desired educational outcome $\Theta*$, and that $x_{1\iota},..,x_{n\iota}$ represent the school's quality characteristics which could be anything from the qualifications of the teachers to the use or non use of a particular teaching method. Alternatively the SDU might be a clinic Π_ι represent the number of patients treated and $x_{1\iota},..,x_{n\iota}$ the proportions of different treatments used. Suppose those characteristics change. Can we use $\mathbf{I^0}$ to represents the school or other SDU in the base period and use the relationships in (6) and (7) to calculate a quantity index for the quality difference?

(10) $$\mathbf{Q^{0t}(x^t,x^0,p^0|I^0,I^t)=C(p^0,}\ \Theta\eta,(x_{1\iota},,x_{n\iota},\Xi^\eta{}_{1\iota},.\Xi^\eta{}_{m\iota})\mathbf{|I^0)/C(p^0,}x_{0\iota},,x_{0\iota},\Xi^\eta{}_{1\iota},.\Xi^\eta{}_{m\iota})\mathbf{)|I^0)}$$

[11] The important concept of "Outcomes" was introduced in a discussion of various volume indicators that might relate to non market output ESA95 (Eurostat 1997). The idea is similar to the concept of impact in the development literature and relates to the actual change in the state of the people whom a particular service or policy is supposed to benefit. Whereas the aggregation of market outputs into a single scalar measure is relatively unproblematic the aggregation of non market *outcomes* is a very difficult issue. It is however a different issue from the one of defining and measuring the *outputs* of service delivery units.

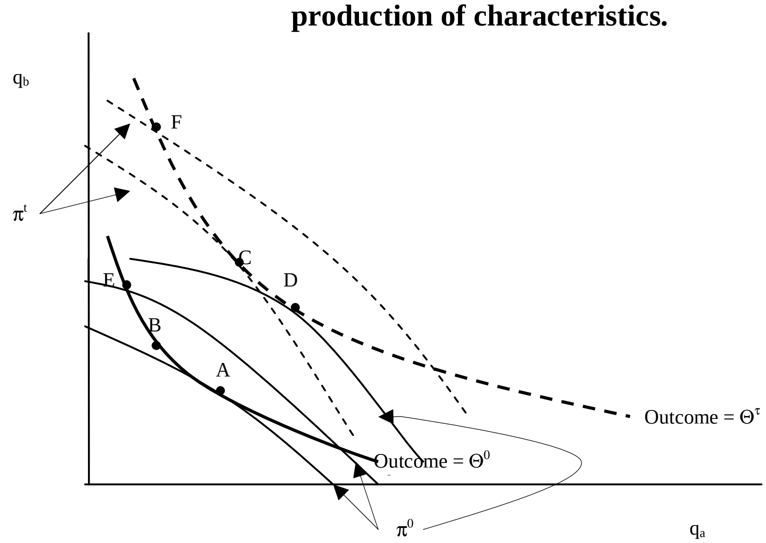
Two main difficulties occur;

- Some household characteristics are fixed so households can't minimise costs by substituting between them and school characteristics
- The observed quantities can no longer be taken as indicating the point of cost minimisation by the household producing the educational outcomes so the relative marginal products at that point can no longer be used to value the characteristics consumed in period t.

The first difficulty is true but irrelevant. Practical calculation of quality adjusted price indices for cars or computers requires strong separability of characteristics on the production or utility function of the firm or consumer to the extent that decisions on the relative quantities of characteristics in the input good being quality adjusted are independent of the relative quantities of other characteristics.

The lack of cost minimisation is more difficult to deal with. However, even if households are not cost minimising, iso-outcome curves still exist. We can imagine households choosing schools to maximise the outcome that they can obtain in each period given the characteristics of the schools available to them (which would of course include the distance from home). The slope of the Z^r curve is still meaningful to them. Similarly even if SDUs are not producing characteristics at the minimum cost required for households to allow the households to achieve the outcomes actually observed, such a minimum cost position still exists. For any observed level of educational outcomes we can construct a unique household iso-outcome curve running through it and a unique school iso-cost curve with a tangent to that iso-outcome curve. Figure 2 illustrates.

Fig 2 Input quantity indices in characteristics space without cost minimisation in production of characteristics.



In formal terms cost functions can still be defined with respect to a reference outcome level $\Theta\eta^r$ but the price line (or hyperplane in the many characteristics case) **p** is replaced by a convex production technology $\boldsymbol{\pi}$ (where changes in the government's own input costs as well as engineering changes are treated as changes in technology). Note that the price lines **p** in Fig.1 can be seen as "technologies" for exchanging one input for another without changing the overall cost. The two sets of parallel price lines $\mathbf{p^0}$ and $\mathbf{p^t}$ represent two different technologies. The different lines within each set represent the same technology even though the total cost of inputs is different for each line. Naturally these fixed price "technologies" show constant returns to scale but this is not required for a well defined cost function provided the reference outcome is fixed.

In the example in of Fig.2 we may have schools producing characteristics at E in period 0 and F in period t. Given those packages of characteristics from schools households are able to produce outcome levels Θ^0 and Θ^t respectively but they could produce better outcomes at no cost to the government. Alternatively the government could cut its costs and move to a more efficient mix of output characteristics without reducing outcomes.

Formally we can define $\mathbf{C}(\boldsymbol{\pi}, \Theta\eta(\mathbf{x}_r,\Xi^\eta_r), \Xi^\eta_r|\mathbf{I})$ as the minimum cost to the government of producing quantities of characteristics that will allow a representative household to achieve a reference outcome level $\Theta\eta(\mathbf{x}_r,\Xi^\eta_r)$ derived from a reference characteristics bundle $\mathbf{x^r}$**,** and a reference set of household characteristics ,Ξ^η_r, *given those reference household characteristics*, a set of SDUs $\mathbf{I}$, and a technology for producing characteristics $\boldsymbol{\pi}$**.** Notice that it is necessary to specify reference levels of household characteristics, Ξ^η_r,, in effect a reference set of households, as well as reference outcomes and reference production technologies in order to get a well defined cost function. We could of course choose any reference households but it seems sensible to choose ones consistent with the reference outcome.We can now define our Laspeyres perspective Allen and Paache perspective Konus indices as;

(11) $$\mathbf{Q_{A,Lasp}^{0t}(x^t,x^0,\pi^0|I^0,I^t)=C(\pi^0}, \Theta\eta(\mathbf{x}_t,\Xi^\eta_t), \Xi^\eta_0|\mathbf{I^0})/\mathbf{C}(\boldsymbol{\pi}^0, \Theta\eta(\mathbf{x}_0,\Xi^\eta_0), \Xi^\eta_0|\mathbf{I^0})$$

and

(12) $$\mathbf{Q_{K,Psche}^{0t}(x^t,x^0,\pi^t,\pi^0,x^t)|I^0,I^t)=}$$

$$\mathbf{(p^tx^t/p^0x^0)*(1/(C(\pi^t}, \Theta\eta(\mathbf{x}_t,\Xi^\eta_t), \Xi^\eta_t|\mathbf{I^t}))/\mathbf{C}(\boldsymbol{\pi}^0, \Theta\eta(\mathbf{x}_t,\Xi^\eta_t), \Xi^\eta_t|\mathbf{I^0}))$$

We can express the two indices in words as

(11a) $\mathbf{Q_{A,Lasp}^{0t}}$=		the minimum cost to the government of producing characteristics that would enable the *households of the current* period to obtain the *outcome observed in the current* period using the characteristics *technology and service delivery units of the base* period **divided by** the minimum cost to the government of producing characteristics that would enable the *households of the base* period to obtain the *outcome observed in the base period* using the characteristics *technology and service delivery units of the base* period
(12a) $\mathbf{Q_{K,Psch}^{0t}}$ =The change in actual spending	*	the minimum cost to the government of producing characteristics that would enable the *households of the current* period to obtain the *outcome observed in the current period* using the characteristics *technology and service delivery units of the base* period **divided by** the minimum cost to the government of producing characteristics that would enable the *households of the current* period to obtain the *outcome observed in the current* period using the characteristics *technology and service delivery units of the current* period

Note that actual current costs at E and F are no longer equal to the minimum costs for those outputs so the theoretical Laspeyres perspective Allen and Paache Perspective Konus quantity indices now differ. Even if governments were cost minimising the indices would still be different because the reference households differ. Finally the logic of (4) also collapses and with it the result on homothetic outcome functions.

