European Commission

JRC TECHNICAL REPORTS

Institute for Prospective Technological Studies
Digital Economy Working Paper 2013/01

# An Appropriate Model for the Estimation of Consumer Time Expenditure Patterns on the Internet

Russel Cooper
University of New South Wales, Canberra Campus, Australia

2013

Report EUR 25836 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for Prospective Technological Studies

Contact information
Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: jrc-ipts-secretariat@ec.europa.eu
Tel.: +34 954488318
Fax: +34 954488300

http://ipts.jrc.ec.europa.eu
http://www.jrc.ec.europa.eu

This publication is a Working Paper by the Joint Research Centre of the European Commission. It results from the Digital Economy Research Programme at the JRC Institute for Prospective Technological Studies, which carries out economic research on information society and EU Digital Agenda policy issues, with a focus on growth, jobs and innovation in the Single Market. The Digital Economy Research Programme is co-financed by the Directorate General Communications Networks, Content and Technology.



JRC79022

EUR 25836 EN

ISBN 978-92-79-28780-0 (pdf)

ISSN 1831-9424 (online)

doi:10.2791/81009

Luxembourg: Publications Office of the European Union, 2013



Printed in Spain

# Table of Contents

## Summary

IPTS recently acquired a consumer internet clickstream database containing the full set of annual (2011) clickstream records for about 25.000 internet users in the five largest EU economies (UK, Germany, France, Italy and Spain). It contains time spend on each webpage and socio-economic characteristics of the internet users.

This study describes a model of consumer Internet time use that is capable of empirical implementation with the clickstream database. There is a natural model of Internet search time that could be developed to fit neatly with the database as it currently exists. The basic structure of such a model has been discussed in detail.

The study recommends that IPTS should not rely on any existing model of Internet time expenditures. There does not appear to be a good contender in the literature that would adequately address the requirements of IPTS nor the realities of available data. A purpose-built model needs to be considered. Features that could be built into a staged development of a purpose-crafted empirical model include:

Stage 1: A stand-alone model of Internet time use using the Internet as a search engine for traditional products and specifically geared to utilizing the IPTS clickstream database

Stage 2: Incorporate a set of additional considerations that could include:

i) A model which integrates Internet search time use with broad sub-aggregates of traditional consumer expenditure
ii) A model which endogenises total Internet time use
iii) A model which distinguishes Internet time that substitutes for traditional consumption from Internet time spent in search of traditional products
iv) A model which accounts for new products that become available following ICT innovations
v) A model which specifically recognizes the role of innovations in ICT in upgrading the quality of traditional goods and services
vi) A model which accounts for ICT- induced externalities and network effects, including consumer complementary use of public ICT infrastructure, and assesses consumer welfare implications

Investigations of approaches to integrate the IPTS clickstream database with regionally and demographically differentiated household expenditure survey data should proceed concurrently with the empirical modelling, especially with the recommended work in Stage 1, so that an extended database will be available for application with the Stage 2 model developments.

## Introduction

One of the objectives of the Digital Economy research programme at IPTS is to estimate the economic value of the digital economy and in particular the internet economy. Almost all existing economic research on this question starts from the market value of digital economy goods and services. The problem with transposing this issue to the internet economy is that many services delivered to consumers over the Internet are free and have no positive market price. Alternative approaches have to be developed that consider the opportunity cost of these free services, for producers and consumers. Here we focus on the estimation of the value of free internet services for consumers, starting from the opportunity cost of the time they spend on the internet. We start with a review of an existing model and then propose an alternative model.

## 1. Review of the Goolsbee-Klenow time use model

Arguably a major contender for a prototype model of consumer time spent on the Internet would be some variation of the model published by Goolsbee and Klenow, 'Valuing Consumer Products by the Time Spent Using Them', *AEA Papers and Proceedings* (May, 2006), pp. 108-113.

The model employed by Goolsbee and Klenow (G-K) treats both internet and non-internet products as requiring time for consumption. Although it would be possible, and advisable, to integrate the IPTS clickstream database with non-internet consumption data available for example from household expenditure surveys, most such surveys do not contain information on time spent on consumption. Hence it is very unlikely that time spent on consuming non-internet purchases will be available in the degree of detail required for an effective integration of data from these sources. For this reason, the value of discussing the G-K model arises more from the perspective of extracting ideas that can be effectively employed in constructing models either utilizing the clickstream database alone (in the first instance) or in generating a model using an integrated database in which the non-Internet data indicates money spent, not time spent.

There are other reasons why the G-K model can serve only in a limited way for current purposes. First, it is highly aggregated. Second, it was designed for a specific purpose - to generate an alternative estimate of an important substitution elasticity from those available at the time in the literature. For this purpose, a number of approximations were made in moving from the model specification to the estimating form. In the context of other requirements of their model, it may not be advisable for users to go down the same approximation path. Finally, there are further approximations involved when Goolsbee and Klenow move from parameter estimation to an application involving a welfare calculation. Again, users in other contexts may not find it advisable to make these approximations.

This section sets out some characteristics of the G-K model. Lessons from examination of the model are employed in the proposal for a purpose-built model that then follows.

### 1.1. The Underlying Preference Structure in the Goolsbee-Klenow Model

The G-K model specifies a direct utility function (DUF) that is a CES function of nested Cobb-Douglas (CD) functions. There are four choice variables in total, $C_I$, $L_I$, $C_O$ and $L_O$. The DUF is:

$$U\left(C_I, L_I, C_O, L_O\right) = CES\left(Y_I, Y_O\right) \tag{1}$$

where

$$Y_j = CD\left(C_j, L_j\right),\ j = I, O \tag{2}$$

Specifically,

$$\begin{aligned} U &= \theta Y_I^{1-1/\sigma} + (1-\theta) Y_O^{1-1/\sigma} \\ Y_I &= C_I^{\alpha_I} L_I^{1-\alpha_I} \\ Y_O &= C_O^{\alpha_O} L_O^{1-\alpha_O} \end{aligned} \tag{3}$$

In the G-K specification, time spent is treated as a complement to consumption - both of the ‘Internet’ product, I and of the ‘Other’ product O. The optimisation problem is specified as a static or atemporal allocational exercise, and $U$ is maximised subject to the budget allocation constraint where the income to be allocated among the four commodities is the wage per unit of time, $W$, less the fixed cost of Internet connection (over the same period of time), $F_I$:

$$P_I C_I + W L_I + P_O C_O + W L_O = W - F_I \tag{4}$$

In this specification, time spent consuming is understood to generate utility. This is true for both the Internet (I) and the Other (O) product.

The Cobb-Douglas specification in the inner nest for the money and time expenditures on any product means that both are necessary for a product to yield any utility at all. On the other hand, the CES specification for the outer nest means that a consumer can in principle live on 'bricks and mortar' alone or, alternatively, on 'virtual consumption' alone. More generally it would be worth investigating whether a generalised CES specification might be more appropriate over the four consumption items: (i) money-spent-on-bricks-and-mortar, $C_O$ ; (ii) time-spent-on bricks-and-mortar, $L_O$ ; (iii) money-spent-on-the-Internet, $C_I$ ; and (iv) time-on-the-Internet, $L_I$. To investigate this rigorously, one would have to specify a more general model than G-K use, and derive the appropriate choice equations for empirical work.

### 1.2. Some initial observations: the G-K concept of time and its value

*Time spent may generate disutility, not utility*

The G-K specification does not cover the possibility that time spent may give disutility, as could be true in some (perhaps even many) cases - e.g. where search is involved, excessive time spent might lead to frustration. In these circumstances Internet-based 'virtual' searches may be less time consuming than real world 'bricks and mortar' searches. In these kinds of cases it may be $\tilde{L}_O - L_I$ that generates utility where $\tilde{L}_O$ denotes the time that would need to be spent on a bricks and mortar search as a substitute for a virtual search for a given good or service. It may be necessary to distinguish two categories of internet products – those for which time spent in use generates utility and those for which time spent in use generates disutility.

*Is the wage the best driving variable for budget allocation?*

Although $W$ is referred to as the wage in the G-K description of the model, in an atemporal allocational model the driving variable is in fact more akin to total expenditure in a given period. This assumes a prior consumption/savings choice has been made. The money available to allocate to consumption may not be well proxied by the wage.

*The G-K model does not have an explicit time budget.*

The G-K model implicitly assumes that time spent is rather like money spent. The time allocations $L_I$ and $L_O$ are fractions of the time period that are spent in consuming the alternative products. But people who are money rich may be time poor so it is possible than blending the decision problem into one overall constraint blurs this possibly important distinction. Wealthy people who consume more luxuries relative to necessities in the money space may in fact consume more necessities relative to luxuries in the time space. This could have interesting consequences which may be worth exploring.

*Aggregation*

The high level of aggregation (one internet product and one other product), while helpful for model exposition, is not ideal for applied work in this context.

*Are there alternative opportunity costs?*

Consideration of a greater range of products also raises the issue of whether a different 'wage' or opportunity cost may be relevant as the price of time spent consuming different types of internet products (and for that matter, different types of bricks and mortar products). One intriguing possibility would be to consider finding the next best option to any particular item consumed and then to use the marginal utility of the next best option as the opportunity cost for time spent on the alternative product.

