Impact of Trade on the Characteristics of the Digital Newspaper Market

New Empirical Evidence from Francophone Africa

Anaïs Galdin

May 22th, 2017

Abstract

This paper provides a first empirical study of how openness to international trade in news products, a phenomenon largely facilitated by the digitalization of newspapers, may impact the characteristics of the information products available to consumers. Based on a case study of the entry of French media firms in Francophone Africa, we build a new dataset of more than 800 000 newspaper articles over the period 2005 to 2017. Using text mining techniques, we construct a five-dimensional set of newspaper characteristics to qualitatively analyze newspaper data. Our evidences suggest that the entry of French digital newspapers, which produce a relatively low share of local news and whose websites are generally uniformly targeting all Francophone African countries, lead to a significant decrease of the diversity of subjects treated by Francophone African newspapers, characterized by a strong and significant increase of the share of local news, and a small decrease of the total number of articles related to France. We also find a slight but significant impact of the penetration of French digital newspapers on two indicators of newspaper format, namely a small increase of the average frequency of publication of Francophone African newspapers, and a very small decrease of the mean wordcount per articles. We believe the specific format of the news displayed on the web, generally significantly shorter than printed articles and with a higher publication rate, could explain the weakness of the effect on these two indicators.[1]

[1]I would like to thank my advisors, Julia CAGE and Thomas CHANEY, for their time, constructive discussions and invaluable advice. I would also like to thank the many French media firms who provided me priceless circulation data and greatly helped me during data collection stage. This data is not directly displayed in the present research due to confidentiality issues. I am also grateful to Thilo KLEIN from the Statistics Directorate of the OECD, who first introduced me to text mining and web scraping methods. All remaining errors are mine.

Contents

1 Introduction and Literature Review

While almost any possible manufacturing sector has been studied and many theoretical developments have been made in the field of International Trade, patterns and consequences of trade in information services are still largely understudied. The recent explosion of information services, and the many difficulties due to measurements and quantification of trade flows, certainly explain why the field is still in its infancy. Many new developments, as the utilization of barcode or credit card data, and the rise of machine learning and text mining techniques, are likely to reverse the trend in the coming years. Among all information services, trade in media products has a fairly short literature, mainly focused on films in western countries and on the cultural hegemony of the United States. Most papers are however theoretical, trying to explain what triggered trade in media products through either the home market effect or the BBAM[2] framework. Few papers focused on newspapers, and none, to our knowledge, tried to assess how trade in newspaper products could affect the very characteristics of the information produced. The rise of the digitalization of information, and the development of data mining, however offer interesting tools to analyze newspaper format and content across borders.

The present paper provides a first empirical study of how openness to trade may impact the characteristics of the information products available to consumers on the digital newspapers market. Using online newspapers data especially allows to provide qualitative analysis of articles content on a large scale, by automatizing the analysis through the use of text mining techniques. Rather than trying to explain the factors driving trade in information products, our focus is mainly to understand its implication for the characteristics of these goods.

The newspaper market provides an ideal environment for assessing the impact of a decrease in trade barriers.

Firstly, as evidenced by printed newspapers, these media traditionally circulate in local markets (geographic distinctness has long been a convenient feature of media markets). But new technologies and the explosion of digital news are eroding the distinctness of local markets, with the main consequence that local products are declining in most media markets. Chyi and Sylvie (2001) were the first (and some of the only) who turned to digital newspapers. Their paper first stresses that the medium's global capacity of online newspapers enables them to serve multiple audiences within multiple geographic areas. They especially draw the distinction between an online newspaper's local market and its long-distance market. Long-distance market is defined as the market where the printed edition, a substitute of the online newspaper, is not easily available.

Secondly, the characteristics of newspapers (in terms of quality and diversity) are important for welfare and have significant political effects (as documented by Gentzkow and Shapiro, 2004 or Della Vigna and Kaplan, 2007).

The true interest in integrating a model of media economics into a model of international trade is to capture the specificity of media markets compared to other type of tradable goods. We believe that trying to explain trade in media products based on home markets

[2]Budget-Based Advantage Model

advantages is neither new nor informative. Understanding the changes induced by a decrease in trade barriers requires to derive the determinants of the demand for specific newspapers characteristics. This could be done by studying not only the supply side (i.e the production and export of media products, with data on prices and quantities), but the demand side and what triggers consumers' tastes and preferences for a given type of information. Inherently, the reasons for which we consume information would explain why some types of news media are advantaged in trading, newspapers content being often created with the will to reach global audiences. This trend is even more likely to be reinforced by the important financing issues on the media markets, which would push media firms to take advantages of large scale economies allowed by the online world in order to realize cost efficiencies (an idea that relies on the most basic models of monopolistic competition, as Dixit Stiglitz, 1977). Gabszewicz and Anderson (2005) also find for instance that a strong advertising demand may lead to little media product differentiation. This is supposed to lead to greater quantities but fewer diversity of goods, and could explain reallocation of trade towards products from more commercialized market (building on Waterman, 1989).

Online news thus present several specific characteristics which are interesting to study from a trade perspective. Firstly, they are commercial public goods (i.e non-rival and non-excludable, with no direct cost for consumers to acquire the good once we accounted for the sunk cost of acquiring a receiving device). Secondly, media content has a high production cost, which allows price paid for distribution rights to dramatically vary across markets. Eventually, the average cost decreases sharply in output. Where these information services are provided in a specific language or cultural context, large markets may therefore be optimal for global production (Krugman 1980, Helpman-Krugman 1985).

Hence, we need to distinguish online news, and more generally information services, from manufacturing. In particular, fixed trade barriers for information services are dissimilar to those for manufacturing. They include for instance developing a marketing strategy to attract consumers and negotiating contracts over the delivery of intellectual property. Once a marketing strategy and standard contract have been developed for one market, they may be easy to replicate in other markets, making the fixed costs of delivering information services primarily global in nature.

This paper thus belongs to the large media economics literature on newspapers and television industries, including seminal articles by Steiner (1957), Rosse (1967) Spence and Owen (1977), Gentzkow and Shapiro (2010) for media biases, or Fan (2013). Most papers of this literature also belong to the field of Industrial Organization, and study the impact of mergers and acquisitions on media markets. In particular, Berry and Waldfogel (2000) show that consolidation of a media sector should increase the absolute number of media varieties, by discouraging multi-products firms to produce medias that compete with each other. Fan (2013) studies the effect of consolidation in the U.S daily newspaper market. She finds that consolidation of the sector leads to an important decrease in the news content quality, content variety and local news ratio, while strongly decreasing reader surplus and increasing publisher surplus.

This paper also enters into the literature on endogenous product choice, which presents important computational challenges (see Crawford and Shum (2006), Mazzeo (2002) or Fan (2013)). Products characteristics are used in this literature as explanatory variables

of the consumer's demand,in structural models. One of the issue with endogenous product characteristics is that the most commonly used instrumental variables for estimating demand with differentiated products are no longer valid. In order to solve this problem, Villas-Boas (2007) and Fan (2013) use demographics in non-overlapping media markets as instruments for media characteristics in overlapping markets. This exploits a key feature of the newspaper industry, namely that the circulation area of printed newspaper only partially overlaps with other newspaper circulation areas. Despite the fact this finding is obviously less binding in the case of online newspaper articles, we expect the underlying idea to hold for digital news, due to the presence of important barriers as language, religions, customs and more generally cultural proximity between areas.

While the International Trade and Media Economics frameworks are still not very integrated, a few seminal papers exist. Rosse (1967) investigates in particular the reason for geographical segmentation among newspaper firms and find that scale economies are the most important determinant of geographical monopoly in this industry. Scale economies would especially be able to overrun the force of consumers' heterogeneous preferences. Siwek and Furchtgott-Roth (1993) support that domestic market size and scale advantages, that were critical to explain the success of U.S movies and television programs in international trade, also allow to explain U.S firms' success in the global software market. Eventually, Frijters and Velamuri (2010) show that declining barriers to entry may create a disincentive for professional media to invest in the produciton of high-quality news.

New information technologies drastically lowered the costs of entry into new and potentially foreign media markets, where online news is supplied as a commercial public good. The distribution system of online publishing therefore leads to a specific type of scale economies, with number of users unrelated to production costs. Variable costs as cost of labor and technical skills remain, but the cost of establishing a broadcasting platform substantially decreases while potential revenue increases. The online world thus offers the possibility to expand the scale of production beyond the printed edition market, to new, foreign markets. On the positive side, reduced industry concentration in the online newspaper markets may allow for the development of media market niches, with the advantage to respond to a greater heterogeneity of tastes. But lower barriers to entry may actually reduce the quality of news available to customers, by decreasing the revenue generated through advertising and consumers' subscriptions, with potentially important consequences for consumers' welfare. A recent NBER working paper by Shiller, Waldfogel and Ryan (January 2017), entitled ""Will Ad Blocking Break the Internet?" especially provides evidences that preventing advertising contents can actually reduce consumers's welfare by lowering the quality of the delivered goods. Several earlier studies provided more general evidence that free entry into differentiated product markets with scale economics and decreasing average costs can result in too many products but too few varieties (see Spence (1976), Dixit and Stiglitz (1977), Lancaster (1979) and Mankiw and Whinston (1984)).

This paper more largely contributes to the literature of trade in cultural goods, which is significantly larger than the one of trade in media products and provides several interesting insights. The most important feature of this literature is that cultural goods are discounted in their raw worth when traveling to a culturally different market. The greater the cultural

distance between the exporting and importing markets, the greater is this loss in value. Local demand for imports in a given cultural product is reduced if values, beliefs, and styles presented in these goods are less locally accepted. Seminal papers in this field are Marvasti and Canterbery (2005), Disdier, Mayer et al. (2010) or Masood (2013). Masood (2013) especially studies diversification, reconcentration and domination due to cultural trade. She focuses on the impact of economic development on the cultural diversity of developing economies, and proves that market's dynamics are in favor of the emergence of a "westernization process". She also finds that an increase in income leads to a diversification of varieties consumed.

Models used in the literature of trade in media products also proved that cultural differences further aggravates the trade imbalance which was driven by disparities in home market size per se (Hoskins and Mirus, 1988; Waterman, 1988; Wildman and Siwek, 1988). Producers from a large domestic market and, hence, with a strong capital power, are better able to offset cultural depreciation in infiltrating markets abroad. As a result, the gain in overseas audiences boosts their investment capability, which in turn snowballs the market-penetrating power. In contrast, small market producers not only are more hampered by cultural barriers in disseminating offshore, but cede home market shares to imported products. It follows that the received impact of cultural distance on the exportability of large and small market media products is asymmetric (see Wildman and Siwek, 1988; Waterman, 2005).