2 Applied economic quantity indices

2.1 Indices of quantity change in National Accounts

The greatest practical problem with actual calculation of the indices described in section 1 is that they require prices and quantities for every single transaction in the economy whereas the data national accountants have available are either aggregates or samples. If they wish to calculate the theoretical indices they have to split the transactions in the flows they are trying to measure into groups for which working with these sources will not introduce too many distortions. Taking $q_{0i}^{t\wedge}$ as the optimum set of values of q_i at t given the prices from period 0 we can split the quantity index as follows;

(13)
$$Q^{0t}(\mathbf{q^t},\mathbf{q^0},\mathbf{p^0})=C(\mathbf{p^0},\mathbf{U(q^t)})/C(\mathbf{p^0},\mathbf{U(q^0)})=\sum_i p_i^0 q_{0i}^{t\wedge} \Big/ \sum_i p_i^0 q_{0i}^{0\wedge} =$$
$$\sum_i p_i^0 q_{0i}^{t\wedge} * q_{0i}^{0\wedge} / q_{0i}^{0\wedge} \Big/ \sum_i p_i^0 q_{0i}^{0\wedge} = \sum_i p_i^0 q_{0i}^{0\wedge} * q_{0i}^{t\wedge} / q_{0i}^{0\wedge} \Big/ \sum_i p_i^0 q_{0i}^{0\wedge} =$$
$$\sum_i s_{0i}^{0\wedge} * q_{0i}^{t\wedge} / q_{0i}^{0\wedge} = \sum_j s_{0j}^{0\wedge} (\sum_{i\in j} u_{0i}^{0\wedge} * q_{0i}^{t\wedge} / q_{0i}^{0\wedge}) \approx \sum_j \overset{\approx}{s}{}_{0j}^{0} (\sum_{i\in j} \overset{\approx}{u}{}_{0i}^{0} * q_i^{t\wedge} / q_i^{0\wedge})$$

where

$s_{0i}^{0\wedge}$,$s_{0j}^{0\wedge}$	are the base period value shares of all individual transactions assuming optimisation[12], and the shares of each group,
$u_{0j}^{0\wedge}$	are within group shares,
$\overset{\approx}{s}{}_{0j}^{0}$ $\overset{\approx}{u}{}_{0i}^{0}$	are estimates of. $s_{0j}^{0\wedge}$ and $u_{0j}^{0\wedge}$.

The universal practise of National Accountants is to estimate $\sum_{i\in j} u_{0i}^{0\wedge} * q_{0i}^{t\wedge} / q_{0i}^{0\wedge}$ using the Laspeyres quantity index $\sum_{i\in j} u_{0i}^{0} * q_{0i}^{t} / q_{0i}^{0}$, based on the observed quantities and shares. It is often stated that this is an upper bound on the theoretical quantity index because $\mathbf{C(p^0,U(q^t))=p^0q^{t\wedge}<=p^0q^t}$. However if the choice of quantities in the reference period is sub-optimal this no longer holds. Note also that estimating subindices for groups independently assumes that the $q_i^{t\wedge}$ in each group are determined independently. This is always true ex-post observed quantities q_i^t but that is not the same thing.

There are two approaches to approximating $\sum_{i\in j} u_i^0 * q_i^t / q_i^0$, *value deflation*, which is analogous to the Konus index approach, and *volume extrapolation*, which corresponds to the Allen.

If the quantity measures of the individual transactions are measured in the same units and economically comparable then the value shares will be approximately equal to the volume shares and the change in the sub index will equal the change in the sum of the

[12] The equivalent Konus/Paache expression is $\sum_i p_i^t q_i^t \Big/ \sum_i p_i^0 q_i^0 * \sum_j s_{tj}^t (\sum_{i\in j} u_{ti}^t * p_i^0 / p_i^t)$

quantities. This is referred to as volume extrapolation. In practise we are unlikely to have aggregates that cover all transactions so a National Accountants must make a subsidiary assumption that the change in the total quantity for the transactions in the group J can be approximated by the change in total quantity of transactions in a subgroup K giving us (14).

(14) $$\sum_{i\in j} u_i^0 * q_i^t / q_i^0 \approx \sum_{i\in j} q_i^t \Big/ \sum_{i\in j} q_i^0 \approx \sum_{i\in k\subset j} q_i^t \Big/ \sum_{i\in k\subset j} q_i^0$$

Note that the subgroup $\sum_{i\in k\subset j} q_i^t \Big/ \sum_{i\in k\subset j} q_i^0$ can only be calculated using quantities of things that are present in both periods. Anything available in the current period only will not affect the estimate. The only opportunity to allow for quality change is to put in an ad-hoc multiplier for the quantities. Even with product sets that are identical over time the requirement that all quantities have the same unit value in the base period is extremely demanding even when transactions are subdivided very finely. For this reason the common approach to estimating $\sum_{i\in j} u_i^0 * q_i^t / q_i^0$ is value deflation which relies on the relationship shown in (15)

(15) $$\sum_{i\in j} u_i^0 * q_i^t / q_i^0 = \sum_{i\in j}\left(p_i^0 q_i^0 \Big/ \sum_{i\in j} p_i^0 q_i^0 \right) * q_i^t / q_i^0 = \left[\sum_{i\in j} p_i^0 q_i^0 * q_i^t / q_i^0 \right] \Big/ \sum_{i\in j} p_i^0 q_i^0 =$$
$$\sum_{i\in j} p_i^0 q_i^t \Big/ \sum_{i\in j} p_i^0 q_i^0 \approx \left[\sum_{i\in j} p_i^0 q_i^t \Big/ \sum_{i\in j} p_i^t q_i^t \right] * \left[\sum_{i\in mt\subset jt} p_i^t q_i^t \Big/ \sum_{i\in m0\subset j0} p_i^0 q_i^0 \right]$$
$$\approx \left[P^0_{i\in l\subset j} \Big/ P^t_{i\in l\subset j} \right] * \left[\sum_{i\in mt\subset jt} p_i^t q_i^t \Big/ \sum_{i\in m0\subset j0} p_i^0 q_i^0 \right]$$

Again we have two stages of approximation, first we assume that the change in the total value of the transactions in the group J can be approximated by the change in total value of transactions in a subgroup M next that the value of the Paache price index calculated for all the transactions in J can be approximated by the ratio at 0 and t of some price index $P^t_{i\in l\subset j}$ based on the transactions in a subgroup L[13]. On the whole values are recorded more often and more accurately than quantities and price changes are much more highly correlated than value changes. The approximations in (15) can therefore be expected to be much better than those in (14). They also have the significant advantage that estimated volume totals in each period can include all the transactions in each period, even if they involve new products. The general superiority of value deflation is reflected in the Eurostat Prices and Volume manual (Eurostat 2002) which states that states that volume extrapolation techniques can only be an ideal or "A" method under certain "very stringent conditions."

Note that in practice, theory can provide no hard and fast guidelines for statisticians seeking to identify good approximations.

[13] Ideally $P^t_{i\in l\subset j}$ would be current weighted. In practise it is likely to be a Walsh index constructed with weights that apply to neither the base or the current period.

2.2 Applied economic indices of price change with hedonic quality adjustment

Both the predominance of the volume deflation method, and the fact that statisticians compiling price indices are usually more familiar with the economic theory underlying their indices than those supplying quantity indices, mean that almost all applied work on hedonic quality adjustment has been undertaken by price statisticians. This section considers the lessons we can learn from their practice and experience.

The first observation is that price statisticians use hedonics quite sparingly. Even the US Bureau of Labour Statistics, probably the most enthusiastic users among official statisticians, only applies them to a small fraction of their inflation measure and have only been using them since the early 1990s. The traditional, and still the commonest response to the problem of appearing and disappearing products is chaining, in effect ensuring that, at least for products with a rapid turnover, the period between any two comparisons is no more than a month and calculating an index for longer periods by multiplying all the monthly indices together. Where they have had to make explicit quality adjustments they have often resorted to simpler techniques such as multiplying by the change in weight to allow for a difference in pack size, adding an amount to the base period price to allow for the cost of the added features in the new product (referred to as option costing), or simply seeking an expert opinion of the value of the new product relative to the old. Nevertheless there are some markets, particularly the market for personal computers but also including clothing, televisions, white goods, and apparel in some countries, where new models[14] are appearing so fast that traditional methods are widely felt to be unsatisfactory.