### 1.3. Optimal Share Equations for Money and Time Expenditure

Direct optimisation of the primal problem (by tedious 'brute force' - by setting up and solving the Lagrangean) would lead to the following four equations (written here in budget share form and using the G-K notation):

$$\frac{P_I C_I}{W - F_I} = \alpha_I \frac{1}{1+\Gamma} \tag{5}$$

$$\frac{WL_I}{W - F_I} = (1-\alpha_I)\frac{1}{1+\Gamma} \tag{6}$$

$$\frac{P_O C_O}{W - F_I} = \alpha_O \frac{\Gamma}{1+\Gamma} \tag{7}$$

$$\frac{WL_O}{W - F_I} = (1-\alpha_O)\frac{\Gamma}{1+\Gamma} \tag{8}$$

where

$$\Gamma = \left(\frac{\lambda_I}{\lambda_O}\right)^{\sigma-1}\left(\frac{1-\theta}{\theta}\right)^{\sigma} \tag{9}$$

and $\lambda_I$ and $\lambda_O$ are Cob-Douglas price sub-indexes:

$$\lambda_j = \left(\frac{P_j}{\alpha_j}\right)^{\alpha_j}\left(\frac{W}{1-\alpha_j}\right)^{1-\alpha_j}, \quad j = I, O \tag{10}$$

*Assumptions Implicit in the Derivation of the G-K Estimating Form*

For reasons that are not made explicit but which appear to relate to data availability, G-K do not estimate the optimal share equations (5)-(8) as a system but instead combine the share equations into a single equation for estimation. It is useful to describe this process because depending upon data availability, there may be other more efficient options, retaining a system of equations for estimation if data permits. Some reduction in the number of equations may be necessary if data does not permit full estimation, and various assumptions may need to be made to give effect to a reduced set of estimating equations.

To illustrate the potentialities and the problems, it is useful to write the budget constraint (4) in share form:

$$\frac{P_I C_I}{W-F_I} + \frac{WL_I}{W-F_I} + \frac{P_O C_O}{W-F_I} + \frac{WL_O}{W-F_I} = 1 \tag{11}$$

from which it is obvious that one of the share equations is redundant. Any three of the four share equations (5)-(8) can be estimated. What is surprising is that G-K sacrifice efficiency by ignoring the fact that they could estimate three equations with cross equation restrictions to get a more efficient estimate of $\sigma$. Their approach effectively is to combine (5), (7) and (8) into one equation:

$$\begin{aligned} \frac{P_I C_I}{W-F_I} + \frac{P_O C_O}{W-F_I} + \frac{WL_O}{W-F_I} &= \alpha_I \frac{1}{1+\Gamma} + \alpha_O \frac{\Gamma}{1+\Gamma} + (1-\alpha_O)\frac{\Gamma}{1+\Gamma} \\ &= \frac{\alpha_I + \Gamma}{1+\Gamma} \end{aligned} \tag{12}$$

which with the use of (11) can be written as:

$$1-\frac{WL_I}{W-F_I}=\frac{\alpha_I+\Gamma}{1+\Gamma} \tag{13}$$

This reduction process may be necessary if data on $C_I$ and $C_O$ is unavailable to allow estimation of the full system of share equations, but otherwise it would not be recommended. If it needs to be used, it reduces the four share equations (5)-(8) to just two equations:

$$\frac{L_I}{1-F_I/W}=(1-\alpha_I)\frac{1}{1+\Gamma} \tag{14}$$

$$1-\frac{L_I}{1-F_I/W}=\frac{\alpha_I+\Gamma}{1+\Gamma} \tag{15}$$

where (14) is a minor rearrangement of (6), while (15) is the replacement for the combination of (5), (7) and (8). In view of the budget constraint, these two equations are not independent. Either one of these equations could be estimated although as pointed out above, if data on $C_I$ and $C_O$ were available it would have been more efficient to estimate three of the four original share equations.

However, instead of estimating any one of (14) and (15), G-K make a further approximation: they assume that $F_I$ is small relative to $W$ and then ignore the term $F_I/W$. This assumption is not innocuous. Elsewhere, G-K make the point that they are going to make use of variation in $W$ to obtain evidence of the size of $\sigma$. They claim that this is possible because with their sample there is considerable variation in $W$. If this is correct then by ignoring the term $F_I/W$ they are now ignoring an important source of variation which could have aided in the estimation of $\sigma$. If the data set is one in which $W$ varies substantially, as G-K suggest is the case, then (given that $F_I$ is relatively uniform across the sample) $F_I/W$ will also vary substantially, even though it is small.

With the assumption that $F_I$ is small relative to $W$, (14) and (15) are further reduced to:

$$L_I=(1-\alpha_I)\frac{1}{1+\Gamma} \tag{16}$$

$$1-L_I=\frac{\alpha_I+\Gamma}{1+\Gamma} \tag{17}$$

Then taking the ratio of (17) to (16), G-K obtain

$$\frac{1-L_I}{L_I}=\frac{\alpha_I+\Gamma}{1-\alpha_I} \tag{18}$$

This equation could be estimated by nonlinear methods though G-K choose to make further approximations to estimate it in a log-linear regression. In the process they treat the prices $P_I$ and $P_O$ as constants, something that may have been necessary with their dataset, where it seems data on $P_I$ and $P_O$ is not available, but which is also limiting in its plausibility.

An approach to nonlinear estimation of (18) would use (9) and (10) to eliminate $\Gamma$. In an explicit series of steps, this gives:

$$\begin{aligned}\frac{1-L_I}{L_I}&=\frac{\alpha_I}{1-\alpha_I}+\frac{1}{1-\alpha_I}\Gamma\\&=\frac{\alpha_I}{1-\alpha_I}+\frac{1}{1-\alpha_I}\left(\frac{\lambda_I}{\lambda_O}\right)^{\sigma-1}\left(\frac{1-\theta}{\theta}\right)^{\sigma}\\&=\frac{\alpha_I}{1-\alpha_I}+\frac{1}{1-\alpha_I}\left(\frac{\alpha_O^{\alpha_O}\left(1-\alpha_O\right)^{1-\alpha_O}}{\alpha_I^{\alpha_I}\left(1-\alpha_I\right)^{1-\alpha_I}}\right)^{\sigma-1}\left(\frac{\left(P_I/W\right)^{\alpha_I}}{\left(P_O/W\right)^{\alpha_O}}\right)^{\sigma-1}\left(\frac{1-\theta}{\theta}\right)^{\sigma}\end{aligned} \tag{19}$$

G-K make a further simplification, by assuming that $\alpha_I=0$, which they then use selectively to eliminate some terms involving $\alpha_I$. This leads to replacement of (19) by the approximation:

$$\frac{1-L_I}{L_I}\approx\left(\frac{\alpha_O^{\alpha_O}\left(1-\alpha_O\right)^{1-\alpha_O}}{\alpha_I^{\alpha_I}\left(1-\alpha_I\right)^{1-\alpha_I}}\right)^{\sigma-1}\left(\frac{\left(P_I/W\right)^{\alpha_I}}{\left(P_O/W\right)^{\alpha_O}}\right)^{\sigma-1}\left(\frac{1-\theta}{\theta}\right)^{\sigma} \tag{20}$$

G-K write this in an alternative form by shifting $P_I$ and $P_O$ into the first term on RHS (20). Thus the G-K version of (20) is:

$$\frac{1-L_I}{L_I}\approx\left(\frac{P_I^{\alpha_I}\alpha_O^{\alpha_O}\left(1-\alpha_O\right)^{1-\alpha_O}}{P_O^{\alpha_O}\alpha_I^{\alpha_I}\left(1-\alpha_I\right)^{1-\alpha_I}}\right)^{\sigma-1}\left(\frac{W^{\alpha_O}}{W^{\alpha_I}}\right)^{\sigma-1}\left(\frac{1-\theta}{\theta}\right)^{\sigma} \tag{21}$$

In log form (with the obvious collection of terms) this is equivalent to their equation (4):

$$\ln\left(\frac{1-L_I}{L_I}\right) \approx \ln A + (\alpha_O - \alpha_I)(\sigma - 1)\ln W + \sigma \ln\left(\frac{1-\theta}{\theta}\right) \tag{22}$$

where $A = \left(\frac{P_I^{\alpha_I}\alpha_O^{\alpha_O}\left(1-\alpha_O\right)^{1-\alpha_O}}{P_O^{\alpha_O}\alpha_I^{\alpha_I}\left(1-\alpha_I\right)^{1-\alpha_I}}\right)^{\sigma-1}$ . G-K further treat the term $\sigma \ln\left(\frac{1-\theta}{\theta}\right)$ as a normally distributed error term, treat $A$ as a constant, and propose to use (22) to obtain an estimate of $\sigma$ .

### 1.5. Further Assumptions for Linear Estimation of the G-K Estimating Form

The G-K regression results for equation (22) treat $A$ as a constant even though it is a function of $P_I$ and $P_O$. G-K also assert that $\alpha_O - \alpha_I$ can be treated as approximately -0.62, though it is not clear that this is a robust statistic and their use of it is hard to reconcile with their assumption elsewhere that $\alpha_I = 0$. This is so because the assumption that $\alpha_I = 0$ together with the assertion that $\alpha_0 - \alpha_I \approx -0.62$ implies that $\alpha_O$ is negative, but this is inconsistent with its use as a parameter in the utility function, where it appears as the power of composite (non-internet) consumption, and would typically be expected to be positive if marginal utility is positive. In fact G-K state (p. 110, para. 1) that they will illustrate that $\alpha_0 - \alpha_I \approx -0.62$, but this illustration may have been eliminated from an earlier version of their work, since it has not been possible to find any such illustration in the paper.