Audiovisual media products are thus cultural goods specific to the values, beliefs, styles, and the like of the target populations. Most consumers intrinsically prefer productions that are fabricated from similar or familiar backgrounds, and that are thus relatively easier to absorb. Consequently, content goods are discounted in their raw worth when traveling to a culturally different market (Cantor and Cantor, 1986; Hoskins and Mirus, 1988; Straubhaar, 1991 or Waterman, 1988). The key of our empirical strategy is the change from printed to online newspapers, who brought the transaction costs of producing news for long-distance markets closed to zero. However, these findings stress that language and culture may affect consumer demand and the cost of marketing goods across borders. The physical cost of delivering online news may be low,but the barriers consumers face in gaining access to them may not be.

The present paper provides a first empirical study of changes in online news product characteristics following an opportunity for firms to trade their media products beyond their initial local market. Our focus is not on the producers of information, but on the receivers, in order to assess what could drive demand for foreign news. In particular, we try to understand whether and how this foreign production could impact the characteristics of the domestic media markets. Do newspaper publishers in importing markets improve or deteriorate the content quality? Do they enlarge or shrink the local news ratio, and do they increase or decrease content variety? We may especially wonder whether the new media products have the capacities to find their places in foreign countries, and how these foreign media markets may be impacted by new information whose content and format may be quite different from the information domestically produced.

We answer these questions through a case study focusing on trade between French media firms (the exporters) and 21 Francophone African countries (the importing region). Relative similarities in languages, legal and administrative environments, and colonial linkages indeed reduce the costs of trading cultural goods for French media firms compared to other regions

in the world. The last few years have witnessed the emergence of one-way trade in media products from France to Francophone Africa, with the creation by French players of digital newspaper targeting Francophone speakers in African countries. Among other, *Le Point Afrique*, *Le Monde Afrique*, *RFI*, *France 24* or *La Croix* entered the market as early as 2014. The quick decline of printed newspapers, the fast digitalization of news and the low growth rate of French media markets are many factors who explained the interest of French media players to get closer to the Francophone African audience. The same phenomenon occured in the United Kingdom (with famous examples as *The Guardian)*, so that the case study could eventually be extended.

This research contributes to the literature on many aspects.

Firstly, the qualitative dimension of its empirics, using big data methods for textual analysis, bridges current data gaps in international trade. The main novelty of this paper for the trade literature is the use of web scraping, text mining and keywords analytics to automatically classify and qualitatively analyze media contents. 831 014 unique newspaper articles dated from 2005 to March 2017 were thereby collected for the purpose of this paper. On the other hand, a trade analysis of media economics permits broader theoretical development in a field that relies primarily today on micro concepts and Industrial Organization. The current understanding of economic forces driving trade in information products is far for complete, and improving this understanding requires overcoming both theoretical and empirical challenges, that involves performing more investigative analysis. Moreover, adopting a trade perspective seems particularly relevant as one of the specificity of the information market is the recent and sharp decrease in trade barriers, largely due to digitalization.

Secondly, most economic researches in media economics focus on the U.S market, and to date, all empirical examinations of media flows have concentrated on U.S. exports. This paper tries to extend the scope and to build a larger scale analysis of market dynamics. The present case study thus incorporates data on exchanges between France and Francophone Africa. This market has witnessed a recent increase of media consolidation and one-way trade from France to Francophone Africa, with the creation of "African" division of French newspapers. Analyzing international trade integration is often challenging in media economics due to the importance of regulatory structures, common languages and accounting practices. Colonial history here facilitates the analysis of these determinants. Moreover, providing cost-efficient news in a long-distance market easily translates to market a *less-than-unique* local product to an audience that would be likely to find it unique, maybe because it cannot easily access substitute goods through other channels than the Internet. It would therefore be relatively easier to please long-distance audiences when there is a lack of media competition in these foreign market.

Such qualitative analysis is currently limited in scope to newspaper media, but the recent development of tools to analyze radio content (as the United Nations *Global Impulse Lab Kampala Project*) opens up opportunities for further researches.

Section 1 provided an introduction and literature review. Section 2 details some background to the entry of French media players in Francophone Africa. Section 3 describes the data, explains the methodology used to collect and analyze text data, and provides definitions of the main indicators of newspaper characteristics that we built for this analysis. Section 4

sets the basis of our empirical framework, describing the model used for estimations. Section 5 provides baseline estimates of the empirical model. Section 6 contains the core estimation results, tackling in particular the issue of endogeneity of the explanatory variable of market penetration by French newspapers. Section 7 concludes and discusses directions for future works.

2 Context of Entry of the French Digital Press in Francophone Africa

"Everybody says it's the new Eldorado, South America of tomorrow. Even if we do not know when." [3]

This section puts in perspective the entry of French media players in Francophone Africa and underlines what motivated the present paper. We provide an overview of what triggered the entry of foreign media, document its recent developments, and specify what characterizes the implementation strategies of French digital newspapers in Africa.

2.1 Population Boom in Francophone Africa

Africa is the largest Francophone region is the world, with thirty-one countries whose official primary or secondary language is French. The IOF (International Organisation of the Francophonie) estimates that the total number of French speakers in the world should move from 220 million in 2012 to 750 million in 2050, among which 90 % should be located in Africa. The organization foresees that 8% of the world's population could therefore speak French in 2050 (against 3 % today), which would place French as the most widely spoken language in the world, before English[4]. This population boom is mainly triggered by West Africa, which is widely Francophone and has a much higher birth rate than the rest of the continent, with 7.6 children per woman in average. Some Francophone countries as Burkina Faso or Niger should see their population multiplied by 2.5 in 2050. Niger counts 700 000 people more every year, and Mali 500 000[5]. This population is moreover extremely young. More than 44% of the population is less than 15 years old (against 16% in Europe). Rejuvenation of the continent,by increasing the percentage of young, hyper-connected consumers, is a bargain for French and other foreign digital media.

Both this explosion of a French-speaking audience and its rejuvenation are at the root of the one-way trade in online newspapers between France and Francophone Africa.

2.2 Digitalization of Africa

Africa indeed took a big step in its digital transformation by almost skipping the step of fixed broadband subscriptions to adopt a mobile digital consumption. More than half of the internet connection in Africa would be provided over a mobile phone according to the International

[3] Yves Bigot, CEO of TV5 Monde, one of the main French media players exporting in Francophone Africa, during an interview given to the magazine *Strategies*, on June 6, 2014.

[4] While spoken by 8 of the world's population today, English is anticipated to account for only 5% in 2050, mainly due to the expansion of African and Asian populations

[5] Source: IOF

Telecommunication Union. Since 2010, the number of mobile cellular subscriptions grew from 187% in average every year. Households spend in average across all Africa 10% of their monthly income in mobile cellulars (against only 3% in developed countries according to the INSEE). The number of fixed secured internet servers is very low, so that one often talks about a *mobile cellular boom* or a *qualitative numerical leap*. This could find several interpretations. Firstly, because the major part of the population is between 15 and 25 years old, and these youth are highly attracted by new technologies. Then, because African populations are very mobile. Eventually, because recurring power outages are less compelling for mobile users. Nevertheless, the internet mobile sector is far from being mature. 3G would account for less than 20% of mobile phone connections and the strong penetration rate of mobile phones is mainly due to multiple subscription by a small proportion of the population.

This mobile revolution starts in 2010, a time when began the privatization of the New Information and Communication Technologies (NICT) sectors. From 2010, the number of internet users increased seven time faster in average than for the rest of the world. Growth of internet users between 2000 and 2012 was 3600% and raised Africa to 167 million users, according to World Statistics. Despite this rapid changes, Africa still stands behind Europe, with an internet penetration rate below 11% across the continent (when the European average is 23%). These figures moreover hide substantial internal disparities and connectivity is far from being ubiquitous on the continent. Internet coverage may vary between 90% in cities to less than 10% in rural areas within a country.[6].

These disparities between rural and urban areas, but also between coastal zones (where are located wireline networks) and inland are due to the absence of the "last kilometers" of network facilities, which are supposed to allow the physical connection to the sources bringing internet access. The development of this facility is a major challenge for the whole continent, but private telecommunication companies on the continent do not see any economical advantage in building these last kilometers of lines in the short-run, and implementation still needs support from local government.

The price of internet is also one of the major barrier to the expansion of fixed internet. The access price to bandwith is still extremely high. For instance, the price of an unlimited ADSL connection in 128 kbit/s in Togo is 45000 CFA per month (70 euros / month), what represents more than one month's salary for the average Togolose employee (mean salary is 42$/month). The unlimited professional 2 Mega access is offered from 960000 CFA, or 1500 euros per month. These prices would be mainly due to the fact a very low proportion of internet traffics are oriented towards Africa. 70%of web traffics are currently directed to foreign markets, contrary to developed countries which display in average 60% of national web traffic. [7]

2.3 Entry of French Media Firms

Most French media still limit distribution of their products inside the hexagon. However, French media markets have a low growth rate (approximately 1% per year), being geographicaly limited. On the contrary, the investment bank Natixis estimated in a study released in

[6]Source: ITU World Telecommunication / ICT indicators database (International Telecommunication Union, UN agency)

[7]Source: TV, Radio, Mobile, Internet : decryptage de la conso media en Afrique, *Africascope*, 2015

September 11th, 2013[8] that the potential growth of the African media markets should be very strong, around +3,2% per year.

This potential new market, in addition of being very dynamic, can be accessed at low physical cost. The digitalization of information allows french media to easily distribute online news from their parent company, without any tariffs or transportation costs. Many French medias saw the opportunity and started entering the market as early as 2011. Radios and television broadcasting groups were among the first to start the expansion, and main French media groups as Vivendi, Lagardere, Canal Plus (via *A+*), TF1 and M6 created programs targeting Francophone Africa.

Many newspapers created or extended digital news platforms dedicated to Francophone Africa. Among them, major players as *Le Point Afrique* in 2014 or *Le Monde Afrique* (by the end of 2013-beginning of 2014), *Le Huffington Post* (which depends of *Le Monde*), la Croix (through *Urbi and Orbi Africa*) in 2015, *RFI* and *France24* , who launched vertically integrated companies in 2015, or *Slate Afrique*. The European news providers *Euronews* also launched a media project for francophone Africa in 2015, called *Africanews*. Some panafrican newspapers, based in Paris, as Jeune Afrique or Afrikinews, also developed online content for Africa and for the diaspora in France.

Our major interest is to understand the diverse strategies of these groups, and especially what kind of content they intend to develop when entering the markets. The motivation is to see whether these new media may impact the characteristics of the Francophone newspapers landscape across African countries, and in what sense.

2.4 Implementation Strategies

We describe here two complementary and instructive examples before starting the analysis.

