INTERNATIONAL INDUSTRIAL ORGANIZATION CONFERENCE
Arlington, may 2008

The digital distribution of online press: Categorization and evaluation of business models[1]

Preliminary version

Marc Baudry* & Fabrice Rochelandet**

* CREM (crem.univ-rennes1.fr/), University of Rennes 1, marc.baudry@univ-rennes1.fr

** ADIS (www.adislab.net), University of Paris 11, fabrice.rochelandet@u-psud.fr

[1] This paper is part of a study supported by French Research National Agency (ANR-05-JCJC-0204-01), which consists in evaluating DRM-based models and alternatives in the fields of digital distribution of cultural goods.

I. Introduction

The case of online press proves important to shed light on informational goods and digital business models. In particular, online press is among the first content industries to be online and since then, to test diversified methods to finance its fixed costs (Mings & White, 2000). Online newspapers are also among the most widely demanded and visited websites. Two types of competition matter as regards the economic analysis of online press. The first type is competition between digital titles on the one hand and press paper on the other hand. The second type is competition between different business models for online contents. The first type of competition has been widely investigated. More specifically, empirical works have addressed the alleged cannibalization effect which states that digital contents are substitutes rather than complements of paper contents, even for a same title that exists in both versions[2]. By contrast, to our knowledge nothing is said about the competition between online websites themselves and success or failure of the different business models.

This paper aims at filling this gap. A first important task for this purpose is to identify the main business models (Section II). Instead of following existing works that make use of few cases considered *a priori* as typical to analyze business models, our methodology builds on a "bottom-up" approach. We first survey a large sample of websites of online press to code qualitative variables characterizing their strategy both in terms of the services and rights associated with contents and in terms of commercialization and financing and, finally, we implement clustering analysis to identify the main business models. In order to compare websites that target a similar market, we focus on the French online press and do not consider other French speaking online titles (Canadians, Belgians, Swiss and Africans). Websites have been selected on the basis of their rapid accessibility through a standard browser by a non specialist of online press. The distances of websites to the centre of each main cluster may be introduced as explanatory variables in an econometric model aiming at explaining the relative economic performance of online titles. This suggests an interpretation in terms of spatial differentiation. Therefore, prior proceeding with an econometric test of the impact of the main identified business models on the performance of websites we develop a simple microeconomic model of the behavior of consumers who have to allocate their time and income between differentiated online titles (Section III). The main interest of the model is to give microeconomic foundations to the audience share attraction model used for econometric purposes. This audience share attraction model which is in essence similar to the market share attraction model commonly used in quantitative marketing serves as a basis to test the influence of business strategies on the performance of websites. Due to the inability to obtain suitable data, a direct test of the influence of business strategies on profit

[2] Some reject the hypothesis of cannibalization for press (Deleersnyder and al., 2002; Kaiser and Kongsted, 2005). Flavián and Gurrea (2006) find that online and newspaper contents can be complementary if they prove sufficiently differentiated according to reader preferences and use situations. Pauwels & Dans (2001) find that print circulation of press titles increases digital visits. But they do not test the inverse relationship. By contrast, many studies reinforce the hypothesis of cannibalization (Filistrucchi, 2005; Gentzkow, 2007; Simon and Kadiyali, 2007 for instance).

is not possible. The adequacy of the audience of websites as a proxy measure of performance is addressed. Estimation results of the model are then presented (Section IV). Because of the alleged role of habits and network effects, a special attention is devoted to the treatment of the dynamics of audience shares.

II. Defining business models

In this section, we first briefly review the existing literature on business model for online press and then propose to overcome its shortcomings by deducing global business models from a case survey. Such a "bottom-up" approach consists in surveying as many cases as possible and then codifying them according to various criteria built from the previous literature. Then, data analysis is applied to highlight clusters of cases that have some consistency.

II.1. Previous literature

Several papers categorise emerging business models concerning the electronic distribution of contents. They all follow the same approach, that is to build general models from specific cases and then look at the role of DRM systems. Mings & White (2000) identify four generic business models through which online newspapers have tried to achieve profitability. The most important ones are the subscription and advertising models.

As regards the subscription model, the key issue is to determine whether people are willing to pay and for what. Offering paying services on an initially free environment or moving from free information to subscription access are not easy tasks but require specific conditions. For instance, the Wall Street Journal has been often quoted as a successful case, but it is in fact specialized in a specific content and readership. Any newspaper does not benefit from such a market position. Many business sub-models based on subscription can be explored. The 'new subscriber' and 'maturation' sub-models consider digital newspapers as a way to initiate and induce young readers to pay when their income will become sufficient. As such, digital versions prove 'loss leaders'. Other sub-models are based on versioning by offering readers information not available in the print channel (in-depth articles, up-to-date news, video complements…) or on personalized services by pre-selecting topics according to individuals' needs. The viability of subscription model has been explored by various empirical studies. Based on random-sample telephone survey of 853 Hong Kong residents, Chyi (2005) finds that very few readers are willing to pay for online news. Using a hierarchical regression analysis, this study shows that age and newspaper use are related to paying intent, whereas income is not.

The advertising model is built on a two-sided market logic. This generic model associates free access to news with ads and financing of the fixed costs by advertisers. Many business models can be implemented from the TV-like mass advertising model to the online personalized and interactive

advertising enabling readers to order products thanks to links incorporated in the ads. Though this model benefits from audience tracing tools and direct linking with consumers, some constraints are to be taken into account to design viable models such as the risk of a too narrowed or dispersed audience, the possibility for readers to avoid or eliminate ads, and the real costs of one-to-one marketing (implying an 'informational logistics').

Two other business models have been identified by Mings & White (2000). In the 'transactional model', digital newspapers play the role of middlemen bringing buyers and sellers together in an electronic marketplace. The ZDNet case illustrates this model: it reviews products and firms, provides technology survey and proposes interactive services enabling readers to interact with columnists and professional bloggers. This model is constrained by the reputation of the provider and the transaction costs associated with the splitting of revenues among all the facilitators of the transaction. The last identified model is the 'bundled model' according to which publishers establish partnership with other publishing and Internet entities. To access newspapers, readers have to pay a fee in addition to the subscription to the online service (for instance, Internet access and cell phone subscription). Network providers benefit from the value that newspapers represent for their subscribers and newspapers have an easy access to the installed base of network providers. However, such a model obviously suffers from the size of the installed base and the revenue sharing procedure.