The hedonic regressions used by price statisticians relate the price of an item p to the quantities of a vector of characteristics **x** i.e. they assume

(16) $$p(\mathbf{x}) = p(x_1, x_2, \ldots\ x_i \ldots\ldots x_n)$$

And estimate the regression with for a sample of products. It has been understood, at least since Rosen's influential 1974 paper (Rosen 1974), that the coefficients of such regressions cannot be interpreted as marginal utilities. Instead the hedonic price surface represents the envelope of the points of tangency of the *value functions* of consumers with varying tastes, incomes, and utility levels and the *offer functions* of producers with access to different technologies, production scales, and required profit rates. If producers' technology is identical the offer functions collapse and (16) is determined by and can be used to derive a cost function for production. Conversely if consumers' tastes are identical the hedonic function can be used to derive them. Diewert (Diewert 2003) following Muellbauer (1974) has provided a derivation from the consumer side that allows consumers to differ in their preferences for other goods as long as they have the same subutility function for characteristics. However this requires strong seperability, for example that relative preferences for the characteristics of cars are not affected by the number of children they have. Sometimes it is necessary to split consumers into groups with similar characteristics before this assumption can be

[14] Throughout this section we shall use the term "product" to refer to a whole category of goods for which an index is being calculated eg personal computers and "model" to refer to a specific instance of that product eg an IBM thinkpad Lenovo t60. All instances of a model are assumed to be identical.

applied comfortably. Some researchers have adopted this approach (see Tauchen and Witte 2001) but it has not been used in official statistics.

Worries about the theoretical underpinnings and of hedonics have lead price statisticians have tried to be very careful about how they use them. Triplett (2004) provides a survey of the different techniques.

The method that places most weight on hedonic functions is the *Characteristics index method* or the construction of price indices that treat the coefficients of the regression equation as shadow prices. These require the collection of data and the estimation of hedonic functions in every period and in real time. They also require that quantity weights for the characteristics are truly representative of the population. For these reasons practical applications are limited to cases where matching is almost impossible. The US house price index produced by the Bureau of the Census is probably the most prominent example[15]. The Bureau runs a nationally representative sample survey covering about 13,000 of the new one-family houses sold each a year and estimates a regression like (17) each quarter.

(17) $$Y_i = b_o + b_1 x_{1i} + .. + b_m x_{mi} + e_i$$

where

- Y_i is the logarithm of the sales price for house i (i=n where n is approximately 3,250)
- x_{1i} is the logarithm of the floor area for house i
- x_{2i} through x_{mi} are the values of qualitative (or dummy) variables (1 or 0)
- b_1, through b_m are the regression coefficients corresponding to each of the characteristics and b_0 is the constant in the regression
- e_i is the unexplained variation (error term)

using a robust, weighted regression technique resistant to outliers. They then estimate a Paache price index (among other things) using the formula.

(18) $$\mathbf{P^{0t}(p^t,p^0\ Q_t)} = 100*[\text{antilog}\left(\sum_{s=1}^{m} b_{st} Q_t\right)/\text{antilog}\left(\sum_{s=1}^{m} b_{s0} Q_t\right)]$$

where Q_t represents the proportions of the qualitative variables and the mean of the logarithm of the floor area in the current time period. Indices are calculated separately for five distinct geographical/house type segments of the market and aggregated to obtain national estimates.

A characteristics price index will give different result from an ordinary goods price index even if there are no appearing or disappearing goods. Many price statisticians find this disturbing. *Time dummy methods* are a way of avoiding the problem. They use hedonic regressions estimated with a semi-log specification using data for two or more periods with a dummy for every period. The change in the dummy is the price index. Because the coefficients for the characteristics must be the same in every period a time dummy method for will give identical results to a traditional method using a geometric mean elementary index when the universe of models doesn't change. Time dummy regressions are used to estimate elementary indices for several products in the German

---

[15] www.census.gov/const/www/constpriceindex.html

CPI using price and characteristics data collected from market research companies. However these methods still require the collection and estimation of new regressions in each period. Any method that produces an index directly from a regression estimate also imposes restrictions on the estimation methodology that may not be desirable. A price index must give most weight to the most significant transactions. If the index emerges directly from a regression, this can only be done by weighting the data that goes into the regression which may give a rather unstable regression dominated by a lot of middle of the road options preferred by the bulk of the population[16].

One method for avoiding the need to estimate a model every period is *Hedonic quality adjustment*. This method uses hedonic equations only when direct comparisons are not possible because an old model is not available and/or a new one must be introduced. The basic technique is to impute a base price for a new model of computer say by applying a quality change factor to the actual base price for the old model it replaces. The quality change factor is calculated as the ratio of the predicted price of the new model and the predicted price old model. This is the most popular hedonic method among statistics offices and is used in the CPIs of the US, France, Canada, and UK. The key operational advantage is that the dataset used for the hedonic regression can differ from that for measuring monthly price changes. This is important as the larger sample sizes and characteristics data required for producing regressions make these datasets much more difficult to collect than a price measurement sample even before we allow for the time required to produce a good regression. The counterpart of this advantage is the danger of producing hugely misleading results from regressions that no longer reflect the current world. In particular it is not sensible to estimate the price for a new model of computer with important characteristics that were completely unavailable or very expensive in the base period. The correct approach here is to calculate a new regression covering a period when this characteristic is more widely available[17]. The UK routinely tests its regressions against current observed prices to ensure that they remain relevant.

All these hedonic techniques are affected by basic issues of theoretical legitimacy, technical quandaries about the appropriate functional form for the regression equation, and practical issues about data collection. Working with them in practice has also raised several issues;

Classification turns out to be extremely important. The question of how new a product has to be to be considered “new” for instance can have a significant effect on results. If a product is merely of higher quality and hedonic methods are applied then the welfare gain from the appearance of that new quality is captured. If it is not then all that is captured is the price fall (if any) as producers increase production after the new product is introduced. The difference can be substantial. Moran (Moran 2006) for example shows the BLS in the United States treats wide screen and plasma screen TVs as merely higher quality versions of traditional tube TVs while the ONS in the UK treats them as a different product and obtains a substantially lower price fall. Careful subdivision of the market into segments for which stable and reasonable regressions can be obtained is

[16] This argument assumes that rarely purchased products are rarely purchased because they appeal to particular, refined tastes. If they are simply bad value for money then we would not want them in a regression for estimating the efficient frontier of the consumers’ consumption possibility set.

[17] There is a particular difficulty that new computer features tend to become widely available in one brand before they are available to the market as a whole. There is then a difficulty in disentangling the affects of these new features from brand effects.

also important. Indeed when qualitative or dummy variables are responsible for most of the variation in price segmenting the market and calculating different indices for each segment may give results similar to hedonics.

Estimation procedures have to be chosen to fit both the approach to using index construction and the market for which they are being used. A regression that will be suitable for hedonic quality adjustment is inappropriate for a characteristics index.

Chaining or the calculation of indices for widely spread periods by multiplying those for short periods together to ensure that the market has not changed too much between estimates should be applied wherever possible.

Data collection procedures have an important effect on methodological decisions. If data on characteristics is collected from a form then only information on characteristics listed on that form will be gathered. These may be relatively few and new characteristics may well not show up. If information is collected from the internet on the other hand it will be possible to test models with many variables and new characteristics will show up quickly but it is impossible to know whether offers for sale are popular or indeed whether anyone is taking them up at all.

Checking the continued relevance of regressions is vital. A characteristic that is highly valued in one year may become completely worthless in the next.

2.3 Applied indices of input quantities for "other individual non market services" with hedonic quality adjustment

What do the examinations of current practise in sections 2.1 and 2.2 tell us about how to apply the theoretical indices developed in section 1.3?

2.3.1 Disaggregation and the choice of expenditure categories

It is clear from the analysis in section 2.2 is that it is necessary to assume optimisation at some level in order to measure public sector outputs within a national accounts framework. If, in the notation of (13), we cannot use actual expenditure to estimate $s_{0j}^{0\wedge}$, we cannot begin to break down the national accounts problem into manageable chunks. In effect we have to assume that the share of GDP that the government is spending on a category of expenditure in the reference period represents the optimum outcome of a legitimate political process even if the way they are spending the money is sub optimal. In effect we are assuming that politicians have decided to spend the right share of GDP *given* the current level of allocational inefficiency in service delivery.