Slate Afrique has been a precursor. *Slate* is a pure player, entirely free, operating exclusively online, and entirely financed through digital advertising. The African version has been the first launched, in 2011. It expanded quickly across the continent, by recruiting many local correspondents in order to be present in every country, to better cover local news and to produce quality content. However, in 2013, Slate Afrique was still not cost-effective. The main observation is that it was launched too early, reproducing almost identically in Africa the French format and its business model. As a consequence, in 2013, the management drastically reduced the number of journalists and changed its editorial strategy: less original content was produced and a greater share of news resulted from the aggregation of French news. Since this new orientation strategy has been adopted, French IP adresses became more important and African connections decreased.

Not taking into account the complexities of the African continent, its diversity, divergences in cultures and the fact the digital transition takes time thus lead to a partial failure of the business started by *Slate Afrique*. Its main mistake was probably due to its unsustainable business model, relying entirely on digital advertising while the digital advertising market was in its infancy in 2011- and is still today. The example of the African website *maliactu.net*, which has a monthly turnover of 8000 euros despite being one of the most important in Mali,

[8] On the basis of previsions by the World Bank and the IOF

is striking. The analysis of the digital advertising markets in Africa is severely constrained by the lack of reliable figures. Most countries do not have centralized tools to collect data nationally, and no institution is in charge of collecting data for digital environments. The only data currently available are Facebook or Twitter data, which are still the best proxies to evaluate a brand awareness. Consequently, few investors decided to blindly enter the market.

Le Monde first wished to develop a printed newspaper targeting Central Africa and Morocco, before turning to a purely digital version by the end of 2013 and creating *Le Monde Afrique*. Entirely hosted on the website of the parent company, the African version keeps some autonomy in terms of financing and editorial line. The group has partnerships with *Media 24* in Morocco, *Algerie Focus* and *Burkina24*. Institutional partners are very present in implementing the project in Africa, provide 25% of funding and are engaged for three years. The goal is to go from an entirely free version to a two-speed version, following the current model of *Le Monde*: providing many articles for free and more investigative journalism through paid membership. This still seems a challenge today. The site publicly releases statistics, and has currently 4 to 5 million monthly visits and 10 million page views, which is quite low. They target much higher figures by the end of 2018. But consumers mainly come from France (between 60 and 70% of total traffic) and only 15 to 20% of internet users are from Africa. Its director, Serge Michel, however stresses an interesting pattern. Since the creation of*Le Monde Afrique*, the media group *Le Monde* has reached a new audience, who was very marginal beforehand: African diasporas, in Europe or abroad. Thus, an important number of *Le Monde Afrique* readers also consume *Le Monde*.

2.5 Concluding Observations

These examples seem to prove two main things. Firstly, one should avoid the trap of the "soft power" derived from colonialist perspectives, let alone considering the African continent as a group of uniform cultures. They necessarily have to adapt to expectations of these new audiences if they want a breakthrough. Secondly, business models of these media cannot be simple "copy-paste" of strategies adopted on their national territory. Some french media firms took the bet of diversifying and to partner with local stakeholders, as did the group Lagardere. But other medias acted as if Francophone Africa was an uniform entity to which they could distribute content without cultural distinction. In 2014, Yves Bigot from *TV5 Monde* announced he wanted to reinforce in *TV5 Monde Afrique* the share of programs specifically dedicated to Africa, which was then only 20% to 25% of the total programmation. *Canal Plus* still currently broadcasts the same panel of programs in all Francophone African countries where it is implemented.

But will an Ivorian be as attracted by a program taking place in Senegal than by a local program?The same kind of question is applicable to digital newspapers. To what extent do the French media firms diversified their content? What proportion of news would only be a low-cost replication of French cultural content? And how does the entry of potentially new types of content from France could impact the characteristics of newspapers domestically produced across Francophone Africa? We try to provide first answers to these many questions in the present paper.

3 Data

This section provides an overview of our sample, lists the different data sources used, summarizes the methodology and data cleaning decisions we have carried out, and defines the major variables created for this analysis.

3.1 Sample of Countries

This paper defines Francophone Africa based on the classification established by the International Organisation of the Francophonie (IOF). The IOF suggest there are 21 African countries for which French is a primary official language, and 4 countries from North Africa in which French is widely spoken. More specifically, there are:

- 11 Countries for which French is the unique official language (West and Central Africa);
- 6 Countries for which French is the co-official language (East and Central Africa);
- 4 countries where French is widely spoken (North Africa)

The share of French speakers in these countries nevertheless widely varies (going from 17% to 67% of the total population). We take into account these divergences across countries in our econometric analysis.

For the purpose of this paper, we focus on 21 Francophone countries for which we collected enough online newspapers data (excluding Comoros, Djibouti, Equatorial Guinea and Seychelles from the above list of 25 countries). This sample corresponds roughly to 115 millions of people .

3.2 Building a New Text Dataset : Definitions and Methodology

One of the main empirical contribution of this paper is the building of a brand new database, which includes digital newspaper articles from Francophone Africa, relying mainly on web scraping and text mining techniques.

Collection of Newspaper Articles

We first built different extensions of R and Python scripts, which allowed us to crawl and parse HTML articles and their metadata based on URLs of websites' main domains. These scripts had to be slightly adapted to match each different version of newspaper websites.

We started using these algorithms before realizing the difficulty of scraping historical data from Francophone African websites: most of them do not store old articles. Some only have online resources for the last month, other for the last week. A couple of websites store articles over the past year. Older articles data may actually be available even if not displayed on the websites, by using deep crawling techniques (and a certain number of tech startups offer today these services), but we did not have sufficient resources for carrying such kind of analysis. A solution for extending our reach was then to rely on news aggregator websites as well (mainly Factiva and Proquest). This solution reduces our initial newspapers sample size (because only main newspapers are registered by news aggregators), but increases our time

series dimension, which is necessary to assess the potential effects of an entry shock of French newspapers in African markets.

For each of the 21 Francophone African countries of interest, we thus targeted the main newspapers whose articles were both available directly on their websites (for the most recent articles), and on news aggregator platforms for the oldest. Main newspaper websites are defined as those providing general news, with circulation at the national level (in opposition with smaller, local newspapers), which are often available in printed format and whose printed national circulation (when known) is also supposed to be among the highest in the country. We tried to rely at least, when feasible, on the five main national newspaper websites.

We assume studying at least the newspapers with the largest circulation figures provides a representative overview of the newspaper characteristics in a given country. While this could seem like a strong assumption, it also has the advantage to make these newspapers comparable across countries Aggregating all kinds of newspapers could actually biased our estimates, because we could for instance collect a disproportionate amount of sport newspapers or tabloids for some countries, whose characteristics are obviously very different from General News websites. These main national newspapers are also more comparable with their new French newspaper competitors, which are for most of them produced by the main national newspaper firms in France. A precise list of newspapers used in this paper, decomposed per countries and regions, is available in Annexe C.1.

The original database of African newspapers we build includes among other, for each newspaper, the main domain, URLs, RSS feeds and page ranks. It also includes information about whether the newspaper exists in printed version, whether its scale of production is national or local, frequency of publication for the printed version, and eh main category of news provided (General News, Sports, Business-Eco, Tabloids...). The database was mainly built based on information provided by the sites *Onlinenewspapers*, *abyznewslinks.com.* and *w3newspapers.com*[9] , which are three websites recording media by countries (digital and printednewspapers, but also audiovisual media). We also derived more precise information about each newspapers by individually consuting each website. A more exhaustive list [10] of the online newspaper landscape for each Francophone African country has been collected for the purpose of this research, and is available upon request.

Text Data Analysis

We then built a parsing algorithm, which allows to read each HTML file in R, extract its main components and the relevant metadata, and convert it as a corpus. For each article, the algorithm gets for instance the main text, the headline, the date and time of publication and the author, if available). It also provides a measure of wordcount. Once the data grouped in a corpus, we apply text mining techniques to identify keywords frequencies. The keyword analysis algorithm relies on traditional pre-processing procedures: we use stopwords to remove extremely common words, and words stemming to reduce words to their linguistic roots,

[9]See http://www.onlinenewspapers.com/africa-newspapers.htm , http://www.abyznewslinks.com/afric.htm and https://www.w3newspapers.com

[10]Bearing in mind that the exhaustiveness of the list is always limited by the information available online, and taking into account the fast evolution of online contents. The latest version of this list is dated from February 2017.

so that all sentences referring to the same topics are considered as equivalent by our algorithm.

Some articles collected from news aggregator websites already have information about the main topics treated in the text and the geographic area of interest. We build on these categories to create a new taxonomy of newspaper topics and geographical areas, derived from Dow Jones intelligent Indexing (which provides hundreds of news categories). Based on the relative frequency of each topics within our sample of newspapers, we refine these categories into eleven groups described in Table 1. Other classifications of media topics exist and are roughly similar to ours, as the one provided by the International Press Telecommunications Council (IPTC). Our script then identifies the most frequent keywords within each article based on these lists of topics, and automatically classify media contents. The relative frequency of the occurrence of each topic is eventually computed across time, across newspaper firms and across countries. A comprehensive description of the distribution of topics per country (in terms of absolute frequency and percentage) is provided in Table C.1 in Annexes.

Table 1: **Taxonomy of Newspaper Topics**

Categories	
Arts/entertainment	International relations / global issues
Corporate/ industrial news	Risk news and disasters
Crime/legal action	Sciences/humanities and technologies
Domestic politics	Society/community and lifestyle
Financial and economic news	Sports
Health	

For the purpose of the present study, **831 014 unique newspaper articles** were analyzed, for 22 countries, France included. Among them, **734 834 unique articles are for 21 francophone African countries**. All regressions use monthly panel data, with panel clusters at the country level (**762** unique monthly observations since 2014; **1268** since 2005).

French newspaper articles are only used for comparison purpose (keywords analytics, share of news referring to France and other newspaper characteristics). These articles are not included in our regression before Section 6, so that the evolution of the characteristics of newspapers in each African country following the increase of the penetration of French media firms is purely analyzed from the point of view of domestically produced content. This is particularly important for instance to analyze the evolution of the percentage of articles referring to France (French news websites targeting Africa normally do not contain news referring to France, and including these articles for each African country would bias the analysis of the evolution of African newspapers).

3.3 Additional Data

In order to empirically estimate the impact of the penetration of French newspaper firms on media characteristics in Francophone Africa, we compile this new qualitative dataset of newspapers characteristics with several additional data sources. These many different controls variables address a range of possible confounding errors, controlling especially for unobserved

newspapers characteristics or market-level tastes.