There are several other issues with the G-K research results that probably should also be noted. One puzzling issue is that in their final presentation of results they appear not to use the theoretical model they have developed but instead opt for 'linear demand' and 'log demand' specifications. The nested CD-CES model they had developed is not totally consistent with either of these, although it could be thought of as (approximately) a log demand type of model. However, in their footnote 7 on p. 112, G-K state that one should be wary of the log-demand model, and they seem to prefer the linear demand results which are really not at all consistent with their model specification.

The two final caveats that G-K offer are also important to keep in mind. The first of these relates to the use of the wage as the opportunity cost of time spent on the Internet. A possible alternative could be use of a 'price' based on the marginal utility of the next most attractive activity. Their second caveat concerns their use of a composite 'other' product for non-internet consumption. One might add that internet time is also a composite of alternative consumption items. Both internet and non-internet products could be (ideally) disaggregated and this could allow for interesting patterns of complementarity and substitutability to be explored.

In view of the fact that the IPTS clickstream database has a much greater degree of disaggregation than the G-K model while at the same time it does not contain information on time spent consuming non-internet items, and in view also of the variety of issues raised above, it has to be concluded that the G-K model offers some useful lessons but only limited overall guidance at this stage.

## 2. Proposal for a purpose-built model of Internet search time

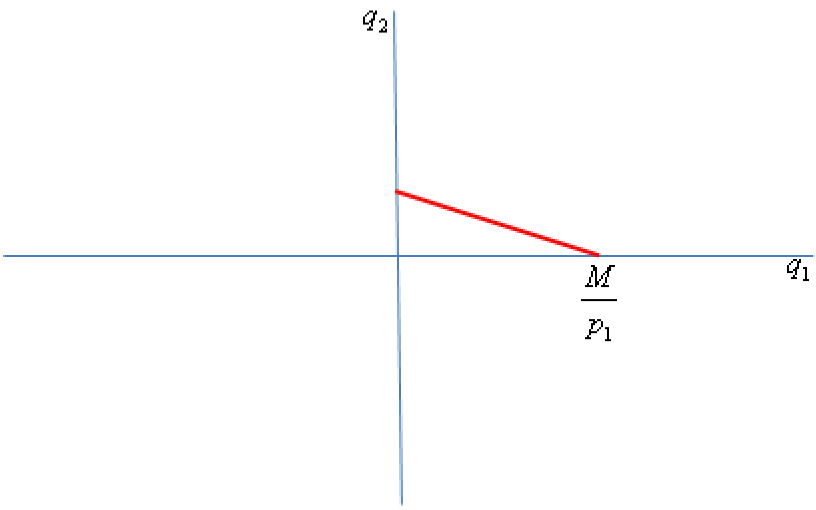
The purpose of this section is to introduce a suggestion for a model which concentrates on consumer Internet search time aimed to produce information for general consumer choices.

A fully integrated model of consumer Internet time choice together with traditional product choice would be complicated by the need to recognize that some Internet time use could consist of search time, which could be more productive than traditional non-Internet search time, but which ultimately still leads to a non-Internet purchase, while other time spent on the Internet could be a direct substitute for a traditional purchase. In general, this would be a very complex modelling task that is beyond the scope of this report. The following model proposal provides suggestions for a simple model that treats all Internet time as search time that is used to obtain the necessary information for later traditional purchases. This also leads to concentration on a model that is most compatible with the IPTS clickstream database. It might be thought that some information in the clickstream database relates more to time spent on consumption rather than search. This could be an interpretation of the 'Entertainment' category, for example, at least where it applies to time spent on web sites related to music, videos and so. However, there is some evidence that time spent on this category is still search time for items that are ultimately downloaded. The subsequent time spent offline consuming the entertainment is not recorded.

## 2.1. Concepts of relevance to the modelling

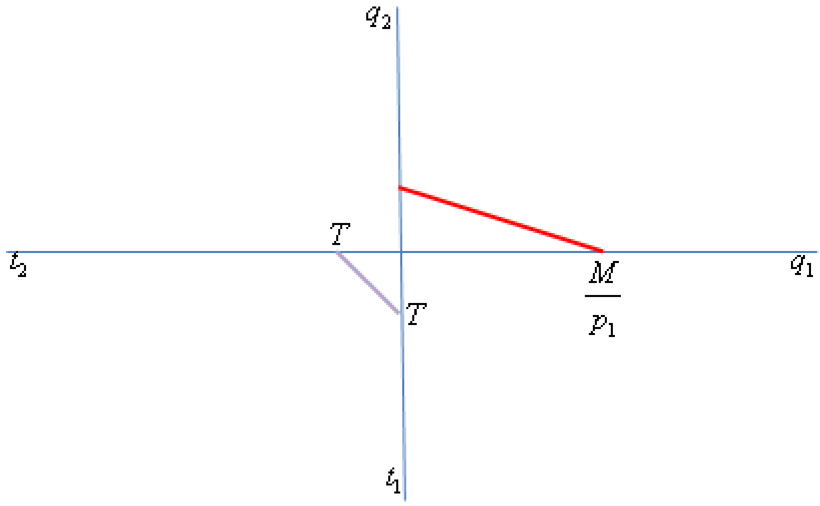
The purpose of this diagrammatic section is to highlight modelling concepts and related issues that will need to be addressed more fully in an empirically oriented computable model. Figure 1 introduces the notation for a consumption allocation problem for a consumer with available total expenditure $M$ and a choice of two commodities at prices $p_1$ and $p_2$.

Figure 1: Budget Constraint – traditional formulation




Next, in Figure 2, the Internet search time allocation problem is introduced by a simple modification to the traditional diagram. The modification makes use of the lower left quadrant to model search time, which is assumed to ultimately produce information relevant for solving the traditional consumer allocation problem which will continue to be exhibited in the upper right quadrant of the diagram. This stylised representation of the model is designed to illustrate a simple situation in which an amount of time $t_1$ can be allocated for searching for information to aid in the choice of $q_1$, while time $t_2$ can aid in the choice of $q_2$. The total amount of time available is $T$ and because there is a one-to-one time trade-off in the allocation of time in search of information on either product, the time budget line has a slope of 45 degrees.

Figure 2: Time constraint for search – allocation problem

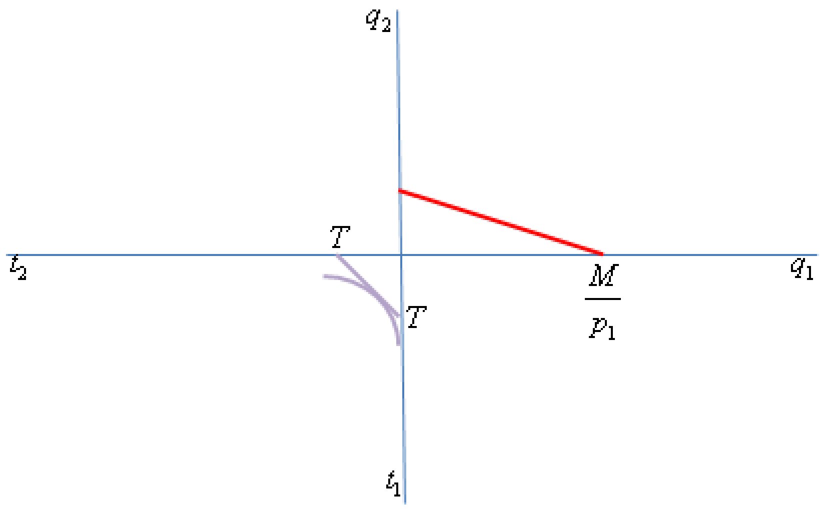



The production of information is considered in this approach to be a non-market activity, once access to the Internet is achieved. In principle, the total time allocation $T$ is something that the consumer must choose. At this stage it is useful to invoke a separability assumption that treats the time allocation problem in a two stage decision making manner, analogously to the separation of the total expenditure and commodity share decisions that is the norm in traditional consumer demand modelling. Given that $T$ is decided at stage 1, the concern now is to model the time split at the second stage.

Hence, thinking of the lower left quadrant in non-market production terms (viz. production of information for own use by the consumer), with factor inputs $t_1$ and $t_2$, given that the consumer has available a total time budget for search, $T$, the optimisation problem can viewed as needing to choose the highest possible production isoquant that can be reached with time budget allocation $T$.

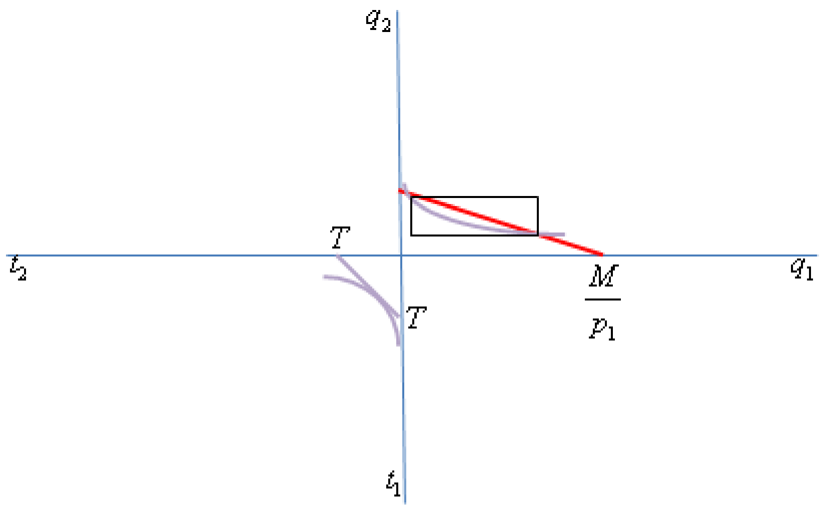
Figure 3 illustrates the outcome of this optimisation by drawing an information production isoquant in the lower left quadrant which is just tangential to the time budget line.