If we use actual costs to estimate $u_{0j}^{0\wedge}$ for non market transactions on the other hand, or work with $s_{0j}^{0\wedge}$ that are too finely disaggregated, we are assuming that government expenditure in the base period is fully optimised. An Allen approach to the growth in the quantity of output produced with the correct numerator will then be biased downwards as the denominator will be too high. This is of some practical importance as cost weighted activity indices have been widely proposed as the way of estimating non market output (see Eurostat ESA95). If the proportional level of inefficiency in the current and base periods is identical the bias will cancel out but if the government becomes more efficient by, say treating depression with more drugs and less psychotherapy, then the downward bias will remain[18]. This problem with cost weighted activity indices is actually quite well known, especially in the health literature, but the fundamental cause of over segmentation is rarely stressed and many practitioners appear to believe that more segmentation always implies better measures. Segmenting the market into categories that are too specific has the additional disadvantage of increasing the chance of new categories will arise. Categories that are too broad however, will be difficult to create robust hedonic regressions or a single scalar measure of output for.

2.3.2 Hedonic methods and functional forms.

Turning to the hedonic regressions themselves we seem to be more fortunate than price statisticians in that we do not have to use a single regression to obtain information about both the production of characteristics and their use. Instead we can estimate separate relationships for the production of observable outcomes by households and of observable characteristics by the government. Estimating these relationships, however still requires some troubling assumptions. Suppose for example we use a Cob-Douglas production function to estimate relationships like these

(19) $$\ln(\Theta_\eta) = a_o + \Sigma_{i=1:m} b_i \ln(x_{\eta i}) + \Sigma_{i=1:n} c_i \ln({\Xi^{\eta}}_i) + e_\eta$$

and

(20) $$\ln(C({\Pi^{r}}_\iota)) = d_o + \Sigma_{i=1:o} d_i \ln(x_{\iota i}) + e_\iota$$

where

Θ_η is the educational outcome achieved by each of H pupils
$x_{\eta i}, x_{\iota i}$ are those characteristics of the school the pupil attends affecting Θ_η
${\Xi^{\eta}}_i$ are the characteristics of pupil and/or their household which affect Θ_η
$C({\Pi^{r}}_\iota)$ is each of the school's total costs
e_i is the unexplained variation for pupils
e_ι is the unexplained variation for schools

There is an implicit assumption that the observation for each household represents a choice of the school with the characteristics offering the best educational outcome given the household's characteristics and also that all pupils turn school and household characteristics into educational outcomes in the same way. In effect we are assuming that any difference in educational outcomes for pupils with the same characteristics attending the same schools is random. On the school side we are assuming that variations in the costs of each school at any given time are entirely due to differences in the mix of characteristics they choose to offer and that all schools produce characteristics with the same technology. I.e. any difference in costs for schools producing the same characteristics is random. These assumptions would appear very restrictive in a market setting. In analysing government activity however it is difficult to see how we could say anything at all about whether or one situation was better than another without assuming there was some group who would order possible patterns of government activity in the same way. Similarly the suggestion that differences in costs are entirely caused by the unique unobservable circumstances of each school precludes any analysis at all. In many respects therefore non market hedonics is easier.

[18] The example is a hypothetical one often cited in the non market output literature. I have no idea if drugs are actually a more cost effective treatment for depression.

There are also some unique problems. Cost minimisation requires total cost and total outcome are respectively concave and convex functions of characteristics. Such requirements are difficult enough to obtain for individual firms but for us there is also the added complication that we are dealing with effects for many schools. Consider the estimated relationships from (21) and (22).

(21)

$$\text{Total predicted outcome of base households} = \sum_{\eta=1:H} e^{a_0} \prod_{i=1}^{m} x_{\eta i}^{b_i} \prod_{i=1}^{n} \Xi_{\eta i}^{c_i} = \text{the sum}$$

$$\text{of school effects weighted by household effects} = \sum_{\tau=1:\mathrm{T}} \left(\prod_{i=1}^{m} x_{\eta i}^{b\sim_i} \right) \left(\sum_{\eta \in \tau} \prod_{i=1}^{n} e^{a\sim_0} \Xi_{\eta i}^{c\sim_i} \right)$$

$$= \sum_{\tau=1:\mathrm{T}} K_\tau \prod_{i=1}^{m} x_{\eta i}^{b\sim_i}$$

and

(22)

$$\text{Total predicted cost} = \sum_{\tau=1:\mathrm{T}} \left(e^{d_0} \prod_{i=1}^{m} x_{\tau i}^{d\sim_i} \right)$$

where $a^{\sim}_o$, $b^{\sim}_i$, $c^{\sim}_i$ and $d^{\sim}_i$ are the estimated values of a_o, b_i c_i ,d_i and K_τ is the total of all the household effects in school τ.

A given change to a characteristic will have different effects on total output, and total costs, depending on the school in which it occurs. This makes it much harder to ensure that there is no characteristic in any school so effective in improving outcomes compared to the cost of producing it that optimisation results in a ‘corner solution’ piling everything into it .

2.3.3 Index construction

Whatever approach we take to estimating hedonic regressions there are still several ways to calculate indices. Tables 1. and 2. present some dummy data.

Table 1 Constructing a volume index for three schools and two characteristics

	Period 1				**Period 2**			
Schools	SDUa	SDUb	SDUc	Total	SDUa	SDUb	SDUc	Total
FactorX1	2	3		5	2	4	6	12
FactorX2	4	5		9	4	6	7	17
Pupil effect	20	25		45	20	15	7	42
Real outcome	42	79		121	43	50	34	127
Real costs	35	60		95	23	85	150	258
Parameters								
	$b_1=0.4$	$b_2=0.4$			$b_1=0.4$	$b_2=0.4$		
	$d_1=1.3$	$d_2=1.3$		$d_0=2$	$d_1=1.1$	$d_2=1.1$	$d_2=1.1$	$d_0=2.5$
Fitted Values								
Fitd. outcome	45.95	73.9		120	45.9	53.5	31.2	130.6
Fitd. costs	29.86	67.6		97.5	24.6	82.4	152.6	259.7
Relationships								
$\Theta_1^{\sim}=20x_{a1}^{0.4}x_{a2}^{0.4}+25x_{b1}^{0.4}x_{b2}^{0.4}$					$\Theta_2^{\sim}=20x_{a1}^{0.4}x_{a2}^{0.4}+15x_{b1}^{0.4}x_{b2}^{0.4}+7x_{c1}^{0.4}x_{c2}^{0.4}$			
$C_1^{\sim}=2(x_{a1}^{1.3}x_{a2}^{1.3}+x_{b1}^{1.3}x_{b2}^{1.3})$					$C_2^{\sim}=2.5(x_{a1}^{1.1}x_{a2}^{1.1}+x_{b1}^{1.1}x_{b2}^{1.1}+x_{c1}^{1.1}x_{c2}^{1.1})$			

N.B Parameters for the functions have been assumed rather than estimated from the actual outcomes and costs.

Table 1 explains the set up. The example is simplistic but still provides some insight into the proposed indices. Households turn characteristics into outcomes using a constant technology but the technology the government uses to produce characteristics changes in two ways. The exponentials on the inputs fall slightly, implying increased technical efficiency, but the multiplier e^{d_0} increases, implying a rise in the general price of inputs[19]. In the second period the government pours resources into a new school with a relatively small pupil effect. This is a highly inefficient way of promoting outcomes even though the actual outcomes of the school's pupils are higher, and its costs lower, than the fitted ones. The result is an explosion in costs that leaves the government producing way above the lowest cost.

Table 2 Minimum cost solutions to achieving various outcomes with different technologies and pupils

	Characteristics Produced						Costs	
	X1a*	X2a*	X1b*	X2b*	X1c*	X2c*	Minimised	Actual
Outcome of Period 1 (fitted =119.8, actual=121)								
Pupils Period 1, Technology of Period 1 (Allen denominator)								95.00
fitted	2.40	4.19	2.57	5.00			95.56	
actual	2.45	4.20	2.63	5.02			98.70	
Outcome of Period 2 (fitted=130.64, actual=127)								
Pupils Period 2, Technology of Period 2 (Konus deflator denominator)								258.00
fitted	3.56	6.77	1.33	11.99	0.96	5.58	151.61	
actual	3.37	6.67	1.24	11.98	0.90	5.57	140.26	
Pupils Period 2, Technology of Period 1* (numerator for both indices)								
fitted	4.55	4.28	1.96	7.23			157.54	
actual	4.40	4.12	1.84	7.18			143.71	

Indices based on actual outcomes and costs		**Actual Changes,** Outcome **=** 1.05
Allen=(143.71/98.7)=1.46	Konus= (258/95)* (143.71/140.26) = 2.78	Cost = 2.72, Deflated Cost = 2.17, X1 =2.4, X2=1.9

*This calculation assumes Period 2 Pupil effects are divided proportionately between schools a and b

Table 2 shows the minimum cost solutions used for calculating our two indices. The Allen divides the minimum cost of achieving the outcome of the second period with the minimum cost of achieving that of the first, with all calculations made using the schools and technology of period 1. The Konus multiplies the change in actual costs by the minimum cost of achieving the second period's outcome using the first's schools and technology divided by the minimum cost of achieving it using its own. Despite the large rise in the government's input costs, improved physical efficiency in producing characteristics means that the cost of obtaining the second period outcome efficiently is actually lower with the period 2 technology than with period 1's. The Konus deflator therefore does little to dampen the steep increase in the index caused by lower efficiency. If efficiency[20] was improving a Konus approach would mislead in the other

[19] Representing changes in the price of inputs as a single scaler, like the traditional input deflation approach, implies that the relative costs of producing the different characteristics are unaffected by changes in the relative price of the government's inputs. This is unlikely to be the case in practise but "joint production", handling a producer making multiple outputs with multiple inputs, is clearly beyond the scope of this paper.