It is difficult to derive from Mings & White (2000) a systematic method that permits to categorize and evaluate business models. More precise methods have been suggested by recent studies examining digital business models in the field of music and movies. Einhorn & Rosenblatt (2005) stress the role of digital right management systems (DRMs) as a key device for content providers to appropriate value online. They list DRM-based business models in the music industry. Interestingly, they look at the services provided by the different cases they analyse. "Versioning allows consumers to choose among a number of service options instead of being confined to any one." So anticopying tools are necessary for these new market solutions to be efficient by preventing resale and arbitrage between categories of consumers (Meurer, 1997). However, despite the use of DRM, business models have evolved with regard to consumers' tastes. Competing online services are differentiated on the set of services they offer to consumers from burning and transferring options to recommendation, rating and personalised playlist functions. However, the viability of online services depends on other attributes than the sole usability. In this way, Regner et al. (2006) envisages a larger spectrum of digital business models. They make the hypothesis that ICTs permit to design new services by decoupling the functions of payment and rights transfer (scope of authorised uses) from the actual distribution of contents. Thereby, they categorise different models. DRM-based retail models strictly link payments and rights, that is payment conditions both access and rights of use. By contrast, free access model associates the set of uses of contents with a compulsory levy. Between these two cases five models are characterised by a decreasing link between payments and rights: (1) soft DRM; (2) variable pricing; (3) super-distribution; (4) voluntary contributions and (5) complement-based revenues. In the same

way, Gee & Lubomira (2006) focus on the market for DRM systems. They suggest that DRM acceptability and switching costs borne by consumers are challenged by the variety of business models. They classify business models according to the sources of revenue: sales of operating systems, sales of DRM-compatible hardware, sales of contents or Web traffic (advertising), DRM licensing. Business models range from the elementary case of DRM licensing (indirectly associated with contents distribution) to the Microsoft case (which associates the 4 sources of revenue). However, this study only considers one criterion (the source of revenues) and it does not look at alternative business models. From these various studies, we elaborate Table 1 to distinguish business models.

<INSERT TABLE 1>

All the surveyed literature is essentially descriptive and does not take into consideration other competing factors in order to assess the viability and evolution of the listed services. It can not serve to predict what and under what conditions combinations of services, rights, and pricing might be successful. Studies follow the same approach, that is to derive general categories from specific cases. In fact, such a "top-down" approach stipulates more than it demonstrates the existence of models from emblematic cases. For instance, iTunes represents the unit sale model, AOL In2TV the advertising one, Wall Street Journal the subscription one and so on. Moreover, using the media success or the market share of a specific case to demonstrate its importance can be delicate in the current period of technological change. In addition, such an approach suffers from lack of precision because it does not measure intra-model differences and inter-models similarities. Finally, it allows neither to determine the empirical representativeness of such models (only the visible part of the iceberg might be analysed), nor to study the evolution of models by identifying the (potential) factors of success or failure. In what follows, we adopt a different approach that deduces global business models from a case survey and evaluate their economic success by taking into account the sole online market. Such a "bottom-up" approach consists in surveying as many cases as possible and then codifying them according to various criteria built from the above literature. Then, data analysis is applied to highlight clusters of cases that have some consistency. Success of models is then evaluated by adapting the market share attractor method.

II.2. Cluster analysis

With the aim to identify distribution models of online press, a database was elaborated from a survey of 63 websites from which digital contents can be legally obtained. These cases were classified by using about fifty variables, essentially qualitative and binary. The information necessary to produce profiles can be of four types:

- "Rights": This group of variables corresponds to the scope of use, i.e. the set of rights granted to users by providers of digital contents and copyright owners (limitation and duration of usage, portability, ability to share…). Regner *et al* (2006) uses this kind of variables to evaluate the level of contents of digital media ("convenience of use")[3]. By contrast, our approach consists in applying this criterion to differentiate contents.
- "Services": Other variables qualify the services themselves rather than the possibilities associated with contents once obtained. This kind of variables encompasses the size of the online catalogue, the diversity of titles, the presence of premium contents and complementary products (videos…), the degree of customisation of services, the interactivity… all features that are associated to the service of delivering digital contents and perceived by consumers as vectors of differentiation.
- "Financing": We also consider the appropriability methods used by the service providers to recoup their costs and to compensate rights owners. It can be the revenues from sales of contents or sales of complementary/ancillary products. It can also be the receipts from advertising and the exploitation/resale of personal data. Currently, there is a crucial debate about the viability of direct appropriability with regard to advertising revenues, which are supposed to be more suitable to the "everything-is-for-free" digital environment.
- "Commercialisation": Another category of variables concerns the marketing and pricing of digital contents. Pricing and versioning according to the set of possible uses; with regard to consumers' tastes; paying or free access to contents ; with or without ads and exploitation of personal data; and so on.

Table 2 displays the complete list of variables. Since the “Financing” and “Commercialisation” methods may be thought of as the counterparts of “Services” and “Rights” offered to consumers, variables are grouped in two categories respectively named “Commercialisation & Financing” and “Services & Rights”. This dichotomy of variables serves as a basis to identify the main strategies of online press providers and to define business models.

<INSERT TABLE 2>

Typologies of cases are created by using successively two classification methods: firstly, Ascending Hierarchical Clustering (AHC) to identify the main clusters of cases on the less subjective grounds and, secondly, Moving Centres Clustering (MCC) in order to consolidate the partitioning obtained

[3] However, in their study, prediction appears to depend mostly on the perception of the researcher. In fact, some individuals might prefer fewer rights but more ergonomic services to singularize themselves from others (for instance, the first consumers of iTunesMS). A more scientific approach requires a specific consumer survey like the survey INDICARE (2005).

thanks to AHC[4]. This two stages classification approach has been implemented separately with the "Services & Rights" variables presented in Table 2 and with the "Commercialisation & Financing" variables. For each stage, the metric used is the Chi-square distance, which is more suitable for qualitative data than the usual Euclidian distance. The guideline to determine the relevant number of clusters was a compromise between a focus on the main business strategies (i.e. not too many clusters) and the need of a sufficient differentiation of clusters to ease the interpretation (i.e. not too big clusters and thus sufficiently numerous clusters). We finally opted for five clusters. The composition of clusters as regards the "Services & Rights" strategy and the "Commercialisation & Financing" strategy are respectively synthesised by Tables 3 and 4. Key variables that are statistically either overrepresented or underrepresented in a cluster compared to the whole set of cases are also reported in Tables 3 and 4. These variables help characterising and comparing the clusters.