Demographics

Country demographics used in this paper are mainly extracted from the *World Development Indicators* of The World Bank and include: percentage of francophone population, adult literacy rate (percentage for both sexes, population of age 15 and above), adjusted net national income per capita, GDP per capita (current USD), educational attainment, population median income, population median age and urbanization (percentage of urban population), all of which positively affect the demand for newspapers in the economic literature, excepted median income. Educational attainment for population of age 25 and above is controlled at two levels: (i) percentage of people having at least completed lower secondary education, and (ii) percentage of people having at least completed upper secondary education. These two indicators are those displaying the greater variances across countries in our sample, and those supposed to better reflect basic educations levels.

Bilateral Trade Data

Bilateral distances between each Francophone African country and France are calculated as the sum of the distances between the biggest cities of both countries, weighted by the share of the population living in each city. Data for these variables are extracted from the *CEPII GeoDist database*[11]. Variables related to cultural distances, as common official language and colonial dummies, are also available from the CEPII, which makes available a "square" gravity dataset for all world pairs of countries, for the period 1948 to 2015[12]. We use this data as proxies for measuring more precisely bilateral trade barriers. Data on independence dates come from Head, Mayer and Ries (2010) and the *CIA World Factbook*. Data on entry costs and time required to start a business come from the *World Development Indicators* provided by the World Bank .

Internet Penetration

Several sets of data have been used to control for audiences reach. Mobile cellular subscriptions, fixed broadband subscriptions, secure Internet servers and internet users (all per 100 person, since 1990) are extracted from *Internet World Stats* and the *World Bank database*. Figures on the absolute number of internet users, Facebook users, the relative rank of Wikipedia France for each francophone African country and the number of active social media users (in million) are from the *US Census Bureau*, *Internet World Stats, Facebook*, and the *GSMA Intelligence*. Data on protection on intellectual property rights is compiled using the Ginarte-Park patent protection index and the *Global Competitiveness Report* measure of the strength of IPR protection.

[11]Mayer, T. and Zignago, S. (2011) Notes on CEPII's distances measures : the GeoDist Database CEPII Working Paper 2011-25

[12]This dataset was originally generated by Keith Head, Thierry Mayer and John Ries to be used in the following paper: HEAD, K., T. MAYER AND J. RIES, 2010, "The erosion of colonial trade linkages after independence" *Journal of International Economics*, 81(1):1-14. (formerly CEPII discussion paper 2008-27)

Web Traffic Measures and Audiences Reach

Due to the need for advertisers' finance, media markets generally require good audience measurements and therefore provides good data on circulation for each products or on products characteristics. Reliable circulation data for printed newspapers are often available. For online news, the exact number of online users is more complicated to obtain, because no single reliable technique exists to accurately measure the identity and the size of the audience (see Greenstein, 1998; Gugel, 1997; Kirsner, 1997; Rich, 1997). This data could imperfectly be captured trough web scraping, using page ranks or "likes" and sharing of articles in Twitter or Facebook.

Most online newspaper firms are nevertheless able to provide statistics on "page views" (the number of requests for HTML pages, used by advertisers as a measure of online audience size) and "unique users" (the number of site visitors, often derived by cookies) statistics. Most newspaper firms are also able to provide the breakdown of users by local and long-distance markets.

We therefore directly collected data on French newspaper websites' audiences from French media firms. Most of these newspaper companies provided us audience statistics, with data geolocalization (i.e. total number of visits, total number of pages and total number of unique users per Francophone African country). Data provided are either disaggregated per day or per month. For the purpose of the present paper, we used monthly data, that is the lowest common denominator across all our newspapers data. A majority of French media firms also provided audiences data disaggregated by distribution channels (site webs, application, mobile web). Due to confidentiality issues, this data is aggregated at the country level in this paper, and we provide no statistics on individual newspaper firms.

The information collected in terms of the circulation of French online newspapers in each African country goes back to 2014, the date at which most of them officially entered the market. This date of entry however varies across newspaper firms.

Concerning online newspapers for each Francophone African country, official audiences data have been directly derived from websites when official Google Analytics measures are provided. Official circulation data are however missing for a couple of French media firms and for a significant proportion of African newspaper websites. Whenever this is the case, we inferred web traffic measures using online platforms which provide sites' traffic over time based on their URLs. We used mainly two sources: Similar Web and SEM Rush[13]. These websites provide data on total number of monthly visits, total unique visitors, total pages viewed, and average total minutes per visits for each website, based on their main domains. It also provide data about the relative weight of each country in the percentage of total traffic for a given website. The data can thus be split by country. Similar Web and SEM Rush data includes sites of all sizes. SEM Rush seems to provide the most accurate data of all the tools we examined, but the sample of newspaper websites available to us was more limited compared to Similar Web. We thus only used estimates from the latter platform in the present paper, but we cross-checked each site with SEM Rush in order to detect any abnormality.

However, the present study does not exclusively rely on web traffic measures provided by these websites, due to several weaknesses. Firstly, accuracy is questionable for all of them, and we can seriously cast doubts on the absolute reliability of any of these tools based

[13]Other similar platforms for web traffic measures exit but have not been used for this study due to limited access to their ressources (mainly Alexa, QuantCast and Comscore MMX-Platform)

on comparisons with official data. However, such data may be useful for cross-countries comparisons. Secondly, these platforms do not always provide statistics for the smaller sites (which was not an obstacle for this paper, but could be for a largest-scale research).

We therefore inferred circulation data for each newspaper websites in our sample based on the official circulation data we collected. The simplified step are the following:

(i) For every online newspapers, included the one for which we have official data, we collected Similar Web data.

(ii) Whenever we have the two type of data for a newspaper i, we compute the multiplier which allows to go from Similar Web data to the Official data. This coefficient is simply given by:

$$\varpi = \frac{\text{Similar Web circulation figures for newspaper i}}{\text{Official circulation figures for newspaper i}}$$

(iii) For all newspapers for which we do not have any official data, we then multiply the circulation measure provided by Similar web by this coefficient ϖ. This "inferred official rate" is not assumed to be the real one, but it allows us to have a more precise idea of the real circulation figures, based on the bias between the official rate and Similar Web rate. Above all, it allows to obtain comparable circulation figures across countries, and to determine which countries consume a relatively higher share of french online newspapers, and how this share evolves over time.

3.4 Definition of the Indicators of Newspaper Characteristics

We derive from our database of newspaper articles several indicators of newspaper characteristics for each national markets. For each of these countries, we compute monthly value of this newly created indicators, as we have circulation data at the month level. The final dataset used for regressions thus contains **762** unique monthly observations since 2014, and **1268** since 2005.

Our main empirical question is to assess the impact of the market penetration of French digital newspapers on the characteristics of online newspaper articles domestically produced in each Francophone African country c. These indicators of newspaper characteristics are thus used as dependent variables in our estimations, and our main regressor is the variable of market penetration of French newspapers for each pair of country-month (i.e. the total number of visits of French newspaper websites per population located in each Francophone African country c across our sample).

The different dependent variables of newspaper characteristics used in this paper are described in details below.

The dependent variable Y_{ct} is **a 5-dimensional matrix of newspaper characteristics**, composed of:

1. *Two variables that refer to the newspaper articles format*, and which have been used in the media literature as subcomponents of a newspaper quality index. These indicators are:

(a) **Wordcount per article**. We use in our estimation the log mean wordcount per article per month (for printed newspaper, the log of the number of pages is commonly used);

(b) **The log frequency of publication** (based on the number of articles released by month).

2. *Three variables that refer to the specific content of the articles released*:

(a) **A Local news ratio**, which captures the relative share of articles referring to domestic, local topics, and whose construction is described in detail below;

(b) **A French news ratio**, which captures the relative share of articles produced by a given Francophone African country c that refers to France

(c) **A measure of content variety**, which captures the diversity of topics treated by newspapers in a given country c each month.

Figures of the evolution of these five indicators of newspaper characteristics can be found in Annexe D.

Different compilation of similar indicators (excepted the French news ratio, that we purely introduced to see whether the entry of French newspapers had an influence on the number of topics related to France in African newspapers) has been used in the literature as indexes of newspaper quality[14]. However, these indicators are not computed as we did, as most of the literature interested in newspaper quality does not use qualitative text data.[15] Fan (2013)[16], who studies endogenous newspaper characteristics in the United States, uses for instance he percentage of local-news staff over the total number of staff as a proxy for local news, and define newspaper variety as $100[1 - \sum_i (\text{share of staff in section} i)^2]$, the sum part being analogous to the Herfindahl- Hirschman Index for industry concentration.

High quality news is assumed more costly to produce for a media firm, but also to attract more audience. To construct a real measure of newspaper's quality, we would ideally add data on the log of the number of staff and the number of reporters (following Berry and Waldfogel 2003). However, there is a real scarcity of data on the number of journalists in Africa. Many Francophone African countries do not officially release this information, and the data we collected for our sample of interest is very imprecise. Due to the importance of missing data for this indicator across countries, we decided not to incorporate it in the analysis.

Based on our sample of newspapers, we nevertheless compile some preliminary measure of the number of unique reporters for each country. This is done by automatically identifying in each article the name of the author (when provided). These estimates can be found in Annexes, Table C.1. However, an important number of articles does not contain this information (resulting in NA or very low number of journalists in the table). Therefore, this information must be taken carefully.

[14] See in particular Fan (2013) or Berry and Waldfogel (2003)

[15] For qualitative analysis of text data, see in particular Gentzkow and Shapiro (2010), "What drives media slant? Evidence from U.S. daily newspapers." *Econometrica*, 78, 35–71.

[16] Fan (2013), "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market", *American Economic Review*, 103(5)

Following Fan (2013), a more rigourous measure of news content quality should also depend on the news hole (non-advertising space). There is no direct data on the news hole, but it could be computed from the difference between the annual number of pages (or wordcount) and the annual advertising quantity. While Fan (2013) uses such an approach for media outlets in the U.S market, the latter data is not available for our case study targeting francophone Africa and onine news. This could be an interesting way to extend the analysis in the future, as some new measures of advertising quantity becomes available for websites. These measures are principally based on Optimal Character Recognition (OCR) techniques, that allows to detect for each article the proportion of space allocated to advertising, and are developing fast.

Measure of Content Variety

Content variety could be measured through diverse channels. Based on the Herfindahl-Hirschman Index (HHI), it could especially be proxied by a decreasing measure of concentration of news topics. We mentioned Fan (2013) works with a similar index, using as reference the number of reporters for each section of news. We also use here on the Herfindahl-Hirschman Index of market concentration, but using a qualitative analysis. More specifically, we rely on the relative proportion of news dedicated to each topic category, using the list of topics we created and presented in Table 1 (See also Figures B.1 and B.2 in Annexes for the relative distribution of each topic over time, aggregated over all African countries).