Figure 3: Time allocation for given T determines best information isoquant




Now Figure 4 shows a mapping between a production information isoquant and the consumer's digestion of this information in understanding the trade-off between products that is possible while retaining a given level of utility.

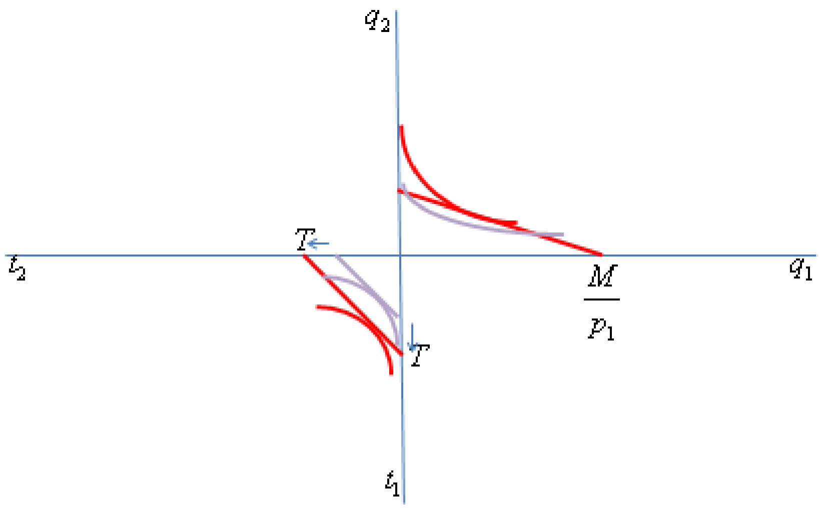
Figure 4: Information isoquant reveals pattern of indifference curves
– narrows range to search for optimal choice




The situation shown in Figure 4 is one in which the information production isoquant in the lower left quadrant reveals an indifference curve which is drawn in the upper right quadrant. That is, the information that the consumer acquires includes relevant assessments of the quality of the products sufficient to have an appreciation of their relative merits in generating consumer utility. However, the consumer's optimal choice of search times, and the highest information production isoquant that can be reached, is dictated by the prior assignment of a total amount of search time (the stage 1 decision taken as given here) and this is not necessarily sufficient to reveal the optimal consumer choice. That is, the given isoquant reveals an indifference curve that could be achieved, but it does not reveal full information about all opportunities. Hence it does not reveal the optimal indifference curve for the given monetary budget $M$ and relative prices $p_1$ and $p_2$. Nevertheless, with the given amount of information, the range of suitable choices is narrowed. The rectangular box in Figure 4 designates the 'grey area' which the consumer knows contains an optimal choice. However, the optimal choice is not known with precision.

Hence, as Figure 5 suggests, a greater amount of time spent in search may reveal more, and ultimately will give information suitable for making an optimal choice for a given budget.

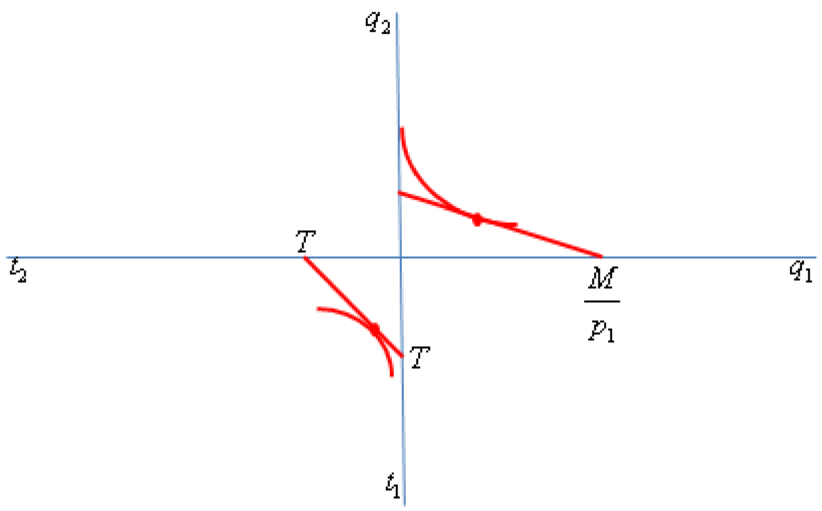
Figure 5: Greater search time – higher information isoquant – reveals higher indifference curve




This suggests that the total search time $T$ is endogenous, being determined by the amount of time required to reach the isoquant that ultimately reveals the indifference curve that is just tangential to the monetary budget constraint. Ultimately, it will be desirable to model this 'first stage' decision. However, analogously to the case in consumer theory, this will be more complex than the allocational decision. In traditional consumer theory the total expenditure choice decision is an intertemporal one, involving a consumption/savings choice, while the allocation of total expenditure among various categories of commodities is treated as an atemporal choice. This is then modelled by a simpler optimisation problem. The same applies here. It seems wise to attempt the simpler problem first, by invoking separability in the decision making process, and leaving the issue of endogenising $T$ to later research.

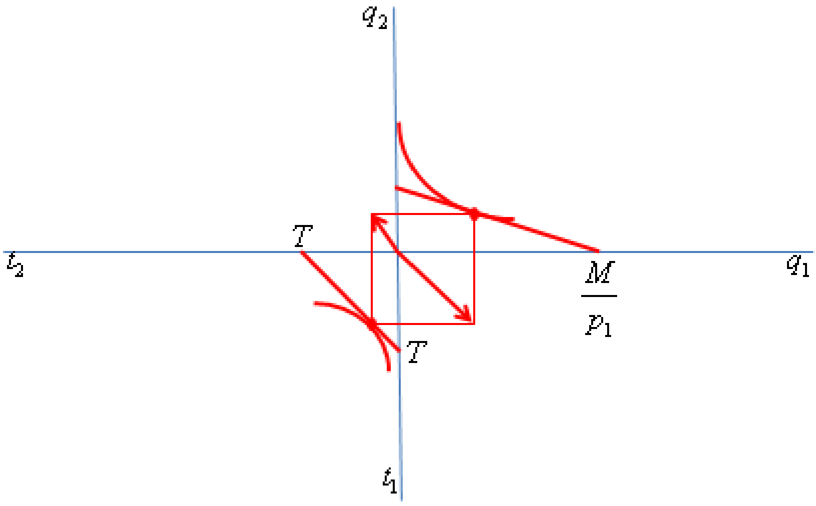
In order to concentrate on the search time allocation problem at this point, the total time $T$ is assumed at this point to be sufficient to generate an optimal choice for a given budget, as shown in Figure 6.

Figure 6: Optimal amount and allocation of search time
- leads to optimal consumption choice



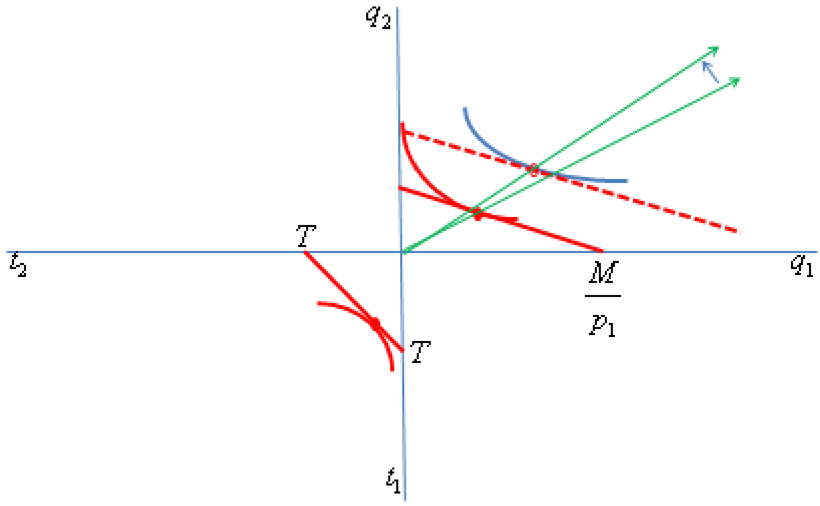

At the optimum, it is possible to determine the relationship between the time allocation required and the resulting optimal consumption choice that can be made. This can be thought of as the shadow value of search time – measured in terms of the consumption available as a result of the allocation of search time. These shadow values are depicted in Figure 7.

Figure 7: Shadow value of search time
- defined as optimal consumption per unit of search time




This definition of shadow value is sensitive to the units of measurement. The important point is not the difference between the shadow values (so measured) for the two search times but rather their role in describing how the circumstances change. Figure 7 has depicted the situation for the initial optimal choice. To illustrate the effect of a change in circumstances it is useful to consider a situation in which one traditional good is a necessity while the other is a luxury. Figure 8 depicts the case where the expansion path is biased towards increasing the proportion of good 2 consumed relative to good 1.