<INSERT TABLE 3>

<INSERT TABLE 4>

Both types of strategies exhibit a strong asymmetry in terms of the sizes of clusters with some big clusters encompassing a large number of cases and the remaining clusters corresponding to small clusters or even a sole atypical case. There are three atypical cases in terms of the "Services & Rights" strategy: *Relay* which is a website that provides legal copies of magazines in exactly the same form than their paper version, *L'Humanité* which is the digital mirror of the newspaper and imposes no legal limits in terms of rights of use[5], *La dépêche du Midi* which essentially departs from other website due to the specific application required to read contents displayed in a proprietary format. The two main clusters in terms of the "Services & Rights" are associated with opposite strategies. Cluster S&R4 gathers websites providing general interest but obsolescent contents and that do not exploit techniques differentiating digital contents from paper contents (interactivity, digital multiproducts…). Conversely, cluster S&R5 gathers websites that fully exploit all these techniques. As regards the "Commercialisation & Financing" strategy, there is one atypical website (*Virginmega* which sales copies of press magazines per unit) and a small cluster (C&F2) that essentially gathers websites of well-known daily newspapers (*Le Monde*, *Les Echos*, *La Tribune*, *Le Parisien*) financed by a wide range of techniques of direct sales methods (per unit sales, package sales, subscription, 2nd and 3rd degree discrimination). Cluster C&F4 is characterised by an overrepresentation of the same techniques of direct sales except per unit sales and 3rd degree discrimination but also an overrepresentation of the

[4] Indeed, the main drawback of AHC is that it doesn't necessarily produce the lowest intra-group inertia for a number of clusters fixed *a priori*. To solve this problem, we apply MCC with the centres calculated through the AHC.
[5] The reader that is unused with French newspapers has to be advised that *L'Humanité* is affiliated to the French Communist Party. The strategic choices made by this website are thus susceptible to be better understood on political strategic grounds rather than economic strategic grounds.

exploitation of personal data. These two clusters may be opposed to clusters C&F3 and C&F5 which do not widely use direct sales. What departs cluster C&F5 from cluster C&F3 is the use of advertising.

<INSERT TABLE 5>
<INSERT TABLE 6>

A rapid look at Table 3 and 4 suggests that those cases that are gathered in a similar cluster as regards the "Services & Rights" strategy are often gathered in a same cluster as regards the "Commercialisation & Financing" strategy. This is confirmed by Table 5 where the cases are dispatched with respect to which cluster they belong for both types of strategy. The hypothesis that the two classifications are independent is statistically rejected (the Chi-square statistic associated with the test amounts to 35.2287). Therefore the results of the classification in terms of "Services & Rights" and "Commercialisation & Financing" have finally been used as inputs for a third cluster analysis that aims at identifying global business strategies. How the resulting five global clusters (denoted respectively G1 G2 G3 G4 and G5) are dispatched according to the clusters for "Services & Rights" and "Commercialisation & Financing" strategies is also reported in Table 5. It clearly appears that these global clusters are associated with a unique cluster in terms of "Commercialisation & Financing" but eventually multiple clusters in terms of "Services & Rights". The cases composing each global clusters are mentioned in Table 6 which also displays those strategies that are over-represented or under-represented in each global cluster compared with the whole set of cases. The largest global cluster (G5) is essentially defined by specificities in terms of "Commercialisation & Financing" (over representation of C&F5 and under representation of C&F4). We thus conclude that it gathers websites oriented toward indirect value creation through advertising and denote it the "ads business model". The second largest cluster (G4) is associated with S&R5 (opposite to S&R4 which is under-represented) and C&F4 (opposite to C&F3 which is under-represented). It may be qualified as the "innovative business model" in the sense that both techniques differentiating digital contents from paper contents in terms of "Services & Rights" and a wide range of methods of direct sales that differentiate online press from paper press are used. The global cluster G3 is opposite to G4 and is thus qualified "old fashion business model". Finally, G2 is close to G4 except in terms of "Services & Rights" where the use of specific applications to read contents is over-represented and G1 is specific to *Virgin Mega*.

III. A model of audience shares

The behaviour of consumers facing differentiated web sites of online press is examined in this section. The focus is made on the allocation of both budget and time to derive a multinomial

econometric model of audience shares with microeconomic foundations. We first present the microeconomic model and then turn to some econometric developments.

III.1. The microeconomic model

At each date t, the allocation of time and budget to each website of online press is obtained as the result of a two stages optimization program. In the first stage, consumers have to choose how to dispatch a given time X^t between the visits of I websites given their budget constraint. In the second stage, consumers have to arbitrate between working time that determines their income on the one hand and leisure time, including the time X^t devoted to online press, on the other hand. Since our interest is on the determination of the audience shares of the different websites rather than on the determination of the time of leisure devoted to online press, we only detail the first stage.

A key distinction is made between the time devoted to search and visit a website i, denoted by x_i^t, and the quantity of relevant information and/or distraction really provided. This quantity is given by $q_i^t = x_i^t / \alpha_i^t$ where $\alpha_i^t > 0$ is a conversion parameter that may vary across websites and time, depending on the adequacy between websites characteristics at time t on the one hand and the consumer's taste on the other hand. The time constraint implies that the sum of these times amounts to the total time X^t devoted to the consumption of online press. Visiting a website i induces an expenditure of $p_i^t \geq 0$ per unit of time that may differ from one website to another one[6]. As a result, the budget constraint states that the total expenditure $\sum_{i=1}^{I} p_i^t x_i^t$ induced by the consumption of online press plus the expenditure induced by the consumption of a quantity y^t of an aggregate good used as the *numéraire* amounts to the income R^t of the consumer at each date t. For computational convenience, we assume that the utility function has a quasi linear form of the entropic type. More precisely, it is linear with respect to the quantity y^t of the aggregate good and it is entropic with respect to the quantities q_i^t. We thus have $U\left(q_1^t, \cdots, q_I^t, y^t\right) = \sum_{i=1}^{I}\left(q_i^t\left(1 + \ln \beta_i^t\right) - q_i^t \ln q_i^t\right) + y^t$ where β_i^t ($i \in \{1, \cdots I\}$) is a preference parameter associated with each website and that eventually depends on the website characteristics. This utility function is decreasing and concave with respect to q_i^t as long as $q_i^t < \beta_i^t$ which is systematically satisfied if $\beta_i^t > X^t / \alpha_i^t$. Substituting $y^t = R^t - \sum_{i=1}^{I} p_i^t x_i^t$ from the budget constraint and x_i^t / α_i^t for q_i^t in the expression of the utility function, we obtain that the optimal allocation of time X^t is the solution to the following optimization program:

---

[6] $p_i^t = 0$ if the content provided by website i is for free at date t. Note also that payment is generally based on the number of articles viewed rather than on the time of lecture. However, we can reasonably assume that for a given consumer the time of lecture of each article is almost constant so that a payment based on that time is equivalent to a payment per article viewed.