The Herfindahl-Hirschman Index (HHI) used as a measure of content variety is defined as:

$$H = \sum_{j=1}^{N} s_j^2$$

where s_j is the relative share of topic j, and N is the total number of online news articles. This Herfindahl index thus ranges from $1/N$ to 1.

We rather use the Normalized Herfindahl Index, that ranges from 0 to 1 and allows to better take into account the total number of articles released in a given country over a month. It also allows to control for our arbitrary number of topics' categories. For instance, we can have 10 articles and 10 different topics in a month for a given newspaper, and the following month 1000 articles and 11 different topics (the maximum number of topics' categories created), very evenly distributed.

This normalized index is computed as:

$$\begin{cases} H^* = \dfrac{H - 1/N}{1 - 1/N} \quad \text{for} \quad N > 1 \\ \\ H^* = 1 \quad \text{for} \quad N \quad = 1 \end{cases}$$

where H is the normal Herfindahl Index, and N is the total number of articles in the market.

Local news ratio

A second qualitative index used in our matrix of newspaper characteristics is a local ratio, which allows to capture readers-taste for local news. We define local news as news referring to domestic, local areas for each country, at the regional or city level :

$$\text{Local ratio (monthly)} = \frac{\text{Local references}_{ct}}{\text{Total articles}_{ct}}$$

where Local references$_{ct}$ is the total number of articles whose topic refers to local news for country c in month t, and Total articles$_{ct}$ is the total number of articles released in country c in month t.

For each article in our sample, we have several geographic identifiers. If a local keyword identifier appears once, we consider the article refers to local topics. However, one should be careful as we want to measure the proportion of *domestically* produced articles in country c referring to local areas *within* country c. For instance, it can happen that a newspaper produced in Ivory Coast releases an article about Abidjan, its capital. We would consider this article as a "local news". But we can also have articles from Egypt or Mali referring to Abidjan, which we should not consider as local news as the locality does not directly refer to the country producing the news.

In order to prevent this kind of mistake, we associate to each newspaper its country identifier. We then associate to the full list of geographical identifiers, for every country an region in the world, two geographical indexes, described in Table 2:

Table 2: **Identifiers for computing ratio of local news**

Geographical Levels	**Locality for Francophone African countries**
1: Continent	1: City or province (local component) within country
2 : Regional group (supranational)	2: Francophone African country
3: Country	3: Other African country, city or region
4: Province or state (subnational)	4: Other non-African countries/ continents (foreign news)
5: City	5: France

The first column, "geographical levels", purely refers to the level of the geographical keywords, and not to its relation to the country producing a given newspaper i. For instance, "France" would have level 3, "Paris" 5, "Aquitaine" 4. The second column gives a local index defining the relationship of each area with our sample of f Francophone African countries. For instance, any city or province belonging to our sample of African country is given a 1, these francophone countries are given a 2, any other non-Francophone African country, as Ghana, is given a 3 and so on.

Based on these two indexes, we then wrote a script that matches the country identifier of each newspaper and these indexes. A local news thus correspond to any news released by newspaper i and for which one of the geographical keyword matches the country identifier of newspaper i, has a geographical level (first column) of 4 or 5, and for which the locality (second column) is 1. A dummy equal to 1 is given to such kind of news. The final index for each

country is thus the monthly sum over all local news, divided by all articles released in a month.

French news ratio

The same process is used for computing the French news ratio. A dummy equal to 1 is created for any news having a locality index equal to 5 (column 2). We then compute for every francophone African country in our sample the total share of articles referring to France.

This indicator as no particular interest for the characteristics of Francophone African newspapers by themselves. It was purely introduced to see whether the entry of French newspapers had an influence on the number of topics related to France in African newspapers, as a contrast with the ratio of local news.

Main Explanatory Variable: Market Penetration by French Newspapers

The variable of market penetration is computed monthly for each country from 2014, using official circulation data provided by French newspapers and circulation rate extrapolated from Similar Web. We have three main kinds of circulation data: the total sum of visits of french newspaper websites per population in francophone African country c (monthly), the total sum of pages view and the total sum of unique users. We decided to use as final index of market penetration the total sum of visits of french newspaper websites, and performed robustness checks to test the validity of our results to alternative definitions of market penetration. We find our results are robust to different specifications using the total sum of pages view and the total sum of unique users. Despite being of different amplitudes, these different measures are indeed following a similar trend over time (see Figure A.1 in Annexes).

The measure used in this paper is the logarithmic sum of total visits of French newspaper websites by population from each Francophone African country c, rather than the direct share of news produced by French newspapers and consumed in country c. Market penetration is log-linearized. The reason why we do not directly rely on market shares is that we often miss accurate information about the exact circulation figures of all domestic newspapers available online. Firstly, our sample of newspapers, while trying to be as broad as possible, does not include an exhaustive list of domestic newspapers for every African countries. Secondly, although we know the relative ranking (in terms of visits) of each newspaper among all national newspapers, we miss reliable data about the exact share of the domestic readerships captured by each given newspaper i. Eventually, because online news is supplied as a commercial public good (non-rival and non-excludable, i.e there is no direct cost for consumers to acquire information once we account for the sunk cost of acquiring a receiving device), multiple newspapers readerships is possible and indeed assumed. Measures of unique users does not allow to capture the identity or uniqueness of these users *across* newspapers.

For instance, there are relatively small Francophone African countries for which we only have 2 or 3 online newspapers, while we have data for 9 french newspapers' websites. Using market shares based on this data can obviously distort the proportion of French newspapers for these countries (we indeed get some absurd values, like a 90% market penetration for a couple of countries). We minimize this bias by using the log of the total visits of French newspaper's websites for each francophone African country, while jointly controlling for population growth and the evolution of internet penetration. This allows to capture the

relative increase of market penetration by french firms.

Figure 1 displays the average number of visits per month of French newspaper websites by populations from Francophone Africa. We can clearly see the increasing trend over the period 2014-March 2017.

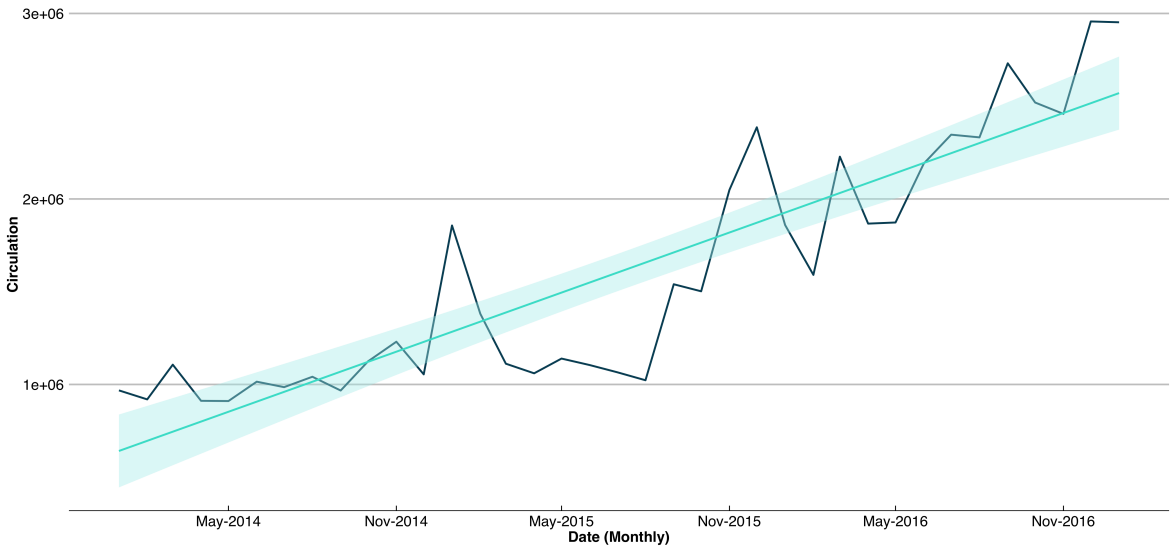

Figure 1: Market penetration of digital french newspapers in francophone Africa
Average over francophone African countries, 2014-2017

4 Empirical Framework

4.1 Main Identification Equation

We estimate throughout this paper several extensions of the following baseline equation:

$$Y_{ct} = \beta \, ln \, MP_{ct} + \delta \, X_{ct} + \varphi \, W_c + \gamma_t + \varepsilon_{ct} \quad (1)$$

for country $c = 1, 2, ...n$ and time $t = 1, 2, ...T$

Y_{ct} is the matrix response variables of endogenous newspapers characteristics describe in the previous Section, composed of (i) a measure of content variety based on the Herfindhal-Hirschman index adapted for concentration of newspaper topics $y_{1ct} = H^*$, (ii) a French news ratio y_{2ct} ,(iii) a local news ratio y_{3ct}, (iv) the log wordcount per article $ln \, y_{4ct}$, (v) the log frequency of publication per month $ln \, y_{5ct}$.

For a given time t, the matrix response variable of newspaper characteristics is given by:

$$Y_c = \begin{pmatrix} y_{1c} & y_{2c} & y_{3c} & y_{4c} & y_{5c} \end{pmatrix} = \begin{pmatrix} y_{11} & y_{21} & y_{31} & y_{41} & y_{51} \\ y_{12} & y_{22} & y_{32} & y_{42} & y_{52} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{1n} & y_{2n} & y_{3n} & y_{4n} & y_{5n} \end{pmatrix}$$

We try to assess the causal effect of the variable of market penetration $\boldsymbol{MP_{ct}}$ on newspapers characteristics across Francophone Africa. It represents the monthly logarithmic sum of visits of French newspaper websites per population in each given Francophone African country c, as described in details in the previous Section.

$\boldsymbol{X_{ct}}$ is a set of time-varying controls that captures African country-specific tastes for newspapers. This includes demographics and a constant term. Demographics include especially educational level, median income, median population age, literacy rates, percentage of french speakers and urbanization. It also includes controls that aim to capture the level of internet penetration per country, as the percentage of fixed broadband subscriptions, internet users or time required to access electricity. These sets of controls are aimed at capturing differences in demographics which correlates with the matrix of newspaper characteristics. Two interesting stylized facts from the media economics literature are in particular consistent with heterogenous preferences and tastes for newspapers variety. The first is that audience size is an increasing function of the number of choice. The second is that within the same market, different demographic groups make different choices.

$\boldsymbol{W_c}$ includes a set of time-invariant controls aiming to provide a measure of bilateral cultural distance between country c and France, beyond the share of French speakers in a country c. This includes log weighted distance between France and a given African country c, and a set of dummies for pre-colonial and colonial history: a dummy if the country-pair was ever in a colonial relationship, a dummy for common religion, a dummy for common legal origins after transition, and a dummy for conflicts.