[20] It is important to stress that the inefficiency we are talking about in period two has nothing to do with slacking or incompetence by service producers in the way they produce characteristics. Actual costs are

direction. Note that the normal difficulty of volume extrapolation failing capture the effects of new characteristics, doesn't apply. If the new characteristics result in improved outcomes then the minimum cost of achieving those outcomes with period 1 technology will rise. The Laspeyres perspective Allen approach therefore appears unambiguously better as a measure of non-market output *provided* the outcome measures in the two periods are comparable.

The indices we have just calculated correspond to the price statisticians' "Characteristics Price index" method. In effect we are quality adjusting every unit. Annother approach might be to measure the quantity change for service delivery units with unchanged characteristics purely by throughput, assuming that if the characteristics of the unit didn't change then the proportional inefficiency in costs didn't change either. However it does not seem likely that there would be many SDUs with completely unchanged characteristics in practice so some sort of characteristics approach is probably inevitable.

One of the controversial issues in price hedonics is when where and if it is legitimate to compare imputed with actual values, in effect how to treat residuals. In our case both the numerator and denominator are forecasts and there are always actual outcomes to use as references, however the residuals controversy still emerges in a different form. Both theoretical indices require us to estimate the cost of achieving a period 2 outcome for period 2 pupils with period 1 technology. If the school population is unchanged we can make our estimate with pupils attending the same schools they attended in period 2, although it will still be necessary to run a regression to estimate pupil specific effects if these are important and we believe they represent permanent but unobserved characteristics of the pupils rather than transitory random effects (good or bad days). What to do with the pupils from the schools that are new in the second period is less obvious but allocation of pupils to schools has an effect on the outcome and therefore the index.

One option would be allocating pupils from new schools to those present in both periods in proportion to the numbers in the second period. However this would leave no pupils for schools which close. Closing schools were presumably high cost for the output they produced so this might artificially lower the minimum total cost of producing the period 2 outcome and therefore the index. Probably the best approach would be to calculate the proportional allocation of pupil effect between the closing and continuity schools in period 1 and ensure it is maintained in the calculation for the period 2 pupils. However not all pupils provide the same proportional effect on outcomes. It will also probably be sensible to allocate pupil effect in proportion to pupil effect rather than pupils in proportion to pupils.

2.3.4 Operational issues

One of the great lessons of price statistician's work on hedonics is that collecting data for monthly estimates and model building can be both expensive and time consuming for results that sometimes make little difference to published estimates. We therefore have to give careful consideration to what information should be collected when.

---

very close to the production function and technical efficiency has increased. The inefficiency is in the mix

How often it would be necessary to estimate the effect of characteristics on outcomes would depend on how often those relationships change. While it does not appear likely that the effects of a given characteristic on outcomes will change very fast new characteristics may appear quite frequently in an area such as health. It might, however be possible to incorporate these into an outcome equation based on say clinical trials rather than rerunning surveys for every technique. In any case a Laspeyres approach implies that we will always have time to make such estimations. This may be important because some outcomes may take a long time to show themselves. It is quite easy to imagine both treatment techniques and medical treatments who's effects only become apparent years later.

The relationship between characteristics produced and costs is likely to change more often. However extensive information on the characteristics or activities of service delivery units however is already captured in administrative reporting systems. The main data difficulty in estimating cost regressions is more likely to be separating expenditure on immediate service delivery from investment and investment type spending (training, research etc.).

By far the most burdensome additional data gathering need will be timely current period information about pupil or patients characteristics that are relevant to outcomes. In many cases it may be necessary to collect data not only about the average characteristics of the user population but the characteristics of those using each facility. If these characteristics are at all relevant to outcomes however, any measuring or monitoring system that uses outcomes at all, for quality adjustment for instance, without monitoring changes will produce highly misleading results.

The most important factor determining the current period/real-time burden of compiling hedonic measures will be whether or not we have to allow for unexplained individual pupil or patient specific effects on outcomes. Facility or SDU specific effects are less significant as most facilities will be the same in both periods. If unexplained household effects are important it will be necessary to run regressions in the current period in order to estimate the cost of reaching a particular outcome with a particular population in the base period. If they can be ignored it would be possible to.

1. Look at characteristics of SDUs and households in the base period
2. Take the fitted outcome in the base period
3. Calculate the minimum cost of achieving it
4. Look at characteristics of SDUs and households in the current period
5. Take the fitted outcome in the current period
6. Calculate the minimum cost of achieving it

The role of the outcome measure is would then be as a diagnostic to ensure that he outcome regression was still working.

It is important to note that the importance of large household residuals in a single regression using sample data not necessarily indicate permanent unexplained household effects or that these effects are important at the national level. Identifying whether observed residuals are significant or not is a difficult technical question but exactly the sort of question that econometric methods are designed to answer.

3 Example

This section presents a more extended exploration of the feasibility of constructing output indices based on actual data. It is, however important to realise that his is still an exercise to examine a technique as opposed to a serious attempt at policy analysis and no policy implications can be drawn from the results.

3.1 The Data

The data in this exercise is drawn from an expenditure tracking and service delivery survey of the secondary education sector carried out by the Government of [an Asian developing country] in 2004. The survey was based on a nationally representative random sample of 219 schools of three types: government secondary schools, non-government secondary schools and non-government madrasas. Two features of the survey stand out. First, a detailed consumption module based on the Household Income and Expenditure Survey allows a robust measurement of the student households' welfare. Second, all Class 9 students present in the schools on the day their school was surveyed, took a learning achievement test in [the language] and Mathematics. The exercise was carried out by [] on behalf of the Government of [the Asian developing country] and analysed by Al-Samarrai1, Antoninis, Carraro and Rawle, from here on in referred to as AACR. (See FRMP 2005 and AACR 2005).

AACR fitted equations to estimate the determinants of student performance in the sampled secondary schools, defined as the score in the [language] and Mathematics tests for each type of school. However for this exercise we will only look at the estimates for Maths performance in government secondary schools. Table three shows the significant relationships. **F**urther information on AACR's estimation procedures and results is given in annexe 1.

Table 3 Summary of Significant Factors Affecting Pupil Performance in AACR regressions

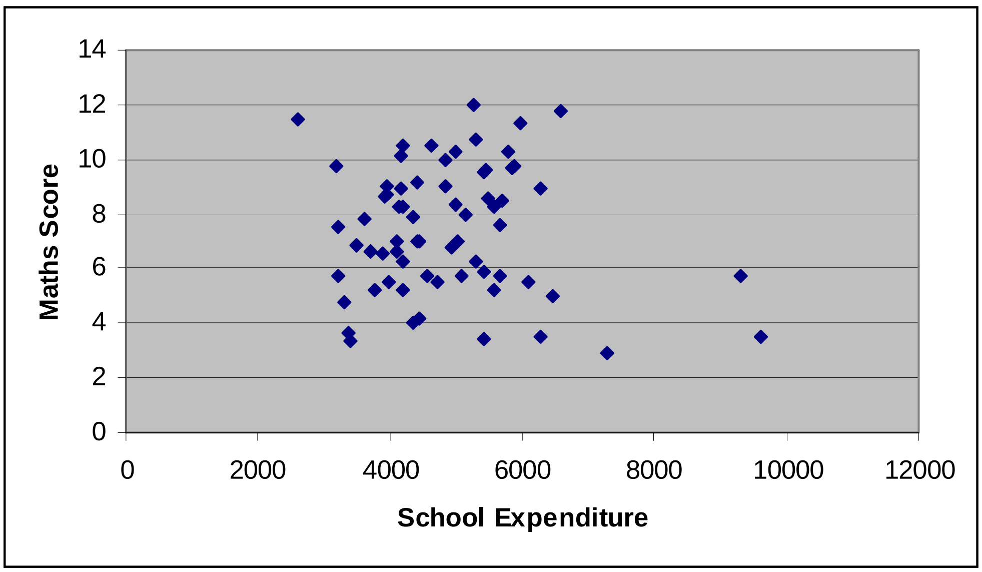
	Maths	
	Individual	**School**
Individual—Not specific to school choice		
Ravens* Progressive Matrices score	0.43	0.41
Number of years repeated in school	-5.6	-5.58
Household *per capita* expenditure (Tk; log)	-10	-9.62
Individual - Specific to school choice		
Household education expenditure on student (Tk; log)	6.2	
Science stream	6.33	6.7
Class 6 Mathematics and [lang] average school test score	0.79	0.88
School Specific		
Single gender school		-6.69
Annual school non-salary expenditure per student (Tk; log)		8.52
Head teacher has received managerial training		2.89
Proportion of school teachers with professional qualification (%)		0.1
Rural location		3.6