$$\underset{\{x_1^t,\cdots,x_I^t\}}{Max}\sum_{i=1}^{I}\left(\frac{x_i^t}{\alpha_1^t}\left(1+\ln\beta_i^t\right)-\frac{x_i^t}{\alpha_1^t}\ln\left(\frac{x_i^t}{\alpha_1^t}\right)\right)+R^t-\sum_{i=1}^{I}p_i^t x_i^t \qquad (1.a)$$

subject to the time constraint

$$\sum_{i=1}^{I}x_i^t = X^t \qquad (1.b)$$

The solution for x_i^t ($i\in\{1,\cdots I\}$) is derived by standard optimization methods (see e.g. Anderson and al, 1992). Once expressed as a fraction of the total time X^t devoted to online press, the solution for x_i^t yields the audience share $S_i^t = x_i^t/X^t$ of the website i:

$$S_i^t = \frac{\alpha_i^t \exp\left(\left(\ln\beta_i^t/\alpha_i^t\right)-p_i^t\right)}{\sum_{j=1}^{I}\alpha_j^t\exp\left(\left(\ln\beta_j^t/\alpha_j^t\right)-p_j^t\right)} \quad \forall i\in\{1,\cdots,I\} \qquad (2)$$

Due to the quasi linearity of the utility function, the optimal audience shares do not depend on the consumer's income. More surprisingly, the optimal audience shares are invariant with respect to the total time X^t devoted to online press. If consumers differ in terms of income but not in terms of their preference parameters α_i^t and β_i^t, it then follows on that expression (2) also yields the average audience shares across all consumers and is a consistent aggregation of individual audience shares. In this sense, the system of audience shares formed by expression (2) constitutes an extension of a popular econometric model in marketing research which is the market share attraction model (Cooper and Nakanishi, 1988). Accordingly, the audience share of website i at date t can be expressed as the ratio between its attraction $A_i^t \geq 0$ and the sum of attractions for all websites:

$$S_i^t = \frac{A_i^t}{\sum_{j=1}^{I}A_j^t} \qquad (3.a)$$

with

$$A_i^t = \alpha_i^t \exp\left(\left(\ln\beta_i^t/\alpha_i^t\right)-p_i^t\right) \qquad (3.b)$$

The microeconomic model developed above sets the theoretical foundations of this specification which is generally introduced in an *ad hoc* way in order to satisfy the market share theorem (see Bell *et al*, 1975)[7].

III.2. The econometric model

The key components of the attractions in the audience share model (3) are not directly observed but may reasonably be considered as closely related to the characteristics of websites used to define business models for online press in the previous section. More precisely, we consider that the

[7] The market share theorem states that only specifications in terms of the ratio between each attraction and the sum of attractions satisfy the following two basic conditions characterising a consistent share model: 1) Each share takes a real value between zero and one 2) The sum of shares systematically amounts to one.

preference parameters α_i^t and β_i^t and the per unit of time cost p_i^t that differentiate websites from each others are correctly approximated by a functional form of the distance of each website to the centre of the business models obtained in the previous section[8]. Other explanatory variables may eventually be introduced. The only restriction on the functional form of attractions is that they take positive values to make sure that the share is also positive. A standard specification is the following Cobb Douglass functional form:

$$A_i^t = \exp\left(\sum_{k=1}^{K} \theta_k \ln z_{k\,i}^t + \omega_i^t\right) \tag{4}$$

where $z_{k\,i}^t$ denotes the value of regressor k (with $k = 1,\cdots,K$) for website i at date t and ω_i^t is a Gaussian random term with mean zero and standard deviation σ_i. This random term typically measures the influence of unobserved characteristics.

Habits and network effects are presumably responsible for some kind of histeresis in the dynamics of audience shares. In order to take account of such a histeresis, the lagged value of the audience share of a website is introduced in the list of regressors that appears in its attraction. The exact consequence of this introduction is highlighted by examining how it affects the forecast of the audience share. For this purpose, the Delta Method is used to compute the expected value of audience shares at date t given audience shares at the previous date. If we denote by $g\left(\omega_1^t,\cdots,\omega_1^t\right)$ the expression $A_i^t\left(\omega_i^t\right) \Big/ \sum_{j=1}^{I} A_j^t\left(\omega_j^t\right)$ of the audience share for website i at date t, the linear approximation of $g\left(\omega_1^t,\cdots,\omega_1^t\right)$ when $\left\{\omega_1^t,\cdots,\omega_I^t\right\}$ is a vector of zeros is given by

$$\begin{aligned} g_i\left(\omega_1^t,\cdots,\omega_1^t\right) = \quad & g_i(0,\cdots,0) \\ & + g_i(0,\cdots,0)\left(1 - g_i(0,\cdots,0)\right)\omega_i^t \\ & - \sum_{j\neq i} g_i(0,\cdots,0)\, g_j(0,\cdots,0)\,\omega_j^t \end{aligned} \tag{5}$$

Given that the ω_i^t are independently distributed (but not necessarily homocedastic) with zero as expected value, the expectation and variance of $g\left(\omega_1^t,\cdots,\omega_1^t\right)$ are given by the two following expressions:

$$\mathrm{E}\left[g_i\left(\omega_1^t,\cdots,\omega_1^t\right)\right] = g_i(0,\cdots,0) \tag{6.a}$$

$$\begin{aligned} V\left[g_i\left(\omega_1^t,\cdots,\omega_1^t\right)\right] = \sigma_i^2 \; & g_i(0,\cdots,0)^2 \left(1 - g_i(0,\cdots,0)\right)^2 \\ & - \sum_{j\neq i} \sigma_j^2 \, g_i(0,\cdots,0)^2 \, g_j(0,\cdots,0)^2 \end{aligned} \tag{6.b}$$

According to (6.a), the expression of an audience share with null random terms yields a relevant forecast of this audience share, while the prediction errors are assessed by (6.b). If the previous audience share is the sole explanatory variable in the model and its coefficient amounts to 1 then,

[8] Since most of the characteristics of websites are qualitative, the Chi square distance is used.

according to the functional form (4) and the definition of audience shares, the forecasted audience share exactly amounts to the previously observed audience share. In this sense, the model corresponds to a random walk dynamics and there is a strong histeresis effect in the dynamics of audience shares.

Though the structure of audience or market shares is closed to that of a multinomial Logit model it departs from this type of model as regards its interpretation. Multinomial Logit models are used to analyse which one of several incompatible items is chosen by a consumer. The specification of the probabilities of the different outcomes is then almost similar to expression (3.a) with (4) but yields the theoretical audience shares obtained across a high number of consumers, each of them choosing only one of the different items. By contrast, expression (3.a) with (4) states that each consumer dispatches his consumption among all items. As a result the explained variables are not qualitative variables indicating the choice made by a consumer as in a multinomial Logit model but are directly given by the observed shares for a consumer or a group of consumers if a consistent aggregation of individual shares is applied. After some algebraic modifications, audience or market share attraction models are estimated by standard econometric methods. More precisely, we use the base brand estimation procedure proposed by Fok Franses and Paap (2001). Following these authors we subtract the natural logarithm of the audience share of a base website to the natural logarithm of the audience share of the other websites so as to obtain a set of $I-1$ log-linear equations that can be easily estimated by maximum likelihood. This method implicitly consists in a normalisation of the base website attraction to unity. With the Cobb Douglas specification (4), the equations to be estimated take the form

$$\ln A_i^t - \ln A_I^t = \sum_{k=1}^{K} \theta_k \left(\ln z_{k\,i}^t - \ln z_{k\,I}^t \right) + \left(\omega_i^t - \omega_I^t \right) \qquad (7)$$

where the residual term $\omega_i^t - \omega_I^t$ is Gaussian with zero as expected value and $\sigma_i^2 + \sigma_I^2$ as variance if the random shocks ω_i^t and ω_I^t are independently distributed. Note that we do not necessarily assume homocedasticity. The reason for this is that according to (6.b) the variance of audience shares differs from one website to another one even if the random terms ω_i^t ($i \in \{1, \cdots, I\}$) are identically distributed. In order to avoid a too high influence of the specification of audience shares on their variance we introduce a counterbalancing effect by allowing the variance of random terms to differ.