$\boldsymbol{\gamma_t}$ is a year fixed-effect, which controls for time-varying differences in our vector of endogenous newspaper characteristics that are common across countries.

$\boldsymbol{\varepsilon_{ct}}$ is the vector of error for country c at time t, and captures among other unobservable household-specific tastes for newspapers and unobservable trade costs for online newspapers between France and country c. It is assumed to be i.i.d. Standard errors are clustered at the country level in all following regressions.

We already made several identification hypothesis in previous sections, and we will add new ones throughout this paper.

We generally expect the different dependent variables of newspaper characteristics to behave differently to the entry of French digital newspapers. We first believe the entry of French firms could lead to a relative homogenization of Francophone African newspapers content. In particular, in the event that French players do not produce significantly differentiated contents for each country, and based on the finding that the share of local news in French newspapers targeting Africa is very low compared to newspapers in each African country (see summary statistics in Annexes A.1 and A.2), the reaction of Francophone African media could be to focus relatively more on local news. Similarly, we can assume they would jointly decrease the share of news related to France on their websites. Both these features would contribute to increase the Herfindahl Index of concentration of newspaper topics, and thus to decrease the relative variety of news produced each month. A potential scenario is the one of a reallocation of topics between French newspapers and domestic newspapers, where domestic newspapers would focus more on national and local news, with maybe more investigative reporting, and French newspapers would relay more general subjects related indifferently to all African countries.

In terms of the particular format of Francophone African newspaper articles (as measures through wordcount and frequency of publication), we can expect that these characteristics to vary jointly with the characteristics of the new set of newspapers available from French players. For instance, if the mean wordcount in French articles is really higher than in African articles, we could maybe see an increase of wordcount across African articles over time. However, as the entry of French newspaper websites is still recent and their influence is not expected to be particularly strong[17], this effect is likely to be weak. Moreover, the specific format of the news displayed on the web could be such that all newspapers eventually have no significant differences in format. Articles tend in particular to be significantly shorter than in printed format, with a higher amount of short breaking news. Moreover, this trend reinforces over time in France and most Western countries.

The frequency of online publications is also rather high compared to printed newspapers in France. But teams of reporters working for the African version of newspapers' websites is often very limited in size, and this could lead to a rather moderated amount of articles released. This could thus not impact the dynamic of production of online newspaper articles across Francophone African countries.

4.2 Control for Initial Conditions

French newspapers provided us official data starting from 2014. However, we cannot assume the total number of visits of french websites by populations from francophone African countries was null before that date. Indeed, some french newspapers created a totally new website targeting specifically Africa. But other newspapers only added subdomains to their main domain (URL) to include pages targeting more specifically these populations. In any case, the African audience for French-owned websites cannot be assumed to be zero before data was available, and the move of audiences from the root URL to the new subdomains targeting Africa is certainly neither exclusive nor automatic. Assuming an entry shock in 2014 would surely strongly skew the results.

Consequently, we carry out the analysis of the effects of the penetration of french newspaper firms on newspaper characteristics in francophone Africa over the period 2014 - March 2017.

For most of African digital newspapers[18], we however collected articles as soon as 2010 or 2012. For a couple of francophone African countries in North Africa (Tunisia, Morocco and Algeria), we even have newspaper articles dated from 2005.

Hence, we are interested in including this additional information in our analysis, to capture press features in each national market before the entry of french newspapers. Controlling for any *trend* in digital news characteristics across markets especially prevent us from over or under-estimating the impact of entry of french companies.

In order to include these characteristics for 21 African countries over nine years as controls

[17] Circulation figures seem to indicates they account for a rather low share of all newspapers consumed by African populations, lower than French television programs for instance)

[18] We define a newspaper website as belonging to country c if it is hosted by a company located on the national land of this given country

in our multivariate regression analysis[19], we reduce their dimensionality by using Principal Component Analysis (PCA). Principal Component Analysis produces a correlation matrix between variables and allows to obtain a reduced number of composite indexes from them. This also addresses multicollinearity concerns because the fewer remaining variables included in the PCA are orthogonal (uncorrelated).

We perform two discrete steps in sequence for using PCA as controls. Firstly, we find weighted linear composite variables. Secondly, we use some of the principal components as PCA calculated variables for running regressions.

Table 3 displays the principal components and the associated eigenvalues for our dependent variables of newspaper characteristics. We use the PCA correlation matrix instead of the covariance matrix as our five variables are measured in very different units and we wish to ignore these differences between variances.

Table 3: **Principal Components for Outcome Variables, years before 2014**

	Comp1	Comp2	Comp3	Comp4	Comp5
Eigenvalue	2.597	1.033	.690	.435	.245
Proportion of variance	0.519	0.207	0.138	0.0870	0.0490
Cumulative	0.519	0.726	0.864	0.951	1.00

	Principal components (eigenvectors)				
	Comp1	Comp2	Comp3	Comp4	Comp5
Local News ratio	.453	.315	-.610	.386	-.417
French News Ratio	.279	.695	.654	.037	-.095
Concentration of topics (HHI index)	-.507	.233	-.022	.751	.352
Frequency of publication (monthly)	.377	-.602	.4234	535	-.171
Mean wordcount (monthly)	.563	.016	-.137	-.011	.814

The first matrix provides eigenvalues for each of the generated principal components. The relative magnitudes of the eigenvalues indicate the amount of variance they account for. The second correlation matrix gives the eigenvectors, which are the coefficients of the Principal Components. PCA displays the principal components in unit normalization: the sum of squares of the principal loadings equals 1. The higher the component loadings, the more important that variable is to the component[20] .

Based on the criterium that our eigenvalues for the correlation matrix are greater than one for components 1 and 2 (i.e. those components have variances greater than the average), we only retain these two components for our subsequent regressions. This is confirmed by plotting the eigenvalues of the correlation matrix (Figure 2). The first component explains 53 percent of the variance between variables, and the second one 21 percent. The first two

[19]We cannot include these variables by simply running our panel data regressions, as all observations for which our main panel regressor is missing are automatically dropped - that is all observations before 2014.

[20]However, the specific sign of the component loadings is not important.

components account for 73 percent of the variance in the data and summarize then most of the variation of the original variables.

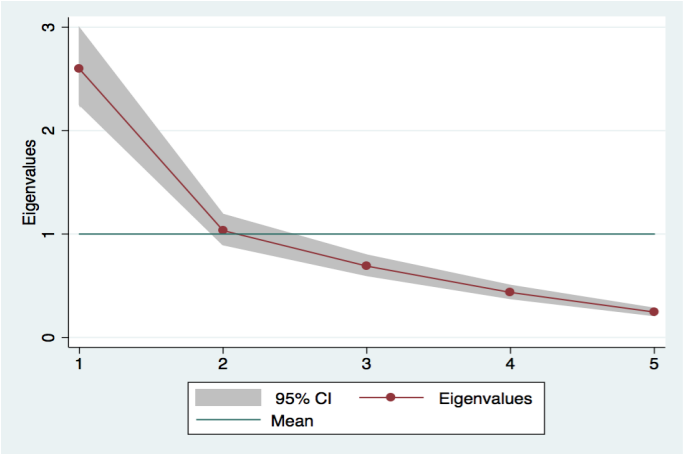

Figure 2: Eigenvalues after Principal Component Analyzis

We can thus rewrite our estimating equation as:

$$Y_{ct} = \beta \, ln \, MP_{ct} + \delta \, X_{ct} + \varphi \, W_c + \gamma_t + I_{ct} + \varepsilon_{ct} \tag{2}$$

with $c = 1, 2, ...n$ and $t = 1, 2, ...T$

where the vector of the two first principal components controlling for initial conditions, for a given country c at a fixed time t, is given by:

$$I_{ct} = \begin{pmatrix} I_{1ct} & I_{2ct} \end{pmatrix} = \begin{pmatrix} i_{11t} & i_{21t} \\ i_{12t} & i_{22t} \\ \vdots & \vdots \\ i_{1nt} & i_{2nt} \end{pmatrix}$$

5 Baseline Estimates

We present in this section the results of our baseline regressions, i.e the measured impact of the market penetration by french digital newspapers on each of the elements of the matrix of newspapers characteristics across Francophone Africa. All following regression tables present results for monthly observations, with panel clusters at the country level (21 clusters for a total of 762 unique monthly observations since 2014; 1268 since 2005). The following results exclusively analyze the evolution of the characteristics of digital newspapers produced across Francophone African, and thus do not include French newspaper articles in the sample. The precise list of newspapers analyzed for each country is given in Annexe C.1. Section E of the Annexes provides all regression tables without controls.

5.1 Multivariate Multiple OLS Regression Estimates

Because our model includes 5 different outcome variables, we first start by estimating a multivariate multiple OLS regression model, which allows to incorporate several y in a sole OLS regression[21].

Multivariate multiple regression estimates the same coefficients and standard errors as one would obtain using separate OLS regressions for each outcome variable. The residuals from Multivariate regression models are assumed to be multivariate normal, which is analogous to the assumption of normally distributed errors in univariate linear regression.

Multivariate multiple regression being a joint estimator, it also estimates the between-equation covariances (we can see in particular how highly the residuals of the several independent OLS equations are correlated). Multivariate multiple regression thus performs the two Principal Component Analysis steps simultaneously.[22] However, multivariate regression is more powerful than PCA, as the weighted linear composite variables 's are formed so as to maximize the regression.

The major change compared to running separate OLS regressions is thus that y is a matrix response variables and not a vector. The multivariate regression analysis makes especially sense as most of our outcome variables are at least moderately correlated over time.

Indicators of newspapers content

We find in Table 4 that the normalized Herfindahl-Hirschman index for concentration of newspaper topics increases on average for countries where the market penetration of french newspapers is higher (i.e. the total number of visits of French newspaper websites by population from a given Francophone African country c). The correlation with market penetration of French newspapers is positive and significant at the 5 percent level. On average across francophone African countries, an increase by one percent of market penetration is associated with a 2.39 percentage point increase in concentration of news topics, all other variables held constant. We can conclude that on average, the diversity of newspapers topics decreases in Francophone African countries following an increase in market penetration by french newspapers.

Forgetting to control for initial conditions of the digital newspaper markets leads to an downward bias of the point estimates (which becomes 1.82, see Notes in Table 4). Taking into account the characteristics of domestic online newspapers in francophone Africa from 2005 to 2013 here increases the coefficient estimate by 0.57 percentage point.

[21] We could also have used Principal Component Analysis (PCA) on the outcome variables and create a sole composite index based on the correlation matrix between the variables. However, running separate regressions for each dependent variable makes clearly appear that some indicator are much more impacted than others by our regressors. We would thus loose a substantial amount of information by reducing the dimension of the dependent variables to a sole multidimensional index.