N.B Coefficients with the wrong sign for teachers' experience and salary expenditure have been set to 0
*This is a test administered to test the student's untaught or raw ability

Unsurprisingly AACR found that performance in the in the SSPS or social sector performance survey test was positively correlated with intelligence as measured by the Ravens progressive score and negatively correlated with the numbers of years of schooling that pupils repeat. Interestingly the log of household per capita expenditure appeared negatively correlated with output, possibly because only a small range of relatively well off students actually manage to get into class 9 in the first place. The amount households spent on education, possibly an indication of parental commitment, was important for individuals. Whether or not students were in the science stream mattered, and the class 6 average score in the same school (meant as a proxy of how well prepared pupils are on entry). As far as school characteristics went single gender schools appeared to do significantly worse in mathematics, rural schools did better, and non salary expenditure and training for the headmaster appeared particularly important. It is important to note however that even if all the non-significant variables are included in the regression the results only explains just over half of the test results. That is there are important pupil specific effects that are not being picked up.

The same schools dataset also contains information on the expenditure of the schools producing these characteristics. As chart 1 shows it is not particularly highly correlated with test results.

Chart 1 Weighted average of maths score in SPP test v expenditure per pupil



Following the methodology outlined in section 2.3, a regression of the form from (21) has been run to look at the correlation with outcome characteristics, and minimum solutions developed.

3.2 Results

Table 4 Effects of school characteristics in outcome and cost Regressions

Lables	Characteristic names	Outcome Regression	Cost Regression
A	Single gender school	-6.69	0.06
B	Annual school non-salary expenditure per student (Tk; log)	8.52	0.17
C	Head teacher has received managerial training	2.89	0.03
D	Proportion of school teachers with professional qualification (%)	0.1	0.0006
E	Rural location	3.6	-0.13
F	Constant		12.9
	R-squared		0.8792
	Number of Schools in Costs regression		67

Outcome and Cost relations for each school

Outcome = -6.69A + 8.52B + 2.89C + 0.1D + 3.6E + pupil specific factors

Ln(Cost) = 0.06A + 0.17B + 0.03C + 0.0006D - 0.13E + 12.9 + school specific factors

N.B. details of the cost regression are given in Annexe 2

Table 4 presents the coefficients from the cost regression alongside those for the outcome regression. It is unweighted because we wish to examine the effect of different characteristics combinations on school costs rather than make forecasts. While the cost relationship performs relatively well at predicting schools' costs it is dominated by the constant term. Annexe 2 gives more details about estimation.

The results suggest that single gender schools, which lower outcomes, actually raise costs while a rural location, which raises outcomes, lowers costs. The implication is that schools could provide exactly the same quantities of beneficial characteristics (actually more) at lower cost. In terms of fig. 2 they are off their iso-cost curves. Consequently the regression can't tell us about their optimising behaviour. Fortunately there are still some beneficial characteristics that appear costly. As well as non salary costs it appears intuitively sensible that hiring head teachers with management training and professionally qualified teachers would also raise costs. For the purposes of this exercise we will simply assume that the single gender and rural location variables remain unchanged[21].

The optimisation problem then becomes one of choosing annual school non-salary expenditure per student, management trained head teachers and the proportion of professionally trained teachers for each of the sixty seven schools to minimise the weighted sum of costs while achieving the reference outcome level. The reference outcome level for the current period itself is simply the weighted sum of the actual test scores which turns out to be 16,813,995. The simplistic but satisfying finding is that cutting non salary expenditure by 40% and putting the funds into hiring head teachers with management training and a fully qualified staff would allow schools to maintain this score while achieving a 4.2% fall in total costs.

[21] In a real world analysis we could try estimating efficient production frontiers.

Table 5 Optimising in period 1

	Observed	
	Cost /1,000	**Score /1,000**
Observed Situation	1,305,180	16,808
Optimised Values		
Minimised cost	1,250,068	16,808
Maximised output	1,305,180	17,476

Note though that comparisons across time will be affected by changes in the pupil effect. It is therefore particularly troubling that individual pupil effects play such a large part in explaining outcomes. It is likely that a lot of what appears as a pupil specific residual in the AACR regression would be revealed as school specific with an alternative estimation strategy, however this would only reduce the problem, not eliminate it. Note also that even in an example as simple as this the fact that each school has a different weighting means that moving a given amount of total pupil effect from a low weight school to a high weight one will raise output without changing costs and so reduce the minimum cost of hitting a given output target. The assumed calculation and distribution of pupil or other household specific effects therefore becomes an important part of calculating non market output from the Allen/Laspeyres approach. The position is similar to that when there is inconsistency over time in the output measure as is the solution of resorting to the Konus/Paache approach.

3.3 Measurement and monitoring

As well as providing growth measures and identifying opportunities for reducing costs, outcome and cost equations can be used to design a framework for monitoring and performance measurement. The target is obvious. Ideally service delivery units would produce characteristics that would maximise outcomes given the amount of money they are allocated, the characteristics of the households that will use their services, and the outcome and cost equations that will be available to them. In terms of our [asian dev country] example we would want them to choose the annual school non-salary expenditure per student, management trained head teachers and the proportion of professionally trained teachers for each school to maximise the weighted test score for a given set of households, outcome, and cost equations for their secondary education budget.

If we knew exactly what the household characteristics, output and cost equations would be in the second period we could calculate the maximum score and give it to the ministry of education as a target. We could also calculate all the optimised combinations. Imagine a situation in which the schools in our sample were offered a ten percent budget increase and the outcome and cost parameters don't change. The maximum possible score they could achieve is about 12.5 % more than they are getting at the moment. Because the minimum cost of achieving any particular score increases with the score aiming for this target will also be equivalent to maximising the Allen quantity index. Table 6 shows the observed and optimised results. If we knew the current parameters we could set targets in terms of characteristics or parameters and the effect would be the same.

Note that because we are not starting from an efficient position maximising an index of cost weighted characteristics will not produce an optimal solution. Even if we were in a cost minimising position to start with the ideal mix of characteristics with the new budget will not be the same as with the old if there are economies of scale in producing any of them. Note also that as both technology and households are unchanged the deflator term in the Konus deflator is equal to one by definition so it will give a value equal to the change in current costs, not a particularly desirable target.

Now consider a situation where the schools are offered the same budget increase but face increases in the cost parameters of materials and teachers' wages, slightly offset by an improvement in pupils learning abilities (which could be innate or due to say increased income). The maximum score the schools could achieve with the budgeted resources is now 93% of the base score. Again the Allen index will rise with output so maximising it will provide the right incentives.

Table 6 Monitoring Example

	Period 1			Period 2		
	Coefts.	**Cost /1,000**	**Score /1,000**	**Coefts.**	**Cost /1,000**	**Score /1,000**
Observed Situations						
Baseline		1,305,180	16,808			
85% cut in non salary spend					1,435,698	15,616
Maintaining target characteristics for period 1 coeffts					1,986,601	19,171
Outcome and Cost Parameters						
Average Pupil Effect*	2,849			3,174		
Non-salary expenditures (log)	0.167706			0.217706		
% qualified teachers	0.000622			0.000822		
Optimised Values						
Minimised cost		1,250,068	16,808			
Maximised output		1,305,180	17,476		1,435,698	15,623
Maximised output, Period 1 coefficients					1,435,698	18,916

* The average pupil effect is provided for explanatory effects only. Actual results of changes to the pupil effect depend on its distribution

Suppose however that the change to the parameters was unanticipated. If the schools maintained their original target characteristics they would exceed their score target but show large budget overruns. If they cut their budgets for non salary expenditure across the board they could meet their budgetary target but at an unnecessarily high cost to performance. The only way to reach the optimum solution in the face of changed parameters is actually to re-optimise. Table 6 shows the different situations.