IV. Estimation results

The model developed in the previous section is estimated to assess the impact of business models on the audience of online press websites. Information reported by *Alexa* is used to obtain data for audience share. We first provide a description of these data and then present the main results as regards the influence of business models on audience shares and their dynamics.

IV.1. Data and estimation method

A measure of performance is required to assess the impact of business models on the audience of online press websites. Since profits generated by online activity is seldom reported by firms a proxy variable has to be found. For this purpose, we take advantage of the so-called spiral effect originally outlined by Furhoff (1973) and Gustaffson (1978) for press paper. The key idea is that profits generated by a title are contingent on audience which in turn depends on the quantity, the quality and the adequacy of contents which are assumed to improve with the financial means. The spiral effect thus implies a strong correlation between profits and audience for which it is easier to collect data. The evaluation of business models proposed in this paper more specifically builds on data collected by *Alexa* for the measurement of the audience of internet sites[9].

Alexa computes traffic indicators based on a three months moving average of aggregated historical traffic data from millions of *Alexa* toolbar voluntary users. The main indicator developed and published by *Alexa* is "Traffic Rank" which yields the position of the site in interest with respect to all the sites on the web. The main drawback of this indicator for the present study is that we are interested in the position of each site of online press in our database with respect to the other sites in the same database, not all sites on the web. Therefore we rather used a combination of the indicators "Reach" and "Page Views per User". The "Reach" indicator measures the percentage of all internet users who visit a given site. The "Page Views per User" are the average numbers of unique pages viewed per user per day by the users visiting the site. Note that multiple page views of the same page made by the same user on the same day are counted only once. Once multiplied, the "Reach" and "Page Views per User" indicators yield a measure of the total number of visits of pages belonging to a same site. A main advantage of this last measure is that it takes account of the fact that a user may visit a site only once but views numerous pages of this same site which means that the impact of the site is higher than if it contains a sole page. The product between "Reach" and "Page Views per User" has been computed for each online press site in our database and then divided by the sum for all sites in the database so as to obtain a measure of the audience share. These audience shares serve as a basis for the development of an econometric analysis of the performance of business models identified in the first section.

The causality effect in interest for this study is that of the choice of a business model on performance as measured by audience. However, some business models may reveals more suitable for low audience sites of online press and conversely. A strategy of free access and financing by the sales of complementary products for instance may ease the launch of a new site with low audience while per unit sales and advertising better suit to a well established title. As a result, the causality effect may be from the audience to the choice of business model. More generally, there exists a risk of cross causality and simultaneous endogeneity of audience and business strategy. This risk is addressed by using the usual instrumental variable. More specifically, a time lag is introduced between the audience

[9] See http://www.alexa.com/site/help/traffic_learn_more

share to be estimated and the survey data used to generate the business models clusters. Indeed, one may reasonably think that business strategies do not change drastically on a short time interval (a three months lag is used) so that past but recent business choices are correlated to current business choices but not endogenous when explaining the current audience share. In other to take account of ubiquitous effects of a business strategy, the basic model defined in (7) is improve in two ways: firstly, we allow coefficients associated with each strategy to differ according to the audience share, secondly we allow for the standard deviation of the error terms ω_i^t to depend on the audience share.

A differentiation of coefficients on the basis of the audience share introduces a truncation in the econometric model. Indeed, we have to distinguish between websites with an audience share that is lower to an exogenously given threshold and websites with an audience share that exceeds this same threshold. As a result, given that the error terms $\omega_i^t - \omega_I^t$ are independent, the likelihood associated with equation (7) is contingent on the fact that the explained variable $\ln A_i^t - \ln A_I^t$ lies behind or above an exogenous threshold

$$Ln_i^t = \begin{cases} \phi_i\left(h_i^t\right) / \Phi_i\left(k_i^t\right) & \text{if} \quad \ln A_i^t - \ln A_I^t \le \overline{S} \\ \phi_i\left(h_i^t\right) / \left(1 - \Phi_i\left(k_i^t\right)\right) & \text{if} \quad \ln A_i^t - \ln A_I^t \le \overline{S} \end{cases} \tag{8.a}$$

with

$$h_i^t = \ln A_i^t - \ln A_I^t - \sum_{k=1}^{K} \theta_k \left(\ln z_{k\,i}^t - \ln z_{k\,I}^t\right) \tag{8.b}$$

$$k_i^t = \overline{S} - \sum_{k=1}^{K} \theta_k \left(\ln z_{k\,i}^t - \ln z_{k\,I}^t\right) \tag{8.c}$$

The threshold value $\overline{S}$ used thereafter is the sample median of the explained variable $\ln A_i^t - \ln A_I^t$ in our database. ϕ_i and Φ_i respectively denote the partial and the cumulative distribution function of the error term $\omega_i^t - \omega_I^t$ and may differ from one website to another one due to heterocedasticity.

The underlying goal as regards the treatment of heterocedasticity is to compensate the arbitrary influence of the level of the audience share on its own variance. As already outlined, according to (6.b) the variance of audience shares differs from one website to another one even if the random terms ω_i^t ($i \in \{1, \cdots, I\}$) are identically distributed. It also appears that, according to (6.b) heterocedasticy of the random terms may counterbalance the arbitrary influence of the level of the audience share on its own variance. For this purpose, we assume that the variance σ_i of ω_i^t is given by $0.5\sigma^2 \Big/ \left(A_i^{t-1}\right)^{\gamma}$ where γ is a parameter to be estimated. The associated variance of the error terms $\omega_i^t - \omega_I^t$ in (7) thus reads

$$\sigma^2\left(\frac{0.5}{\left(A_i^{t-1}\right)^{\gamma}}+\frac{0.5}{\left(A_I^{t-1}\right)^{\gamma}}\right) \qquad (9)$$

The interest of this specification is that it encompasses the case of homocedastic error terms $\omega_i^t - \omega_I^t$ (when $\gamma = 0$) so that it enables us to test the assumption of heterocedastic random terms. All parameters in the model formed by expressions (8.a) to (8.c) and expression (9) are estimated by a standard maximisation of the log-likelihood except parameter γ which is estimated by selecting in a grid of values ranging from -2.5 to 2.5 with a step of 0.01 the value that is associated with the highest concentrated log-likelihood.