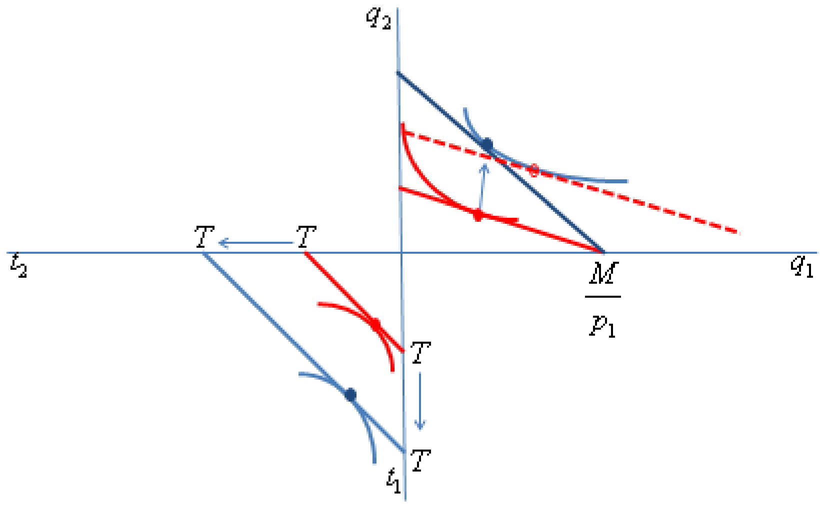
Figure 8: Expansion path of goods
– suggests good 1 is a necessity and good 2 is a luxury




The indifference curves shown in Figure 8 reflect non-homotheticity of preferences. For a given ratio of relative prices, as $M$ increases the point of tangency is not along a single ray from the origin. In the Figure, the point of tangency shifts upwards, so that at a higher income level the proportions spent on good 1 and good 2 turn more in favour of good 2. In traditional terms, good 2 can be described as a luxury and good 1 as a necessity. This description takes it as given that total money expenditure $M$ is more or less synonymous with income. In fact just about all theories of consumption generate this type of close relationship so it is taken for granted here. Later, a more careful look at optimal choice of $M$ may generate guidance in developing a theory of optimal choice of $T$, but this is relegated as a task for further research.

The point now is to use the analysis, complete with a depiction of a non-homothetic preferences situation, to consider the implications of a change in relative prices. An important empirically relevant case to examine is that in which it is the price of the luxury that falls relative to the price of the necessity. It seems reasonable to assert (although empirically this is still a matter for thorough investigation) that the types of innovations that have occurred in ICT have had the following consequence: first, ICT has acted as a general purpose technology that is present as an intermediate good in the production of most if not all products; second, there tends to be more ICT embedded in more sophisticated products; third, luxuries tend to be more sophisticated products than necessities. As a result it is arguable that innovations in ICT, which have led to falls in quality adjusted prices, have had a bigger effect on cost reductions in the production of luxuries. In Figure 9, this is depicted as a change in relative prices via a fall in $p_2$, the price of the luxury good, or at least a lesser price rise in $p_2$ than that which occurs in $p_1$.

Figure 9: Fall in price of good 2 – a luxury
– leads to further search to determine new optimum

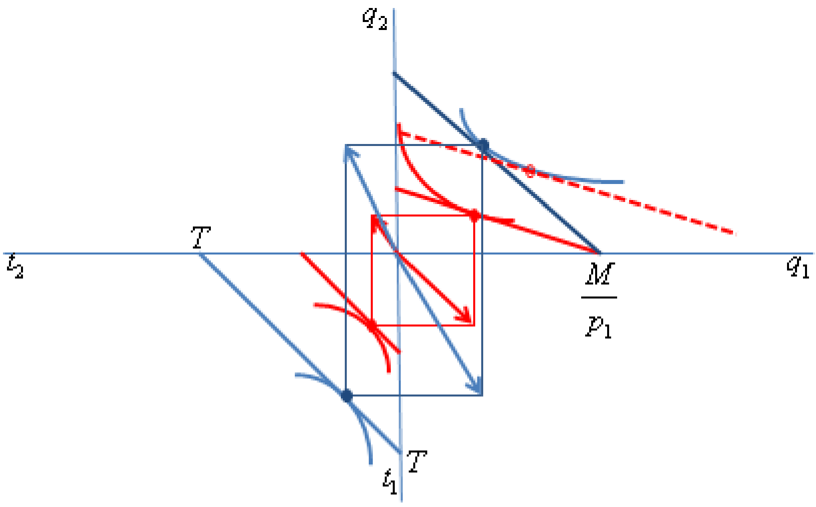



The relative price change is drawn in Figure 9 as if the budget line swings around the good 1 axis intercept. This depicts a situation in which the consumer has his/her endowment stored in units of the necessity. This simple device makes the analysis clearer than general alternatives, but does not change the general results. As depicted, the relative price change leads to an increase in real income (technically, in real total expenditure possibilities). Also as shown, this leads to the opening up of a range of possible consumption choices not previously envisaged. This suggests that finding the new optimum requires an increase in total search time and may also require a change in the proportional allocation of search times.

Figure 9 shows the outcome as the combination of an income effect and a substitution effect. This is simply meant to illustrate that the new optimum, being associated with a higher level of real income, will exhibit a greater proportional consumption of the luxury.

Now recalling the shadow values depicted in Figure 7, one could ask what happens to them. The situation is illustrated in Figure 10 below, where it is shown that the change has led to a reasonably large decrease in the shadow value of good 1 and possibly no change or a best a slight increase in the shadow value of good 2. Relatively, however, there are now more 'bangs for the buck' in searching for good 2.

Figure 10: New search time allocation and consumption levels
- determine revised shadow values

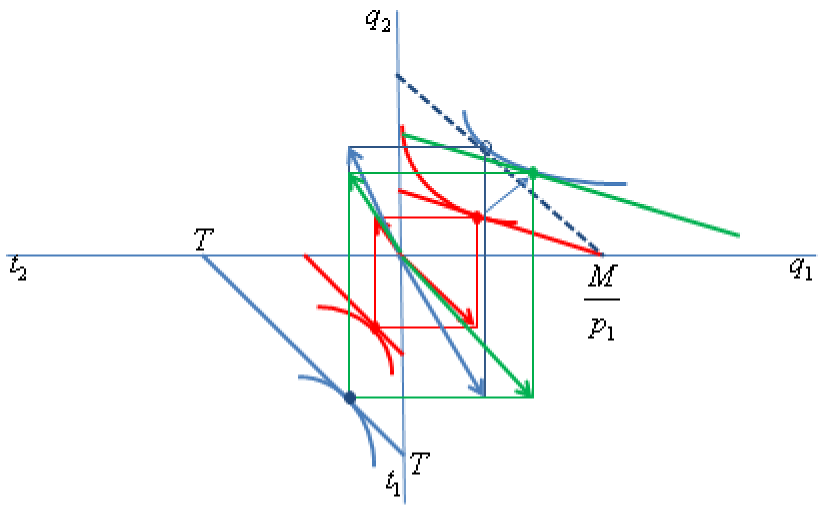



As is probably evident, the direction of change in the shadow values is actually related to the degree of non-homotheticity in both the production technology and the consumption preferences. Figure 10 depicts a situation where the information production time input technology is relatively homothetic, although biased towards factor $t_1$. That is, the expansion path of tangencies in the lower left quadrant is drawn as extending along a ray from the origin. This is simply for ease of presentation. On the other hand the consumption preferences are non-homothetic, with good 1 being a necessity and good 2 being a luxury as drawn (refer to Figure 8 for the demonstration of this).

It is worth noting that, given the definition of shadow values used here – defined in an average sense as the ratio of consumption of a product to the amount of time required to have revealed this consumption as optimal – then if an integrated database of both search times and consumption choices can be developed, these shadow values can be calculated.

It is useful to compare the situation depicted in Figure 10 with that of an equivalent increase in welfare achieved simply by an increase in income with no relative price change. This is illustrated in Figure 11. Both situations show an information production isoquant sufficient to achieve a given level of utility. However, the shadow values of the search times are quite different in the two situations.

Figure 11: An income increase with an equivalent welfare effect - generates a different pattern of shadow values




As Figure 11 shows of course, the consumption points are also quite different even though they generate the same utility.

Now it is useful to strip out the clutter from Figures 10 and 11 in order to uncover the underlying 'information production' technology. This is done in Figure 12. This reveals an interesting modelling possibility. Suppose actual consumption prices are not available but search times and consumption levels are. Then actual consumption choices could be explained by a search time model instead of by a traditional consumer demand model.