It I also interesting to consider the case of an education ministry offered an output target in terms of the procedure outlined in section 2.3.4. To recap the proposed stages in calculating the measure are.

1. Look at characteristics of SDUs and households in the base period
2. Take the fitted outcome in the base period
3. Calculate the minimum cost of achieving it
4. Look at characteristics of SDUs and households in the current period

5. Take the fitted outcome in the current period
6. Calculate the minimum cost of achieving it
7. Test the outcome regression to see of it still works.

A forecast of user characteristics in the second period would allow an ideal second period position to be identified. It could then be specified as a set of characteristics targets *or* a single outcome target. Setting an outcome target however would have the distinct advantage of allowing service providers to respond to changes in the characteristics of their users. If all parties were operating with an agreed outcome function then the outcome changes caused by differences in user characteristics would not cause Service providers to miss their volume output target. Failing users might complain that the outcome relationship used in the base period was no longer valid but this is something that, at least in principle, it is possible to test

4 Conclusions

This paper has

- Shown that, given the existence of a scalar measure of *outcome*, it is possible to specify theoretical quantity indices for the *output* of non market service delivery units in the same way as theoretical quantity indices are specified in the market sector.

- Described in general terms the assumptions that have to be made to estimate these theoretical indices in practice and compared them with the assumptions that have to be made when estimating of quantity indices for the market sector.

- Provided an example based on an actual dataset for secondary schools that demonstrates the feasibility and usefulness of constructing the proposed measures. It also provides further insight into the data that needs to be collected for their compilation and the technical problems that remain to be solved.

Although the paper has shown the potential benefits of non market output measures derived from the theory of economic indices it has also highlighted some difficulties. In particular it appears that, although it is now common to consider government as a service producer, there is still a lot of work to be done developing well behaved production functions for government activities. Without such functions it is difficult to estimate how far the government is from cost minimisation. There are obvious implications beyond creating improved measures for statisticians.

Annother theme that runs through the paper is the definition of quantity indices in terms of optimisation by a representative household. This has some rather strict implications for outcome measures. First they have to be things that households would wish to maximise as an end in themselves rather than intermediate goals like cleanliness of hospitals or numbers of text-books. Also they have to avoid explicit equality targets, indeed it has to be possible to add the outcomes of different households. This is actually the case with the market sector quantity measures in national accounts which are created by simply adding all transactions regardless of who is making them. Many non market organisations however have aims about distribution. Incorporating distributional goals within the National Accounts tradition is difficult but certainly possible (see Sen 1979). The approach that has generally been adopted is to create poverty measures as an addition to the accounts rather than some sort of equality adjusted growth estimate. However extending this approach to nonmarket consumption will not be able to rely on the assumption that the balance between consumption of different "products" represents optimisation by the households in the way that purely market based poverty lines can**.**

Finally it is certainly worth investigating the properties of the suggested functions from a purely axiomatic point of view.

Annexe 1 – AACP regression for GSS

AACR fitted equations to estimate the determinants of student performance in the sampled secondary schools, defined as the score in the [Language] and Mathematics tests. The analysis was complicated by the necessity to control for the non-random assignment of students to different types of schools, which may bias the estimated coefficients for the factors explaining student achievement on the tests. In order to address this problem, a generalised approach of the Heckman two-step procedure to selectivity bias correction was used[22].and a selection term included in the performance regression. Performance equations were then estimated using ordinary least squares with the variance-covariance matrix corrected to account for the tendency of residuals from the to be correlated for students within the same school[23]. Each observation was weighted using sample weights in order to summarise the characteristics of the population. The final regression was estimated using a simple additive form with the characteristics adding to explain the test scores.

The procedure used by AACR to calculate selection terms utilises estimates from a multinomial logit model (MNL) rather than a probit to construct the set of selection correction terms. At first, a reduced form MNL is estimated for the three shool types categories to obtain the parameter estimates. The predicted probabilities for each individual i = 1, ..., N for each school type are computed. The standardised selection probabilities for each individual for each school type are then computed using the inverse standard normal operator are then computed. Finally the following correction term is constructed for each category and school type:

$$\lambda_{\text{i, school type}} = \frac{\phi(z_{\text{i, school type}})}{P_{\text{robability of i attending scool type}}}$$

where f(•) denotes the probability density function for the standard normal.

These selection terms are then added to the xi vector in the regression models estimated separately for the achievement models. A number of instruments have been used to assist in identification of the parameters of the selection effects. These need to be able to shift the probability of school choice but not the level of achievement on the [Language] and Mathematics test scores. In order for this procedure to be valid, the Independence of Irrelevant Alternatives (IIA) assumption of the MNL has to be satisfied. Tests of the IIA assumption based on the Small-Hsiao tests are reported in table A2 and provide evidence that the IIA assumption is not violated[24].

Tables A1 and A2 show their results

[22] The generalised model is described in L. Lee (1983) 'Generalized models with selectivity', Econometrica Vol. 51, pp.507–512. The original model outlining the original two-step procedure was J. Heckman (1979) 'Sample selection bias as a specification error', Econometrica, Vol. 47, pp.153-161. For an education application of Lee's approach, see G. Lassibille and J-P Tan (2001) 'Are private schools more efficient than public schools? Evidence from Tanzania', Education Economics, Vol. 9(2), pp.145-172.

[23] See H. White (1980) 'A heteroscedastic-consistent covariance matrix estimator and a direct test for heteroscedasticity', Econometrica Vol. 48.

[24] The test is described in K. Small and C. Hsiao (1985)

Table A1 Determinants of student performance in Government secondary schools

	Mathematics		[language]	
	Individual	**School**	**Individual**	**School**
Individual—Not specific to school choice				
Father with primary education	1.66	3.70	0.06	2.01
	(5.42)	(5.69)	(3.04)	(3.52)
Father with secondary education	1.93	1.96	3.37	2.50
	(4.48)	(4.64)	(3.56)	(3.58)
Father with higher secondary education or more	5.43	3.68	1.74	0.51
	(4.41)	(4.51)	(2.92)	(2.70)
Student not living with father	4.19	3.95	3.78	4.58
	(5.82)	(5.63)	(4.54)	(4.76)
Age	-0.94	-0.27	-0.73	0.00
	(1.52)	(1.60)	(1.12)	(1.11)
Female	1.81	3.04	-2.28	-0.47
	(3.25)	(3.62)	(3.08)	(2.93)
Ravens Progressive Matrices score	0.43	0.41	0.40	0.37
	(0.17)***	(0.18)**	(0.13)***	(0.12)***
Number of years repeated in school	-5.57	-5.58	-6.29	-6.68
	(3.26)*	(3.27)*	(2.26)***	(2.38)***
Household *per capita* expenditure (Tk; log)	-9.96*	-9.62*	-1.20	0.94
	(5.51)	(5.73)	(3.37)	(4.14)
Individual—Specific to school choice				
Student attendance rate Jan-Jun 2004 (%)	-0.10	-0.04	0.01	0.08
	(0.08)	(0.10)	(0.06)	(0.05)
Hours of private tuition per week	0.08	0.25	0.18	-0.02
	(0.18)	(0.23)	(0.14)	(0.15)
Household education expenditure on student (Tk; log)	6.20	4.43	1.83	2.15
	(2.84)**	(3.11)	(1.87)	(1.81)
Stipend recipient	1.27	1.15	8.55	8.44
	(3.74)	(3.75)	(2.74)***	(2.90)***
Science stream	6.33	6.70	4.28	4.41
	(2.13)***	(2.25)***	(2.09)**	(2.12)**
Class 6 Mathematics and [language] average school test score	0.79	0.88	0.33	0.60
	(0.19)***	(0.25)***	(0.15)**	(0.25)**