Finally, rather than directly use as explanatory variables qualitative variables indicating to which of the five business models G1 to G5 identified with the clustering method a website belongs we use the Chi-square distances of each website to the centre of each cluster. Indeed, preliminary tests have shown that more significant effects of business strategies were obtained with these distances compared with those obtained with qualitative variables.

IV.2. The influence of business models

Estimation results for the two parts model (i.e. when we allow for coefficients to be contingent on the fact that the explained audience share is higher or lower than the sample mean) with heterocedasticity are reported in Table 7. A first striking result is the highly significant impact of the lagged audience share. Moreover, the associated coefficient is close to one and may even be considered as not significantly different to that value depending on the precision of the test used. The comments following expressions (6.a) and (6.b) then suggest that the dynamics of audience shares is close to a random walk and that habit and network effects are essential factors contributing to the histeresis of audience shares. It follows on that the other variables explain the variation of audience shares rather than their absolute level. A second striking result is that more coefficients are statistically significant when the audience share exceeds the median sample. This is more particularly true for the coefficients associated to the distances to the centre of the global clusters. None of these coefficients is significant at a reasonable level when the audience share is lower than the sample median while distances to the centre of the business models G3 and G4 are significant for audience shares higher than the sample median. The “old fashion” business model G3 has a surprising positive impact on high audience shares whereas the “innovative” business model G4 has a negative though less significant impact. The relative importance of free access to contents (over-representation of C&F3 strategy) probably contributes to its success compared with business model G4 (with an over-representation of C&F4 strategy, i.e. direct sales of contents). However, these results have to be taken with caution. Indeed, the restriction of the coefficients for high audience shares (in the right part of Table 7) to the values

obtained for low audience shares (in the left part of Table 7) is strongly rejected by a log-likelihood ratio test but the inversed restriction (i.e. restriction of coefficients for low audience shares to the values obtained for high audience shares) is not rejected. It thus seems that a single part model is probably more relevant than the two parts model presented in Table 7.

<INSERT TABLE 7>

Table 8 displays the results obtained with the single part model. Results obtained when the assumption of heterocedastic random terms is maintained are reported in left column. The presence of heterocedasticy is confirmed by a log-likelihood ratio test of the restriction $\gamma = 0$ which is rejected. Nonetheless, results obtained with homocedasticity (i.e. when maximising the log-likelihood with $\gamma = 0$) are also reported in the right column. They do not strongly depart from results obtained with a correction of heterocedasticity which may be interpreted as a sign of robustness. The random walk nature of the dynamics of audience shares is confirmed since the coefficient of the lagged audience share is still highly significant and close to one. There are now three distances with a significant impact on audience shares: distances to the centre of G3 and G4 and distance to the centre of G5. The sign of the first two distances do not change compared with the two parts model. The coefficient associated with the distance to G5 is negative and more significant. Moreover, if a 5% confidence level is retained, the only coefficient associated with a distance that is statistically significant is the last one. We conclude that increasing the distance to the centre of the "advertising" business model is clearly detrimental to the increase of the audience share. The "advertising" business model thus seems to be more suitable than the "innovative" business model to create value from online press. Otherwise stated the old recipe consisting in providing contents for free and collecting revenues from the diffusion of ads is probably the best one even for online contents.

<INSERT TABLE 8>

V. Conclusion

Thanks to the elaboration of a database from a survey of 63 websites, a cluster analysis of the strategies adopted by online press providers has been implemented. Among the main business models revealed by this analysis, the "innovative" business that consists in offering a wide range of services in counterpart of direct payment does not prove to be the more successful. Indeed, increasing the distance to the centre of this cluster has a positive influence on the attraction of websites. Conversely, increasing the distance to the centre of the "advertising" business is detrimental to the attraction of websites. Generally speaking, it appears that indirect value creation is better than direct sales of online

contents to enforce the performance of online websites, as long as this performance is correctly approximated by the audience share.

References

Anderson, S.P., De Palma, A. and Thisse, J.F. (1992) *Discrete Choice Theory of Product Differentiation*, MIT Press, Cambridge, MA.

Bell, D.E., Keeney, R.L. and Little, J.D.C. (1975) "A Market Share Theorem", *Journal of Marketing Research*, 12, 136-141

Chyi, H.I. (2005) "Willingness to Pay for Online News: An Empirical Study on the Viability of the Subscription Model", *Journal of media economic*s, 18(2), 131-142

Cooper, L.G. and Nakanishi, M. (1988), *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*, Kluwer Academic Publishers, Boston.

Deleersnyder, B., Geyskens, I., Gielens, K. and Dekimpe, M. (2002) "How cannibalistic is the internet channel? A study of the newspaper industry in the United Kingdom and The Netherlands", *International Journal of Research in Marketing*, 19(4), 337–348.

Einhorn, M.A. and Rosenblatt, B. (2005) "Peer-to-Peer Networking and Digital Rights Management. How Market Tools Can Solve Copyright Problems", *Policy Analysis*, 534, 22 p.

Filistrucchi, L. (2005) "The impact of the internet on the market for daily newspapers in Italy", working paper, European University Institute; Available at: http://www.uni-kiel.de/ifw/konfer/network2004/¯listrucchi.pdf.

Flavian C. and Guerra, R. (2006) "The role of readers' motivations in the choice of digital versus traditional newspapers", *Journal of Targeting, Measurement and Analysis for Marketing*, 14(4), 325–335

Fok, D., Franses, P.H. and Paap, R. (2003), "Econometric Analysis of the Market Share Attraction Model ", *ERIM Report Series Reference No. ERS-2001-25-MKT*, Available at: ssrn.com/abstract=370883

Furhoff, L. (1973) "Some reflections on newspaper concentration", *The Scandinavian Economic History Review*, 21(1), 1-27.

Gee, L. K. and Lubomira, I. (2006) "Fragmentation versus Standardization in the Market for Digital Rights Management Solutions: A Case Study of the Online Market for Audio and Video Content", working paper, available at SSRN: http://ssrn.com/abstract=920711

Gentzkow, M. (2007) "Valuing new goods in a model with complementarity: online newspapers", *American Economic Review*, 97, 713–744.

Gustaffson, K.E. (1978) "The circulation spiral and the principal of household coverage", *The Scandinavian Economic History Review*, 26(1), 1-14.

Kaiser, U. (2006) "Magazines and their companion websites: competing outlet channels?", *Review of Marketing Science*, 4 , article 3.