Figure 12: Relationship between 'production technology' (information) and consumption opportunities

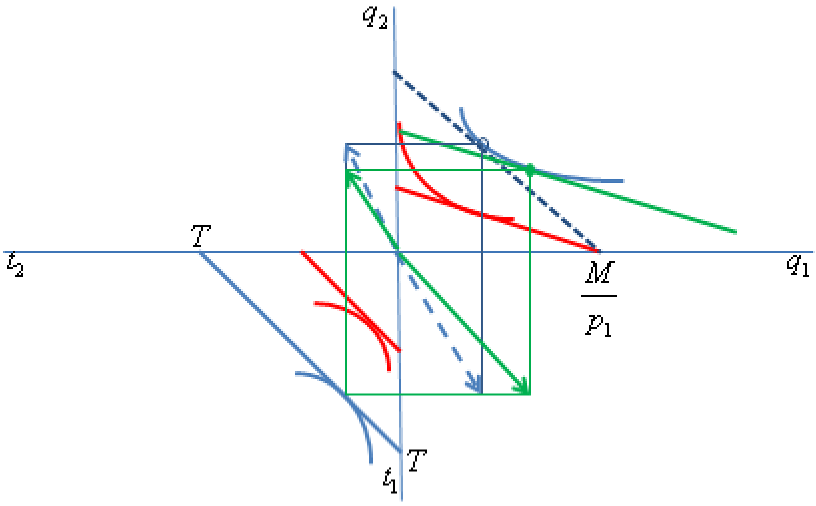
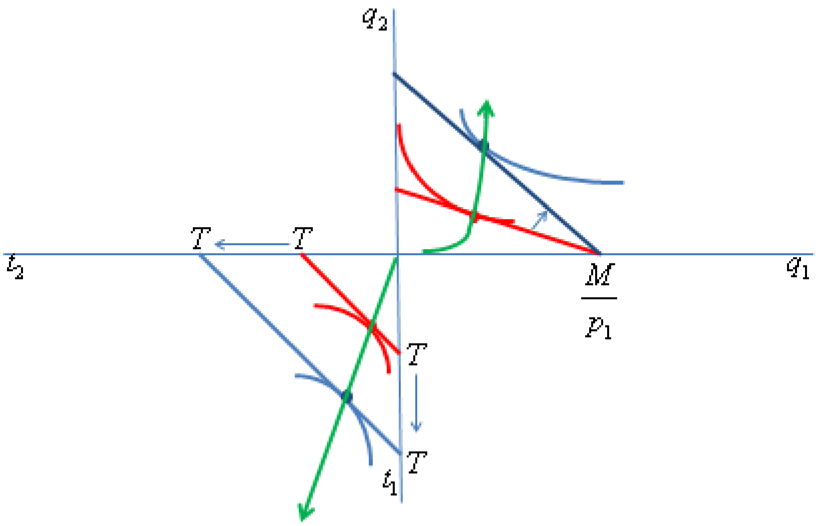



Figure 13 summarises what might be considered the most empirically relevant case. This is essentially a repeat of Figure 9 but with the actual consumption paths shown if this trend in relative price changes were to continue. However, in developing an algebraic model from these considerations it will be necessary to recognize the potentially very complex relationship between Internet search time data, and consumer choice.

Figure 13: Possible paths of search time and consumption shown for initial low consumption of luxury good 2 but with good 2 price falling

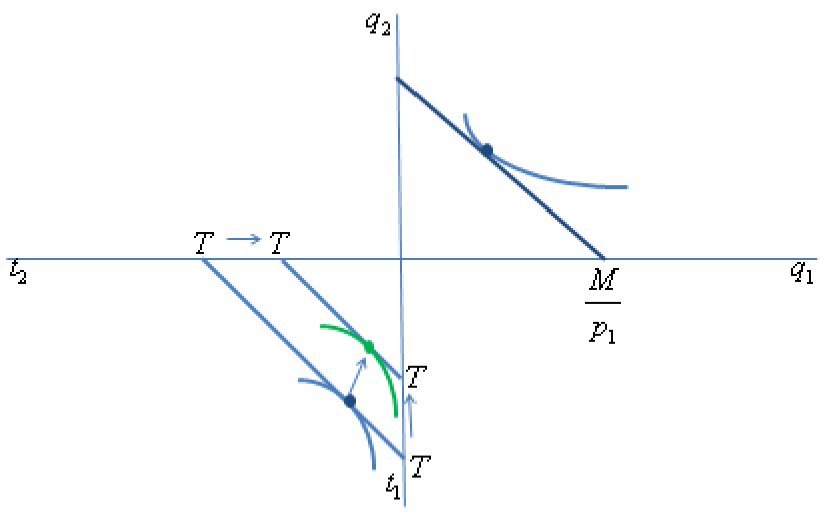


Figures 11 and 12 have demonstrated that the very same time allocation proportions may be associated with different consumption choices even if they yield the same overall utility. Of course, actual relative prices help to distinguish these situations in the traditional consumption case. For household consumer Internet search costs, however, market prices are not available. Nevertheless, it ought to be possible to use shadow values to help discern different outcomes. However, to generate these, matching data from traditional consumption choices would appear to be needed.

An added element of complexity is involved in attempting to endogenise the total time allocation $T$. Figure 13 depicts a situation in which $T$ increases in response to the consumer being in a position to increase their utility. However, this seemingly monotonic relationship between utility and time spent on search will not hold up under all circumstances. Figures 14 and 15 depict two alternative situations which may be empirically relevant from time to time.

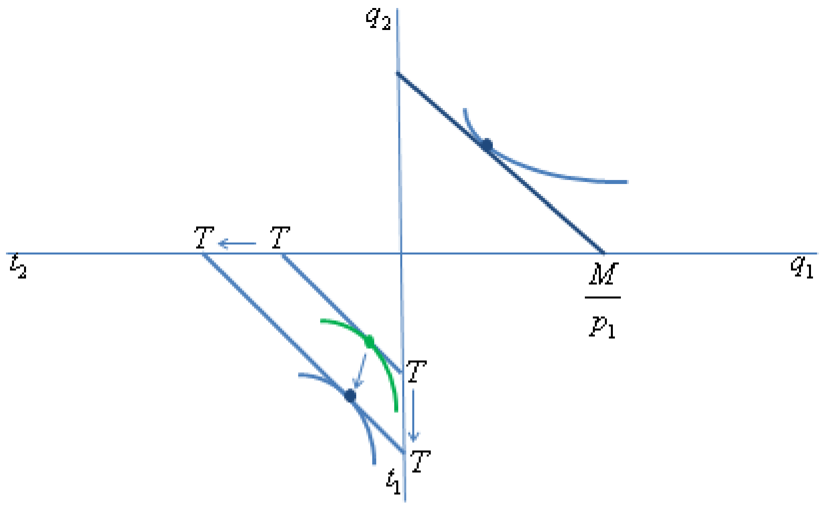
The production of information may be more analogous to the production of a durable good rather than a perishable good. Unless the technology changes quickly, once the information is available it may be useful for some time to come. In the near future, then, there may be less need for search once good information is uncovered. Figure 14 depicts this case.

Figure 14: If there are no major price changes, income advances or innovations in product quality, search times might gradually shrink



On the other hand, with an innovation in a general purpose technology, it may be that the quality of all goods rises more or less evenly. While search time may need to be spent to discover this, the outcome may be that the relevant indifference curve has not changed.

Figure 15: A major increase in search time with no obvious change in consumption patterns may indicate a quality change in the underlying general purpose technology




The important thing is that $T$ changes more rapidly than $M$ in the situations pictured and, moreover, it changes in opposite directions in the two situations. It is clear that the relationship between $M$ and $T$ is an inherently complex one, and endogenising $T$ will require very careful modelling indeed.

## 2.2. Proposed model structure

*A suggested objective function: maximizing an information production function*

The graphical analysis introduced in Figure 3 suggests that an index of the production of information (in effect, a size label for the production isoquant shown in the lower left quadrant) is some function of the two search times $t_1$ and $t_2$. Generalising to a search over $N$ products, a candidate functional form for information production would be:

$$\sum_{i=1}^{N} \gamma_i \ln t_i \ , \ \sum_{i=1}^{N} \gamma_i = 1 \qquad (23)$$

This function is so simple that it is likely to be empirically implausible. However, it will be useful to use it as a starting point for development of a more realistic alternative.

*The 'time budget'*

Now consider the non-market situation where the consumer does not have to pay for additional time spent in searching. There may or may not be an initial fixed cost in obtaining access to a search engine – this will not affect the analysis to follow. At this point it is useful to assume that the total time available for search, $T$ , is treated as exogenous when making the individual product search decisions. It may be regarded as having been chosen by the consumer in a separable two stage decision making process. The consumer's time budget constraint is simply:

$$\sum_{i=1}^{N} t_i = T \tag{24}$$

*Solution in the simple logarithmic case*

Now maximization of (23) subject to (24) (say by a Lagrangean, or primal, technique) gives the optimal choices:

$$t_i = \gamma_i T \tag{25}$$

Or, defining time shares $\tau_i \equiv t_i / T$, one simply obtains $\tau_i = \gamma_i$. This is too simple to be empirically plausible because it implies that no matter how much time is available one will always allocate that time in fixed proportions to the search for information on various products. While the plausibility of this remains to be examined in the case of Internet search times, it seems wise to develop a more general model that does not have this restrictive feature but that contains this as a special case (in the unlikely event that it is found to be reasonable).