[22] i.e., it finds weighted linear composite variables then regress them.

Table 4: **Multivariate Multiple OLS Estimates, monthly**

	Indicators of Newspaper Content						*Indicators of Newspaper Format*			
	Normalized HHI		**French News ratio**		**Local News ratio**		**Mean Wordcount**		**Publication Freq.**	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Market penetration (log total)	2.39***	(0.34)	0.18***	(0.03)	-1.74	(1.14)	-0.05***	(0.01)	0.44***	(0.05)
French speakers (log)	-0.48	(0.49)	-0.36***	(0.04)	12.98***	(1.65)	0.02	(0.02)	-0.44***	(0.07)
Log GDP per cap(current USD)	0.90	(1.28)	-0.57***	(0.11)	28.34***	(4.28)	-0.14**	(0.06)	-0.40**	(0.17)
Urban population (% of total)	0.07	(0.06)	0.02***	(0.00)	-0.52***	(0.19)	-0.01***	(0.00)	0.04***	(0.01)
Broadband subscriptions (%)	0.13	(0.57)	-0.16***	(0.05)	4.03**	(1.89)	0.08***	(0.02)	-0.22***	(0.08)
Internet users (%)	-0.16***	(0.04)	0.00	(0.00)	-0.42***	(0.15)	-0.00	(0.00)	-0.01	(0.01)
1=Colony	5.88***	(1.35)	-0.24**	(0.11)	17.37***	(4.51)	0.03	(0.06)	0.26	(0.18)
1=Common religion	7.81*	(4.28)	-3.25***	(0.36)	138.06***	(14.28)	0.08	(0.19)	-2.76***	(0.58)
1=Common legal origins	94.40***	(22.16)	4.42**	(1.88)	-58.45	(73.89)	6.55***	(0.97)	15.31***	(3.01)
Log distance (pop-wt, km)	-11.05***	(1.71)	0.11	(0.15)	-23.43***	(5.71)	0.08	(0.07)	-0.94***	(0.23)
Scores for components 1 and 2	Yes	Yes	Yes	Yes	Yes					
Observations	645									
adj. R^2	0.944		0.894		0.949		0.999		0.989	

Notes: Robust standard errors in parentheses. The unit of observation is a country-month. Mean Wrdcount and Frequency of Publication are in logs.
$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Coeff. for Market penetration (log total) without controls for initial conditions: 1.82*** (0.54) (HHI) , 0.21*** (0.07) (French ratio), -1.29 (1.26) (Local ratio) , -0.05*** (0.02) (Wordcount), 0.51*** (0.07) (Frequency Publication)
Controls also included: Population median age; Access to electricity (% of population,); Adult literacy rate, population 15+ year,(%); Mobile cellular subscriptions (%); Consumer price index (2010 = 100); GINI index (World Bank estimate); GINI index (World Bank estimate).

We also find a positive and significant correlation between the monthly share of articles from Francophone Africa referring to French news and market penetration of french newspapers. On average, an increase by one percent of the total number of visits of french newspaper websites is associated with a 0.18 percentage point increase in topics related to France in domestically produced newspapers. Controlling for initial conditions or not gives strongly significant points estimates at the 1 percent level (again slightly overestimated without these controls). Coefficients are here again slightly overestimated when missing to account for initial conditions across online newspaper markets.

However, we find no statistically significant correlation at the 5 percent level between market penetration of french firms and the share of local news released on Francophone African newspaper websites. OLS estimates do not allow us to conclude that the increase of visits of french newspaper websites by African populations significantly explain variations in the share of local news.

Indicators of newspapers format

Table 4 also reports a significant correlation at the country level between the increase of market penetration by French newspapers and the frequency of publication (i.e the mean number of articles released by month across Francophone African countries). Coefficients are statistically significant at the 1 percent level. An increase by 1 percent of market penetration by french newspaper firms is associated with a 0.44 percent increase of the number of articles released per month.

Eventually, we find that an increase by one percent of market penetration by French newspapers leads to a slight decrease by 0.05 percent of the mean wordcount per article per month in African newspapers, on average. The point estimate and associated standard error are similar with or without controlling for initial conditions, and significant at the 5 percent level.

5.2 Fixed Effects Estimates for Indicators of Digital Newspaper Format

5.2.1 Potential Bias in OLS Estimates

Multivariate multiple regression results are likely to be biased, as OLS estimates do not control for cross-sectional omitted variables bias neither for measurement errors. We could especially expect omitted variables bias due to non-pecuniary trade barriers, which could be part of what we called *cultural-distance barriers*.

Indeed, while the physical cost of delivering information services may be low (especially for online news), the barriers consumers face in gaining access to them may not be. Language and culture may thus affect consumer demand and the cost of marketing across borders for French firms. We expect these implicit bilateral trade barriers to depend on the percentage of people speaking French in a country, or to the fact the country-pair *France - African country c* was ever in a colonial relationship, share common legal origins or religion. Although we accounted for these covariates in our OLS regressions, we could miss to account for other similar factors of cultural distance. In particular, the Hofstede's cultural index (Hofstede, 1980, 2001) has been used by numerous papers studying trade in cultural goods (Fu and

Lee, 2008; McFadyen, et al., 2003, 2004; Oh, 2001) to quantify exporter-to-importer cultural distance. The Hofstede project indexes the cultures of 65 countries in terms of power distance, individualism, masculinity, and uncertainty avoidance.

While this framework for distinguishing national cultures is frequently adopted in the international trade literature as a proxy for bilateral trade barriers (see Tihanyi, Griffith and Russell, 2005, for a review), the Hofstede's cultural index does not have any available data for the sample of francophone African countries targeted in the present paper.

Ordinary Least Squares regressions neither allow to address the issue of potential measurement errors. For the purpose of this paper, measurement errors could especially be due to circulation data, and in particular to inaccurate inferences of circulation data based on websites as Similar Web or Google Analytics. Official circulation data directly provided by French media firms also often relies on statistics they derive from Google Analytics, which are still believed not to be totally accurate. These issues could also lead to biased OLS point estimates.

We can partially address these threats to internal validity, in particular omitted variable bias, by adding two sets of fixed effects: bilateral country-pair fixed effects, which remove all time-invariant determinants of trade flows like distance or language; and time fixed effects, which control for time effects whenever unexpected variation or special events may affect my outcome variables. Including fixed effects for the panel identifiers allows us to account for heterogeneity over time and across countries.

5.2.2 Fixed Effects Panel Data Model Specification

We use a Fixed Effects model with clustered–robust standard errors at the panel identifier level, because our error terms are likely to be correlated within countries by demographic groups over time. Clustered standard errors allow for intragroup correlation, hence relaxing the usual requirement that the observations be independent.[23] These standard errors are moreover robust to heteroskedasticity.

The regression equation thus becomes:

$$Y_{ct} = \beta \, ln \, MP_{ct} + \delta \, X_{ct} + I_{ct} + \alpha_c + \gamma_t + \varepsilon_{ct} \tag{3}$$

$$\text{where} \quad u_{ct} = \gamma_t + \alpha_c + \varepsilon_{ct} \quad \text{is the error term.}$$

Our vector W_{ct} of time-invariant covariates is replaced in the regression model by a country-specific fixed effect α_c in order to control for all time-invariant unobserved country-specific characteristics. The equation also includes year fixed effects, γ_t, which control for time-varying differences in the matrix of newspaper characteristics that are common across countries. ε_{ct} is the observation-specific error term, which capture purely random or unpredictable effects.

The concern with the multivariate OLS estimation is thus that the estimate of β could be biased if $lnMP_{ct}$ is correlated with α_c the fixed, unobserved characteristics. This is corrected by the Fixed Effect transformation, where we regress the individual-demeaned Y on individual-demeaned explanatory variables. Assuming that there is no time-variant error term,

[23] The observations are thus independent across countries (clusters) but not necessarily within countries.

$$Y_{ct} - \overline{Y}_c = \beta\, ln\,(MP_{ct} - \overline{MP}_c) + \delta\,(X_{ct} - \overline{X}_c) + (I_{ct} - \overline{I}_c) + \varepsilon_{ct} - \overline{\varepsilon}_c \quad (4)$$

We can apply the same demeaning process for time fixed-effects.

We are mainly interested in the time-series information reflected in changes *within* subjects to capture the potential effect of the entry of French newspapers across Francophone African markets. Fixed effects provide within-estimates that remove the effect of time-invariant characteristics such that we can assess the net impact of the predictors on our outcome variables. As fixed-effects do not work well with data for which within-cluster variation is minimal or for slow changing variables over time, we perform a Hausman test. It strongly rejects the consistency of random effects against fixed-effects.

Table 5: **Fixed Effect Estimates for Indicators of Newspaper Format, monthly**

	Mean Wordcount		**Frequency of Publication**	
	without i.c	i.c	without i.c	i.c
Market penetration (log total)	-0.05**	-0.05**	0.20*	0.27**
	(0.02)	(0.02)	(0.10)	(0.10)
French speakers (log)	-0.19**	-0.18**	0.31***	0.31***
	(0.08)	(0.07)	(0.09)	(0.10)
Days to start a business	0.03	0.04	-0.33***	-0.14***
	(0.03)	(0.03)	(0.04)	(0.04)
Broadband subscriptions (%)	-0.19**	-0.17*	-0.17	0.20**
	(0.09)	(0.08)	(0.16)	(0.10)
Internet users (%)	0.04**	0.04**	-0.08***	-0.11***
	(0.01)	(0.01)	(0.02)	(0.02)
Urban population (% of total)	-0.24	-0.21	-0.39	0.25
	(0.18)	(0.17)	(0.31)	(0.24)
Scores for component 1		0.05** (0.02)		0.82***(0.05)
Scores for component 2		0.01 (0.02)		-0.31*** (0.05)
Country Fixed effects	Yes	Yes	Yes	Yes
Time Fixed effects	Yes	Yes	Yes	Yes
Observations	645			
adj. R^2	0.171	0.207	0.212	0.642

Notes: Robust standard errors in parentheses are clustered by country. The unit of observation is a country-month. All the estimations include country and time (yearly) fixed effects. Mean wordcount and Frequency of Publication are log transformed.

$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Columns *i.c* provide point estimates and robust standard errors with additional controls for initial conditions; without i.c do not include control for initial conditions.

Controls also include: Access to electricity (% of population); GINI index;Mobile subscriptions (%); CPI (2010 = 100); Log GDP per capita (current USD); Adult literacy rate, pop. 15+ (%)

5.2.3 Estimation Results

Results from estimating this Fixed Effects model are displayed in Table 5.