Kaiser, U. and Kongsted, H. C. (2005) "Do Magazines' "Companion Websites" Cannibalize the Demand for the Print Version? ", ZEW Discussion Paper No. 05-49

Meurer, M.J. (1997) “Price Discrimination, Personal Use, and Piracy: Copyright Protection of Digital Works”, *Buffalo Law Review*, 45, 845-889

Mings, S. M. and White, P. B. (2000) “Profiting from online news: The search for viable business models”, *in* B. Kahin & H. R. Varian (Eds.), *Internet publishing and beyond: The economics of digital information and intellectual property*, pp. 62–96, Cambridge, MA: MIT Press

Pauwels, K. and Dans E. (2001) “Internet marketing the news: leveraging brand equity from marketplace to marketspace”, *Brand Management*, 8(4), 303-314

Regner, T., Barria, J.A., Pitt J.V. and Neville B. (2006) “Emerging Business Models in Digital Media”, working paper Imperial College London

Simon D. H. and Kadiyali V. (2007) “The effect of a magazine’s free digital content on its print circulation: Cannibalization or complementarity? ” *Information Economics and Policy*, 19, 344–361

Table 1: Business models derived from literature

Appropriability	Source of revenue	DRM/rights	Pricing and marketing	Cases/literature
Direct	Per unit sales ("a la carte")	- Strong and robust protection (trusted computing) - Usage limited to private consumption	- Prices are set according to the elasticity of demand - Strong link between payment, distribution and access to the content.	VOD
		- Less robust protection - Scope of uses>the sole private consumption - Perhaps some tolerance for copying	- Limit-pricing: the price is fixed by taking into account of (1) demand elasticity (2) the cost of getting contents from illegal networks	iTunes MS Varian, 2004 Regner, 2004
			Superdistribution - Viral marketing/pyramidal selling (payment can stimulate the circulation of a given title)	Weed Rosenblatt, 2004
		- No anti-copying protection - Scope of uses maximum - Perhaps watermarking (tracing)	Pricing according to the demand elasticity and competitors' reactivity (including illegal ones)	eMusic iTunes MS-EMI
	Subscription	Ex ante exclusion (filtering)	- payment of a yearly/monthly fee to access freely to the catalogue of content - fixed price not linked to the actual consumption	SVOD Online press
	Voluntary contributions	Weak or no DRM systems	Purely voluntary gift - online tipping - Sponsoring	Jamendo
	Hybrid	- Exclusion - Perhaps some tolerance for copying	- Mix: Fixed price plus variable pricing - voluntary payment according or not to a given price range (the preferences of consumers determine the actual price) - sampling	Magnatune (label) Regner and Barria (2005)
Indirect	(Tying) sales of complementary/ancillary goods to end users	- Tracing uses - ex post exclusion or not (restricting or preventing redistribution in some cases)	- Sales of ancillary goods or services (concerts, hardware, software, subscription to services directly linked or not to the delivered contents…) - Digital contents serve to stimulate the sales of other product. - Tolerance for copying and sharing (all the more since the sales of the complementary good or service are positively correlated with a large circulation of contents.	Online press Grateful Dead (copying+concerts) Gayer & Shy (2004) Connolly & Krueger (2005) Einhorn & Rosenblatt (2005)
	Muti-sided markets - Sales of Web traffic (audience) - Sales of personal information	- Tracing uses - Restricting or preventing redistribution according to the models (for instance, targeted ads need to eliminate redistribution)	- Revenue from agents different from end users	- AOL In2TV, Joost, Pplive… - Microsoft/Google - Online press - Spiralfrog - YouTube, Dailymotion… - Moby (advertising) Gee & Lubomira (2006) Einhorn & Rosenblatt (2005)
	Search for reputation (new artists)	Open access with no or weak conditions (for instance, a form to fill in)	- Increase wages or better job	Online music: independents Online press: Blogs ?

Table 2 : Variables used for the Classification of cases

"Services & Rights" variables		"Commercialisation & Financing" variables	
Way to delivery contents		*Financing*	
Streaming	0 if yes, 1 if no	Direct sales to consumers	0 if yes, 1 if no
Downloadable	0 if yes, 1 if no	Ads linked to contents/service	0 if yes, 1 if no
Physical media (CD, DVD, etc.)	0 if yes, 1 if no	Ads not linked to contents/service	0 if yes, 1 if no
Other (Podcast…)	0 if yes, 1 if no	Targeted ads	0 if yes, 1 if no
Usability of contents		Personal data (exploitation of)	0 if yes, 1 if no
Usable once	0 if yes, 1 if no	Sales of complementary/ancillary goods and services (hardware, press subscriptions…)	0 if yes, 1 if no
Limited usage	0 if yes, 1 if no	Other	0 if yes, 1 if no
Unlimited usage	0 if yes, 1 if no	*Marketing strategies*	
Limited duration	0 if yes, 1 if no	Unit sales	0 if yes, 1 if no
Unlimited duration ("permanent download")	0 if yes, 1 if no	Package sales	0 if yes, 1 if no
Legally Reusable (as an input)	0 if yes, 1 if no	One shot billing (ISPs, mobile phone…)	0 if yes, 1 if no
Legally shareable	0 if yes, 1 if no	Gift formula	0 if yes, 1 if no
Modifiable independently of content provider's will	0 if yes, 1 if no	Subscription	0 if yes, 1 if no
Consumer implication		Viral model	0 if yes, 1 if no
Communication interactivity	0 if yes, 1 if no	*Pricing*	
Interactivity / personalized service	0 if yes, 1 if no	2nd degree discrimination	0 if yes, 1 if no
Loss leader (sampling, front page…)	0 if yes, 1 if no	3rd degree discrimination	0 if yes, 1 if no
Access time to contents			
Immediate access (< 1 minute)	0 if yes, 1 if no		
Delayed access < 1 hour	0 if yes, 1 if no		
Delayed access < 24 hours	0 if yes, 1 if no		
Delayed access > 24 hours	0 if yes, 1 if no		
high-speed constraint	0 if yes, 1 if no		
Supply features			
Exclusive contents/ premium contents	0 if yes, 1 if no		
Catalogue size (importance with regard to competitors)	0 if yes, 1 if no		
General-interest (versus Thematic)	0 if yes, 1 if no		
Obsolescence of contents	0 if yes, 1 if no		
Diversified content providers	0 if yes, 1 if no		
New titles	0 if yes, 1 if no		
Digital multi-products	0 if yes, 1 if no		
Not digital Multi-products	0 if yes, 1 if no		
Terminal / mono-platform (vs. multi-platform) (TV/PC/mobile/portable device)	0 if yes, 1 if no		
Proprietary format (Itunes-Ipod, Atrac Sony…)	0 if yes, 1 if no		
Specific application (to access/consume contents)	0 if yes, 1 if no		
Forwarding charges	0 if yes, 1 if no		
Secondhand market	0 if yes, 1 if no		