*A slight generalisation: entropy adjustment*

A further disadvantage of the candidate functional form (23), quite apart from its likely empirical implausibility, is that it cannot be evaluated for zero search time and that it does not return an index of 'production' that is non-negative. An improvement to rectify the latter defects and also possibly add slightly to empirical plausibility is:

$$\sum_{i=1}^{N} \gamma_i \ln(t_i + \gamma_i) - \sum_{i=1}^{N} \gamma_i \ln \gamma_i \tag{26}$$

The functional form (26) returns zero for $t_i = 0$, $i = 1,...,N$ and a positive value in the presence of any positive $t_i$. The additive term

$$-\sum_{i=1}^{N}\gamma_i \ln \gamma_i = \sum_{i=1}^{N}\gamma_i \ln(1/\gamma_i) \tag{27}$$

does not affect optimal choice of search times but plays an important balancing role in generating the actual measure of information produced. Despite its appearance it is actually a positive term and can be interpreted as 'entropy'. Suppose the search for information on all products is equally productive. (In the two dimensional diagrammatic example, this would correspond to the expansion path for information, the lower arrow in Figure 13, expanding downwards at a 45 degree angle from the origin). In the general case this means $\gamma_i = 1/N$ for all $i$ and the entropy measure (27) is just $\ln N$. In this case it is clear that the entropy increases (at a decreasing rate) with the number of products. This result on entropy generalises to the case where the $\gamma_i$ parameters are not all equal and hence where search for some products is more productive than search for others. It is still the case that having the ability to search for information over a greater number of products adds to the stock of productive information.

*Solution of the simple logarithmic form in the entropy-adjusted case*

Finding the optimal choice of the $t_i$ to maximize (26) subject to the budget constraint (24) is also achievable by the primal (Lagrangean) method. In fact, the solution is:

$$t_i = \gamma_i\left(T+1\right) - \gamma_i = \gamma_i T \tag{28}$$

exactly as for the entropy-neglected case. The equivalence is in fact due to the extreme simplicity of the time budget constraint. The reason for exhibiting this, however, is to lead into another approach based on a transformation of variables.

*The entropy-adjusted logarithmic approach with transformed variables*

Define the artificial variables $\tilde{t}_i = t_i + \gamma_i$. Then the information production function (26) can be recast as:

$$\sum_{i=1}^{N}\gamma_i \ln \tilde{t}_i - \sum_{i=1}^{N}\gamma_i \ln \gamma_i \tag{29}$$

In terms of the artificial $\tilde{t}_i$ variables, the time budget constraint (24) can be recast with the aid of the adding-up identity on RHS (23) as:

$$\sum_{i=1}^{N} \tilde{t}_i = T + 1 \tag{30}$$

Since the entropy term on RHS (29) does not affect the optimisation, the problem in transformed variables, (29)-(30), is technically equivalent to (23)-(25) with the exception that the resource constraint is now $T+1$ rather than $T$. The solution is therefore structurally equivalent to (25), but in terms of $\tilde{t}_i$ and $T+1$, viz.:

$$\tilde{t}_i = \gamma_i \left(T+1\right) \tag{31}$$

Of course, this is just another way of writing the solution to the original simple problem. However, it suggests a transformation of variables that might be capable of further generalisation.

*The use of the concept of the indirect production function*

Unfortunately, plausible generalisations of the function form (26) – or, in transformed variables, (29) – do not lead to easy solution by primal techniques. Fortunately, resort to duality theory suggests a way forward. To this end, it is first of all useful to derive the indirect production function (IPF) associated with the problem as specified above. Here (26) can be interpreted as a direct production function (DPF). Using the transformed variables version (29) for simplicity, substitution of the solution (31) into the DPF (29) gives the IPF as:

$$IPF = \sum_{i=1}^{N} \gamma_i \ln\left(\gamma_i (T+1)\right) - \sum_{i=1}^{N} \gamma_i \ln \gamma_i = \ln\left(T+1\right) \tag{32}$$

The approach in what follows is to generalise the IPF (32) using results from duality theory to ultimately obtain a more empirically plausible functional form.

*A ‘change of variables’ approach: an artificial problem with a generalisable solution*

The IPF (32) is particularly simple, possibly too simple. Additionally, a change of variables is required to set it up in a form suitable for using an envelope theorem from duality theory, such as Roy’s Identity, which is widely used in the analogous theory applied to traditional consumer choices.

Define:

$$V = W_A T \tag{33}$$

where $W_A$ is the consumer's opportunity cost of time.

Although $W_A$ might be approximated empirically by the wage rate $W$ (where $W$ is the wage per hour if $T$ is measured in hours), this is extremely rough and ready and unlikely to lead to sensible empirical results. To see this, observe that the true opportunity cost of time should equal the money equivalent of the utility forgone by not pushing out the time budget line and hence not obtaining the best possible information on characteristics of products that would give an accurate map for consumer preferences (cf Figure 5 above). Especially in the case where a consumer could have put in an extra hour of work for no pay (say, because she/he is salaried), the current wage rate is very unlikely to reflect this opportunity forgone. What would better reflect the value of unpaid extra time worked? Suppose the consumer has decided to work longer in the hope of being promoted. The consumer is then valuing the extra hour of work in terms of its contribution to future income, and a consideration of the difference in the present value of the two future income streams – that with promotion and that without – suitably weighted by the probability of the extra hour's work leading to promotion, would give a more appropriate monetary measure to compare with (the money equivalent of) the current utility foregone. Clearly however, this involves a substantial degree of modelling complexity, and so it is left as a recommended project for future research. Here we turn to another possible approach for dealing with $W_A$.

Now $T = \sum_{i=1}^{N} t_i$, which may be interpreted as a real time aggregator function, so another way to think of $W_A$ is that it is also an aggregator function, in this case a shadow price aggregator. Define shadow prices $\pi_i$ representing the opportunity cost of time spent on search for each product. These shadow prices need not be equal. All that is required at this point for a reasonable definition of the shadow price aggregator function is that it be positive, increasing in prices, homogeneous of degree 1 and satisfy the obvious regularity property that if $W_A = F(\pi_1, ..., \pi_N)$, then $F(W_A, \dots, W_A) = W_A$. Using the definition of the value of time (33) rearranged as $T = V / W_A$, the IPF (32) can be rewritten in terms of $V$ and the shadow price aggregator $W_A = F(\pi_1, ..., \pi_N)$ as:

$$IPF = \ln\left(T+1\right) = \ln\left(\frac{V}{W_A}+1\right) = \ln\left(\frac{V+W_A}{W_A}\right) = \ln\left(\frac{R}{W_A}\right) \tag{34}$$

where, for convenience, $R$ is defined as:

$$R \equiv V + W_A = W_A(T+1) \tag{35}$$

*Equivalent solutions obtained by primal methods, dual methods and change of variables*

The IPF (34) has the same structure and properties as an indirect production function arising from the following problem which is an artificial re-arrangement of the optimization problem under consideration:

By analogy with the definition of $\tilde{t}_i$ in terms of $t_i$ , now consider the slightly more complex transformation of variables::

$$x_i = (W_A / \pi_i)(t_i + \gamma_i) \tag{36}$$

and observe that this and the definition of $R$ given in (35) suggests that the artificial constructs $x_i$ satisfy a budget constraint:

$$\sum_{i=1}^{N} \pi_i x_i = R \tag{37}$$

A suitable choice for a functional form for $W_A$, which will lead to a considerable simplification in (39) below, is:

$$W_A = \prod_{i=1}^{N} \pi_i^{\gamma_i} \tag{38}$$

Note also that from the definition of $x_i$, $\ln(t_i + \gamma_i) = \ln(\pi_i x_i / W_A) = \ln \pi_i + \ln x_i - \ln W_A$. Finally, note that this implies

$$\sum_{i=1}^{N} \gamma_i \ln(t_i + \gamma_i) = \sum_{i=1}^{N} \gamma_i \ln \pi_i + \sum_{i=1}^{N} \gamma_i \ln x_i - \ln W_A = \sum_{i=1}^{N} \gamma_i \ln x_i \tag{39}$$

where use has been made of the logarithmic form of (38) to simplify (39). It follows from this change of variables (from $t_i$ to $x_i$) that the information production optimization problem is the same as maximizing

$$\sum_{i=1}^{N}\gamma_i \ln x_i - \sum_{i=1}^{N}\gamma_i \ln \gamma_i \qquad (40)$$

subject to the budget constraint (37). This is a very simple optimisation problem which has the same form of a traditional consumer optimisation problem with shadow prices $\pi_i$ appearing in the budget constraint, and with an overall 'money' resource $R$ available for implicit allocation on the $x_i$. While this can be solved directly by the Lagrangean method to give the solution in share form:

$$\frac{\pi_i x_i}{R} = \gamma_i \qquad (41)$$

It is also of a form that generates an IPF that satisfies optimality properties implied from duality theory. In fact, as long as $W_A$ is defined by (38), the same solution (41) can equally be obtained from the IPF (34), in its RHS incarnation, using an envelope theorem from duality theory, by application of Roy's identity which in share form asserts:

$$\frac{\pi_i x_i}{R} = -\frac{\partial IPF / \partial \ln \pi_i}{\partial IPF / \partial \ln V} \qquad (42)$$

The equivalence of these two methods - the primal method giving (41) by solving (40)-(37) by a Lagrangean technique; the dual method applying (42) to (34)-(38)) - can be verified by substituting (38) into (34), applying (42) and observing that the outcome is (41).