	Mathematics		[language]	
	Individual	**School**	**Individual**	**School**
School characteristics—General				
Rural location		3.60		11.11
		(4.52)		(3.81)***
Total school enrolment		0.003		-0.002
		(0.005)		(0.003)
Class 9 average daily teaching time		-0.05		0.03
		(0.04)		(0.05)
Single gender school		-6.69		1.09
		(3.52)*		(3.90)
School characteristics—Financial				
Annual school salary expenditure per student (Tk; log)		-14.34		-5.24
		(4.62)***		(4.72)
Annual school non-salary expenditure per student (Tk; log)		8.52		3.88
		(4.90)*		(4.03)
School characteristics—Material				
Class 9 size		-0.05		-0.04
		(0.07)		(0.05)
School characteristics—Managerial				
Head teacher years of experience as teacher		0.05		0.12
		(0.15)		(0.13)
Head teacher has received managerial training		2.89		-3.87
		(2.86)		(2.20)*
Frequency of PTA meetings		-3.86		2.11
		(5.03)		(4.47)
School characteristics—Teachers				
Average years of experience of school teachers		-0.18		-0.78
		(0.45)		(0.26)***
Proportion of school teachers with professional qualification (%)		0.10		0.20
		(0.11)		(0.09)**
Selection term	-5.15	-4.00	-2.62	-3.10
	(6.14)	(6.16)	(4.00)	(4.16)
Constant	22.54	78.39	31.67	-16.67
	65.46	82.34	42.83	66.28
Observations – students	265	265	265	265
Observations – schools	61	61	61	61
R-squared	0.51	0.54	0.49	0.55

Notes: (a) Standard errors are reported in brackets; (b) Robust standard errors, adjusted for heteroscedasticity and the school-level clustering of the data, are in parentheses; (c) One asterisk implies statistical significance at 10%, two asterisks at 5% and three asterisks at 1%; (d) Sample weights are used in estimation.

Annexe Table A2 determinants of school type choice (Govt Secondary School default choice)

	NGSS	DM
Individual characteristics—Not specific to school choice		
Father with primary education	-0.41	-0.77
	(0.54)	(0.56)
Father with secondary education	-0.71	-1.05
	(0.50)	(0.55)*
Father with higher secondary education or more	-0.82	-0.56
	(0.46)*	(0.51)
Student not living with father	0.04	-1.11
	(0.58)	(0.66)*
Age	-0.08	-0.01
	(0.17)	(0.19)
Female	-0.03	0.21
	(0.43)	(0.46)
Ravens Progressive Matrices score	-0.06	-0.07
	(0.01)***	(0.02)***
Number of years repeated in school	-0.21	-0.56
	(0.29)	(0.40)
Household *per capita* expenditure (Tk; log)	-0.96	-1.28
	(0.49)*	(0.51)**
Identifying variables in selection model		
Household has electricity	-1.10	-1.50
	(0.50)**	(0.51)***
Household owns cattle	0.91	1.49
	(0.44)**	(0.49)***
Number of household members currently in higher education	-0.59	-0.99
	(0.34)*	(0.38)***
Constant	14.15	13.67
	(3.62)***	(4.52)***
Observations	945	
Pseudo R-squared	0.13	

The table reports the results of a multinomial logit regression. Small-Hsiao test of Independence of Irrelevant Alternatives: (i) eliminating GSS: 8.26; (ii) eliminating NGSS: 8.01; (iii) eliminating DM: 12.25.

Annexe 2 – Costs regression

Table A3 – Regression of costs on school characteristics significantly affecting school performance

Unweighted OLS regression with robust standard errors. Note t statistics lousy but its only an exercise

	Coef.	**Std. Err.**	**t**	**P>t**	**[95% Conf.**	**Interval]**
Single gender school	0.06112	0.09384	0.65000	0.5170	- 0.12658	0.24883
Rural location	- 0.13470	0.07507	- 1.79000	0.0780	- 0.28487	0.01547
Head teacher received managerial training	0.02912	0.07057	0.41000	0.6810	- 0.11203	0.17027
Annual school non-salary expenditure per student (Tk; log)	0.16771	0.10755	1.56000	0.1240	- 0.04743	0.38284
Proportion of school teachers with professional qualification (%)	0.00062	0.00204	0.31000	0.7610	- 0.00345	0.00470
Total Number of Students	0.00097	0.00007	13.29000	-	0.00083	0.00112
Constant	12.86902	0.76857	16.74000	-	11.33165	14.40638

Number of obs = 67 R-squared = 0.8792
F(6, 60) = 54.33 Root MSE = .24668
Prob > F = 0.0000

References

Allen, R.G.D., 1949. "The Economic Theory Of Index Numbers," *Econometrica N.S. 16, 197–203.*

Al-Sammarail, Antoninis, Carro, And Rawle – "Assessing The Effect Of Secondary Education Policy On Poverty Reductin In [an Asian Developing country]". *Conference Paper CSAE 2006*

Balk 1999 "On Curing The CPI's Substitutionand New Goods Bias" *Bert M. Balk Department Of Statistical Methodsstatistics Netherlandse-Mail Bblk@Cbs.Nl*

Burnstein 1961

Diewert 1981 :"The Economic Theory Of Index Numbers: A Survey " *Essays In The Theory And Measurement Of Consumer Behaviour in Honour Of Sir Richard Stone, Edited By* A. Deaton, London: Cambridge University Press, 1981,

Diewert 2003 "Hedonic Regeressions, A Consumer Theory Approach" *In R.C. Fenestra and M.D. Shapiro (Eds.) Scanner Data and Price Indexes, NBER Studies In Income And Wealth* (University Of Chicago Press Pp. 317-348

Eurostat 2002 "Handbook On Price And Volume Measures In National Accounts"

Eurostat 1995 "European System Of Accounts 1995"

FRMP – "Secondary Education In [] Assessing Service Delivery" *Final Report Sept. 2005*

Heckman J. (1979) 'Sample Selection Bias As A Specification Error', *Econometrica*, Vol. 47, Pp.153-

Konus 1924 Konus, A.A., 1924. English Translation, Titled "The Problem Of The True Indexof The Cost Of Living," Published In 1939 In *Econometrica* 7, 10–29.Konuus, A.A. And S.S. Byushgens, 1926. "K Probleme Pokupatelnoi Cili Deneg"

Lassibille G. And J-P Tan (2001) “Are Private Schools More Efficient Than Public Schools? Evidence From Tanzania’, *Education Economics*, Vol. 9(2), Pp.145-172.

Lee L. (1983) ‘Generalized Models With Selectivity’, *Econometrica* Vol. 51, Pp.507–512.

Mackay K. 2006 "How To Build M&E Systems To Support Better Government" World Bank 2006

Malmquist, S., 1953. “Index Numbers And Indifference Surfaces,” *Trabajos Deestatistica* 4, 209–242.Mcfadden, D., 1978a. “Cost,

Moran UK And US Approaches To Hedonic Price Indices For TVs *Paper For Ottawa Group* 2006

Muelbauer J. 1974 "Household Production Theory, Quality And The Hednic Technique" *American Economic Review* Vol 64 No.6 Pp977-994

ONS: (Http://Www.Statistics.Gov.Uk/Articles/Nojournal/Ukcemga_Strategy_Paper.Pdf)

ONS: “Attkinson Report: Measurement Of Government Output And Productivity For National Accounts” ONS 2005

Rosen S. 1974 "Hedonic Prices And Implicit Markets, Product Differentiation And Pure Competition" *Journal Of Political Economy*, Vol 82 Pp 34-49

Royal Statistical Society 2004 “Working Party On performance Monitoring In The Public Services Performance Indicators: Good, Bad, And Ugly”

Sen, A “The Welfare Basis Of Real Income Comparisons: A Survey” *Journal Of Economic Literature,* Vol. 17, No. 1. (Mar., 1979), Pp. 1-45.

Small K. And C. Hsiao (1985) ‘Multinomial Logit Specification Tests’, International Economic Review, Vol. 26(3), Pp.619-627.

Tauchen H And A.D. Witte 2001”Estimating Hedonic Models Implications Of The Theory”, *Technical Working Paper No. 271* Available At Nber.Org

Triplett 1983 "Concepts Of Quality In Input And Output Price Measures :A Resolution Of The User-Value Resource-Cost Debate" In M.F. Foss Ed The U.S National Income And Product Accounts: Selected Topics, *NBER Studies In Income And Wealth* University Of Chhicago Press Pp 269-311

Triplett 2004 "Handbook On Hedonic Indexes And Quality Adjustments In Price Indexes: Special Application To Information Technology Products" *Sti Working Paper* 2004/9 OECD

Varian 1992

White H. (1980) 'A Heteroscedastic-Consistent Covariance Matrix Estimator And A Direct Test For Heteroscedasticity', *Econometrica* Vol. 48.