Table 3 : Classification of cases for « Services & Rigths »

	OVER REPRESENTED VARIABLES:	**UNDER REPRESENTED VARIABLES:**	**CASES :**
cluster S&R1	- Delayed access (less that one hour) - High-speed constraint -Specific application	- Streaming - Modifiable content - Immediate access - Exclusive/Premium content	relay
cluster S&R2	- Legally reusable	- Streaming - Downloadable	l'humanité
cluster S&R3	- Proprietary format - Specific application		la dépêche du midi
cluster S&R4	- General interest - Obsolescence	- Interactivity/personalised service - Catalogue size - Diversified content providers - Digital multi products	Historia, le canard enchainé, alternatives éco, au féminin, l'étudiant, evene, ça m'intéresse, google news, eurosport, plein champ, actuenvironnement, newsweb, automoto, boursier, coté maison, culture femme, football, géo, goal, largeur, l'autojournal, le mague, le monde diplomatique, lire, Nice matin, sport24, télerama, terraeconomica, votre argent
cluster S&R5	- Interactivity/personalised service - Catalogue size - Diversified content providers - Digital multi products	- Obsolescence	Le figaro, libération, le monde, la croix, les échos, la tribune, le point, le nouvel observateur, l'expansion, le parisien, ouest-France, la voix du nord, le journal du net, l'internaute, infosciences, la vie financière, capital, agoravox, afp, 20minutes, agefi, argusauto, challenge, journalauto, la montagne, l'entreprise, l'est républicain, l'express, technoscience, virginmega

Table 4 : Classification of cases for « Commercialisation & Financing »

	OVER REPRESENTED VARIABLES:	**UNDER REPRESENTED VARIABLES:**	**CASES :**
cluster C&F1	- Unit sales - One shot billing - Gift formula		virginmega
cluster C&F2	- Direct sales - Per unit sales - Package sales - One shot billing - Subscription - 2nd degree discrimination - 3rd degree discrimination		Le monde, les échos, la tribune, le parisien, la dépêche du midi
cluster C&F3		- Direct sales - Ads linked to contents/service - Ads not linked to contents/service - Exploitation of personal data - 2nd degree discrimination	L'humanité, historia, le canard enchainé, ça m'intéresse, agoravox, google news, afp, newsweb, automoto, géo, goal, le monde diplomatique
cluster C&F4	- Direct sales - Exploitation of personal data - Package sales - Subscription - 2nd degree discrimination		Le figaro, relay, libération, la croix, le point, le nouvel observateur, ouest-France, la voix du nord, alternatives éco, capital, agefi, boursier, challenge, la montagne, l'est républicain, l'express, terraeconomica
cluster C&F5	- Ads not linked to contents	- Direct sales - Package sales - Subscription - 2nd degree discrimination	L'expansion, le journal du net, l'internaute, infosciences, la vie financière, au féminin, l'étudiant, evene, eurosport, plein champ, actuenvironnement, 20minutes, argusauto, coté maison, culture femme, football, journalauto, largeur, l'autojournal, le magne, l'entreprise, lire, Nice matin, sport24, technoscience, télerama, votre argent

Table 5: Comparison of clusters

« Commercialisation & Financing »

« Services & Rights »	cluster C&F1	cluster C&F2	cluster C&F3	cluster C&F4	cluster C&F5
cluster S&R1				G4 (1 case)	
cluster S&R2			G3 (1 case)		
cluster S&R3		G2 (1case)			
cluster S&R4			G3 (9 cases)	G4 (3 cases)	G5 (17 cases)
cluster S&R5	G1 (1 case)	G2 (4 cases)	G3 (2 cases)	G4 (13 cases)	G5 (10 cases)

Table 6 : Combined classification of cases

	OVER REPRESENTED STRATEGIESS:	**UNDER REPRESENTED STRATEGIES:**	**CASES :**
cluster G1	- C&F1		Virginmega
cluster G2	- S&R3 - C&F2	- S&R4 - C&F5	Le monde, Les échos, La tribune, Le parisien
cluster G3	- C&F3	- S&R5 - C&F4 - C&F5	Historia, Le canard enchaîné, ça m'intéresse, googlenews, newsweb, automoto, géo, goal, le monde diplomatique
cluster G4	- S&R5 - C&F4	- S&R4 - C&F3 - C&F5	Le figaro, libération, la croix, le point, le nouvel observateur, ouest France, la voix du nord, capital, agefi, challenge, la montagne, l'est républicain, l'express
cluster G5	- C&F5	- C&F3 - C&F4	Au féminin, l'étudiant, evene, eurosport, pleinchamp, actuenvironnement, coté maison, culture femme, football, largeur, l'autojournal, le mague, lire, Nice matin, sport24, télerama, votreargent

Table 7: estimation results for the two parts model
(standard deviation of estimated coefficients are reported in brackets)

	Audience shares lower than the sample median	Audience shares higher than the sample median
Lagged audience share	1.1906*** (0.1665)	0.9213*** (0.0302)
distance to the centre of G1	0.0030 (4.7823)	0.0048 (0.0340)
distance to the centre of G2	0.0668 (1.4606)	0.0547 (0.1213)
distance to the centre of G3	-0.1325 (0.2603)	-0.1391** (0.0792)
distance to the centre of G4	0.0849 (0.3428)	0.0966* (0.0598)
distance to the centre of G5	-0.0525 (0.3169)	-0.0502 (0.0506)
σ^2	0.0252*** (0.0029)	
optimal γ	0.37	
Log-likelihood for optimal γ	-38.9029	
log-likelihood with restriction to the lower case	-113.4031	
log-likelihood with restriction to the upper case	-39.3588	

* coefficient significantly different from zero at 15%
** coefficient significantly different from zero at 10%
*** coefficient significantly different from zero at 5%

Table 8: estimation results for the single part model
(standard deviation of estimated coefficients are reported in brackets)

	With heterocedasticity	With homocedasticity
Lagged audience share	0.9665*** (0.0263)	0.9676*** (0.0261)
distance to the centre of G1	-0.0017 (0.0228)	-0.0034 (0.0222)
distance to the centre of G2	0.0377 (0.0968)	0.0297 (0.0942)
distance to the centre of G3	-0.1109** (0.0605)	-0.1098** (0.0604)
distance to the centre of G4	0.0796** (0.0409)	0.0748** (0.0400)
distance to the centre of G5	-0.0690*** (0.0330)	-0.0735*** (0.0330)
σ^2	0.0398*** (0.0036)	0.0250*** (0.0023)
optimal γ	0.62	
Log-likelihood for optimal γ	-43.7591	
log-likelihood with restriction to homocedasticity ($\gamma = 0$)	-6395.4	
Log-likelihood		-43.6003

* coefficient significantly different from zero at 15%
** coefficient significantly different from zero at 10%
*** coefficient significantly different from zero at 5%