Additionally, since the problem solved here involves a simple change of variables it should not be surprising that (41) and (42) are also equivalent to (25). This can be seen directly by use of the artificial variable definitions $x_i = (W_A / \pi_i)(t_i + \gamma_i)$ and $R = W_A(T+1)$. Put together, these imply $\pi_i x_i / R = (t_i + \gamma_i)/(T+1)$ and applying this to (41) and rearranging leads immediately to (25). That is, the primal method for the original time allocation problem giving (25) and the primal method for the problem involving artificial variables giving (41) are actually equivalent because they are effectively solving the same problem, albeit the second in a translated form.

*An empirically useful generalisation*

The purpose of demonstrating the equivalence of the primal and dual approaches applied to the artificial problem and also the equivalence of the original and artificial primal problems is to now generalise the IPF (34) for an artificial problem for which application of

Roy's Identity is valid. This allows the generation of solutions to a more complex original primal problem by a combination of change of variables and indirect (dual) techniques. Effectively this approach employs more empirically relevant functional forms that are fully consistent with theory, without having to solve an impossibly complex primal optimisation problem.

Note that (34) consists of one overall value-of-time aggregate $V$ (or, via a simple translation, $R$) and one price aggregator $W_A$. This leads to demand equations for the allocation of time that can be referred to as 'rank 1'. A more empirically plausible system would be (at least) 'rank 2'. This can be found by generalizing the IPF (34) to include two separate price aggregators. The new price aggregator $W_B$ aggregates the same shadow prices but with different weights, $\delta_i$ which, like the $\gamma_i$ weights in $W_A$ must sum to unity. Although more complex functional forms could be chosen, such as CES or Translog, the combination of two Cobb-Douglas aggregators is all that is really required to extend empirical plausibility of demand systems greatly. Hence, define:

$$W_B = \prod\nolimits_{i=1}^{N} \pi_i^{\delta_i} \ , \ \sum\nolimits_{i=1}^{N} \delta_i = 1 \tag{43}$$

Also introduce a 'non-homotheticity parameter', $\zeta$ say. Then a suitable modification of the IPF (34) is:

$$MIPF = \left( \frac{R}{W_B} \right)^{\zeta} \ln\left( \frac{R}{W_A} \right) \tag{44}$$

*The empirically generalized solution to the artificial problem*

Now apply Roy's identity to (44) to obtain the solution by dual means to an artificial problem:

$$\frac{\pi_i x_i}{R} = \frac{\gamma_i + \delta_i \zeta \ln(R / W_A)}{1 + \zeta \ln(R / W_A)} \ , \ i = 1, \ldots, N \ . \tag{45}$$

*The empirically generalized solution to the search time allocation problem*

Using the definition of the time shares $\tau_i = t_i / T$ together with the change of variables formulas $x_i = (W_A / \pi_i)(t_i + \gamma_i)$ and $R = W_A(T+1)$, (45) can be unravelled to give the solution of the implicit time share system as:

$$\tau_i = \frac{\gamma_i + \left( \delta_i + \frac{\delta_i - \gamma_i}{T} \right) \zeta \ln(T+1)}{1 + \zeta \ln(T+1)} \ , \ i = 1, \ldots, N \ . \tag{46}$$

The parameters $\gamma_i$ and $\delta_i$ represent time share allocations relevant to the search for information on product $i$ for the time poor (where $\gamma_i$ is most relevant) and for the time rich (where $\delta_i$ assumes increasing importance as $T$ rises). A typical consumer will choose time allocations that are some average of these extremes, and for a person with the amount of time $T$ available to allocate in total, the optimal choice is given by the combination of parameters as constructed in (46).

When additional demographic data is available along with the full IPTS clickstream database, separate equations of type (46) could be estimated for consumers from different demographic groups. The proposed modelling can be summarised as:

## 3. Linking the model to the database

The model proposed here fits neatly with the ITPS clickstream database. Table 1 illustrates this by presenting a histogram representing an example of a record from the database after some pre-processing. The pre-processing consists of extension of data fields, collation into common events, aggregation into categories and (in principle) averaging over consumers in identified income or demographic groups to get a sample of representative data.

Table 1: Histogram of one record from the processed database – with model symbols

| Category ID | Category name | Time Share |
|---|---|---|
| 1 | eCommerce | $\tau_1 = 0.357$ |
| 2 | Search Engines, Portals, & Communities | $\tau_2 = 0.257$ |
| 3 | Telecom & Internet Services | $\tau_3 = 0.133$ |
| 4 | Finance | $\tau_4 = 0.071$ |
| 5 | Computers and consumer electronics | $\tau_5 = 0.058$ |
| 6 | Entertainment | $\tau_6 = 0.040$ |
| 7 | Travel | $\tau_7 = 0.030$ |
| 8 | Home & Fashion | $\tau_8 = 0.028$ |
| 9 | News & Information | $\tau_9 = 0.019$ |
| 10 | Education & Careers | $\tau_{10} = 0.007$ |
| | | |
| | Total time in hours | $T = 7.3$ |

Model based symbols have been added to Table 1 to match the symbols in the system of equations to be estimated. This system of estimating equations is drawn from the model developed as equation (46) and reproduced here in the context of a 10 category example:

$$\tau_i = \frac{\gamma_i + \left(\delta_i + \dfrac{\delta_i - \gamma_i}{T}\right)\zeta \ln(T+1)}{1 + \zeta \ln(T+1)} \text{ , } i = 1,\ldots,10 \text{ , } \sum_{i=1}^{10} \gamma_i = 1 \text{, } \sum_{i=1}^{10} \delta_i = 1 . \qquad (47)$$

Based on the 10 categories of time expenditure given in Table 1, there are 10 equations of type (47) that can be estimated. Actually, only 9 of these equations are independent (because the shares, the $\tau_i$ variables, add to unity). Given a sample of Table 1-type histograms, one can estimate a system of 9 equations simultaneously and obtain estimates of the parameters $\gamma_i$ ($i = 1,\ldots,9$), $\delta_i$ ($i = 1,\ldots,9$) and $\zeta$. Then $\gamma_{10}$ and $\delta_{10}$ may be obtained by the adding-up conditions given with equation (47).

To interpret this time allocation equation and briefly indicate aspects of its usefulness, suppose that the amount of time $T$ that was available to be allocated was very small (say

0.001 hours rather than the 7.3 hours given in Table 1). Then $\ln(1.001) \approx \ln 1 = 0$ and equation (47) implies $\tau_i \approx \gamma_i$. So, for example, the parameter $\gamma_i$ in (47) represents the proportion of time spent on category $i$ if the total time available is extremely limited. On the other hand, as $T$ gets larger in (47), $\tau_i$ gets closer to $\delta_i$. Thus the parameter $\delta_i$ in (47) represents the proportion of time that would be spent on category $i$ if an extremely large amount of time was available. The parameter $\zeta$ controls the speed with which a consumer modifies their time share allocation as the total amount of time available changes. This could differ across consumers in different demographic categories, or could even vary for a given consumer in changed circumstances, such as the development of innovations that might require an adjustment to search times.

If the total time available is small, one would expect to spend a greater proportion of this scarce time on the 'necessities' and less on the 'luxuries'. If a much greater amount of time is available, one would expect to need to spend a lesser proportion on necessities and have a greater proportion of time available to spend on those categories that would be thought of as luxuries. As a result, if after estimation we find that $\gamma_i > \delta_i$ then category $i$ can be described as a necessity. On the other hand, if it is found that $\gamma_i < \delta_i$ then category $i$ could be described as a luxury. Categorisation of this type facilitates forecasting future trends.

## 4. Conclusions

There is a natural model of Internet search time that fits neatly with the database as it currently exists. The basic structure of this model has been discussed in detail. The model explains how total available time would be shared out in searches for information on a variety of products. Probably the most pressing need for further development of this model would be to allow for the endogenisation of total time available. The model presented in this report was motivated by a simple 'rank 1' demonstration and the report then moved to a 'rank 2' model for greater realism in the algebraic representation. It is likely that this will have to be extended to at least a 'rank 3' specification to allow for the empirical complexities in endogenisation of total time available that were illustrated diagrammatically in the report.

Further proposals for extensions to both the model and the database will naturally arise with use. Suggestions include extensions to facilitate analyses that distinguish between different demographic groups. A general objective behind this extension would be to develop a linked model and database that can address consumer welfare considerations, such as determining the value of the Internet and of related innovations in information and communications technology for various demographic categories of individuals and households in society.

European Commission
EUR 25836 – Joint Research Centre – Institute for Prospective Technological Studies

Title: An Appropriate Model for the Estimation of Consumer Time Expenditure Patterns on the Internet

Author: Russel Cooper

Luxembourg: Publications Office of the European Union

2013 – 37 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online)

ISBN 978-92-79-28780-0 (pdf)

doi:10.2791/81009

Abstract
IPTS recently acquired a consumer internet clickstream database containing the full set of annual (2011) clickstream records for about 25,000 internet users in the five largest EU economies (UK, Germany, France, Italy and Spain). It contains time spend on each webpage and socio-economic characteristics of the internet users.
This study describes a model of consumer Internet time use that is capable of empirical implementation with the clickstream database. A natural model of Internet search time could be developed to fit neatly with the database as it currently exists. The basic structure of such a model is discussed in detail.
The study recommends that IPTS should not rely on any existing model of Internet time expenditures. There does not appear to be a good contender in the literature that would adequately address the requirements of IPTS nor the realities of available data. A purpose-built model needs to be considered.

LF-NA-25836-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

ISBN 978-92-79-28780-0

Publications Office