Market penetration by French media firms has significant effects at the 5 percent level on both the frequency of publication per month (number of articles released in a given month) and the mean wordcount per month across countries, once we account for initial conditions. On average, an increase by 1 percent of French newspapers penetration in a given country c would increase the number of articles released by domestic newspapers in this country by 0.27 percent in a month. Forgetting to control for initial conditions leads to an underestimated point estimate, which lowers the impact to 0.20 percent change. The OLS estimate without Fixed Effect was thus slightly biased upward (see Table 4).

On the contrary, a one percent increase in market penetration is associated with a slight decrease by 0.05 percent in mean wordcount per month for domestic newspapers of country c. This latter result is consistent to the addition of control for initial conditions. This result is similar to the one obtained with simple multivariate multiple OLS regression.

Both point estimates for the percentage change of mean wordcount and frequency of publication due to a one percent increase in market penetration by French newspapers show that the amplitude of the variations is significant but fairly low. The percentage of variance in mean wordcount per month explained by the entry of French newspapers is low as well. Consequently, the increase in total visits of French newspaper websites by citizens from a given Francophone African country c does not seem to offer a meaningful interpretation of what could cause mean wordcount to change. We can see the percentage of french speakers in a given country has a higher explanatory power for these two indicators of newspapers format.

These two point estimates do not highly differs from the simple OLS case. They are also similar to Fixed Effect estimates when we fit instead a population-average model. A population-average regression performs a cross-section analysis with adjustment for correlation over time for a given panel identifier. The significance and the amplitude of these two point estimates are thus not highly dependent of variations in our model specification, which is a sign of robustness.

5.3 Fixed Effects Tobit Estimates for Indicators of Digital Newspapers' Content

5.3.1 Fixed Effects Tobit Model Specification

Three of the dependent variables of newspaper characteristics are indexes measured as proportions or percentage. These outcome variables (Normalized Herfindahl Hirschman Index, French and Local news ratio) can only take values in a range from 0 to 100. The greatest issue with ordinary linear regression, with or without fixed effects, is that the model can predict values that are not possible (i.e values below 0 or above 100).

The simplest approach is to perform a linear regression anyway, what we did in Table 4. A generally more accurate approach is to use Tobit model for longitudinal data, in order to

treat proportions as censored continous variables.

We thus fit a two-limits Fixed Effects Tobit model (Long, 1997), with limits 0 and 100. The two-sided censoring constrains our fitted values within this range. This approach works best if there is not an excessive amount of censoring (values of 0 and 100), which is the case here.

We use a Tobit-type model with "Fixed" Effects, following an approach developed by Honore (1992)[24], which censors the dependent variable in such a way that the country-specific effect can be differenced away (what a basic Tobit Model does not, as it estimates Random Effects). We use more specifically the Fixed Effects Tobit with two-sided censoring model that was derived from Honore (1992) by Alan, Honere et al. (2011)[25], and may be described by:

$$Y^*_{ct} = \beta \, ln \, MP_{ct} + \delta \, X_{ct} + I_{ct} + \alpha_c + \gamma_t + \varepsilon_{ct} \tag{5}$$

$$Y_{ct} = \begin{cases} \overline{\tau} = 100 & \text{if } Y^*_{ct} > \overline{\tau} \\ Y^*_{ct} & \text{if } \overline{\tau} \geqslant Y^*_{it} \geqslant \underline{\tau} \\ \underline{\tau} = 0 & \text{if } Y^*_{ct} < \underline{\tau} \end{cases} \tag{6}$$

with $c = 1, 2.., n$ and $t = 1, 2..T$

with Y^*_{ct} being the desired (i.e true) level of any of the variables y_{1ct}, y_{2ct} or y_{3ct} in the absence of the censoring constraint, and c is the lowest possible value that a given y_{ct} variable can take. Here $\underline{\tau} = 0$ and $\overline{\tau} = 100$

Table 6 also reports coefficients for *sigma_u* and *sigma_e*, which are such that:

$$\rho = \frac{\sigma^2_\alpha}{\sigma^2_\varepsilon + \sigma^2_\alpha} \tag{7}$$

where ρ is the percent contribution to the total variance of the panel-level variance component. σ^2_α is the panel-level variance component and σ^2_ε is the overall variance component (respectively *sigma_alpha* and *sigma_e* in Table 6).

We find that ρ is significantly different from zero, hence the panel estimator is different from the pooled estimator. A likelihood-ratio test of this indeed rejects the null hypothesis that there are no panel-level effects.

5.3.2 Estimation Results

We find in Table 6 that the increase of market penetration by French newspapers has a significant positive effect at the 5 percent level on the Normalized Herfindahl-Hirschman index (HHI) for concentration of newspaper topics. On average across Francophone African countries in our sample, an increase by one percent of market penetration by French newspapers is associated with a 1.42 percentage point increase in concentration of news topics, all other

[24] Honoré, Bo E. (1992): "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, Vol. 60, No. 3, Pages 533-565.

[25] Alan, Sule, Bo E. Honoré, Luojia Hu and Søren Leth-Pedersen (2011): "Estimation of Panel Data Regression Models with Two-sided Censoring or Truncation"

Table 6: **Fixed Effects Tobit Estimates for Indicators of Newspaper Content, monthly**

	Normalized HHI		**Local News Ratio**		**French News Ratio**	
	without i.c	i.c	without i.c	i.c	without i.c	i.c
Market penetration (log total)	2.66***	1.42**	0.13	-0.59	0.52***	0.13***
	(1.02)	(0.49)	(1.54)	(1.53)	(0.12)	(0.04)
French speakers (log)	0.49	0.04	3.68	3.52	-0.26	0.04
	(2.27)	(0.94)	(3.42)	(2.94)	(0.24)	(0.07)
Log GDP per capita (current USD)	3.11	-1.03	18.67*	34.05***	-1.33*	-1.02***
	(6.80)	(2.56)	(10.25)	(9.73)	(0.75)	(0.21)
Log distance (pop-wt, km)	-367.49	-6.78	585.47	-84.91**	15.51	1.63*
	(420.17)	(7.86)	(633.46)	(34.07)	(48.22)	(0.86)
1=Common religion	1181.16	-19.12	-2774.52	174.67*	-125.73	-4.98**
	(1187.50)	(21.65)	(1790.28)	(92.06)	(138.98)	(2.23)
Broadband subscriptions (%)	1.21	-3.72**	-12.55***	-6.22**	-0.35	0.37***
	(2.24)	(0.96)	(3.38)	(2.97)	(0.24)	(0.07)
Access to electricity (% of pop.)	3.28	-0.12	-13.56**	-1.00***	-0.64	0.04***
	(3.96)	(0.09)	(5.97)	(0.36)	(0.46)	(0.01)
Days to start a business	1.23	-1.83**	-6.38**	-1.75***	-0.27	0.05***
	(1.66)	(0.07)	(2.50)	(0.33)	(0.19)	(0.01)
Scores for components 1 and 2	No	Yes	No	Yes	No	Yes
Country Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
sigma_alpha	0.00	1.62***	0.00	8.43***	0.00	0.22***
	(0.32)	(0.42)	(0.48)	(2.26)	(0.03)	(0.05)
sigma_e	7.23***	4.04***	10.90***	11.05***	0.78***	0.25***
	(0.23)	(0.13)	(0.34)	(0.36)	(0.03)	(0.01)
Observations	580	580	580	580	580	580

Notes: Robust standard errors in parentheses are clustered by country. The unit of observation is a country-month. All the estimations include country and time (yearly) fixed effects. Normalized HHI is an index from 0 to 100.
$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Columns *i.c* provide point estimates and robust standard errors with additional controls for initial conditions; without i.c do not include control for initial conditions.
Controls also include: Access to electricity (% of pop.); Mobile subscriptions (%); Colony and Common legal origins with France, Adult literacy rate, pop. 15+ (%); Population median age; Internet users; Urban population (% of total);CPI (2010 = 100)

variables held constant. As this index is an index of market concentration (the higher the index, the higher the concentration of topics), the diversity of subjects treated in a given month for a given country decreases by 1.42 percentage point following a one percent increase of french market shares. The absence of control for initial conditions leads here again to an upward bias of the point estimate for market penetration of french newspapers, which increases to 2.66. Taking into account the characteristics of domestic online newspapers in francophone African *c*, from 2005 to 2013, thus reduces coefficient estimates by 1.24 percentage point.

This finding vary very little compared to our previous model specifications. We can conclude that on average, across our 21 Francophone African countries of interest, the diversity of newspapers topics decreases in countries facing a higher rates of market penetration by french newspapers.

The percentage of fixed broadband subscriptions also has a statistically significant negative correlation with the HHI index of concentration of topics at the 5 percent level. An increase by one percent of the number of fixed broadband subscriptions leads to an average decrease by 3.72 percentage points of the concentration of topics. A higher rate of digitalization of news medias thus seems to increase the diversity of topics treated by domestic newspapers across francophone African countries.

Market penetration also seems to have a significant effects on the French news ratio (i.e. the share of news articles referring to French topics in a given month). An increase in the circulation of French newspaper in a given country c is correlated with an increase by 0.13 percentage point of the proportion of topics referring to France (french news ratio). This coefficient is significant at the 1 percent level. Without controls for initial conditions, we find an increase by 0.52 percentage points, and thus it seems we overestimate the impact.

Although this increase is small, it is strongly significant and reflects the relatively small proportions of news referring to France across African countries in our sample (which varies between 0 and 11 percent per month, with a mean of 0.47 and a standard deviation of 0.91, see Annexe A.1). Francophone African newspapers thus seem to display a much higher rate of articles referring to France compared to their initial levels.

However, an increase by one percent in terms of market penetration by French newspapers does not seem to have any effect on the local news ratio (share of articles referring to domestic, local news -at the city or regional level- in a given month, for each francophone African country). Point estimates are not significant, with very high standard errors and important variations compared to the multivariate OLS case.

Two variables seem to better explain variations in the local news ratio.

Firstly, this local ratio seems very sensitive to changes in log GDP per capita. On average, Francophone African countries having a higher level of GDP per capita also have newspapers displaying a greater amount of articles referring to domestic, local news. This could translate the fact this countries can afford to invest more money for local reporting (i.e. for producing more specific subjects), or that they have a greater diversity of online newspapers (with a greater share of local newspapers for instance, or website' subdomains for local news).

Secondly, measures of internet penetration seem to have a significant impact at the 5 percent level on local news ratio. On average, the percentage of fixed broadband subscription is negatively correlated with the share of local news produced. A Francophone African country with a 1 percent higher share of fixed broadband subscriptions has newspaper websites producing a 6.22 percentage points lower share of local news. This could be interpreted as the fact digitalization allows internet users to access a broader category and/or more global news, and that newspapers in countries with easier internet access invest less in local contents. This could be especially true for developping countries in which the rate of internet penetration is still low, but rising fast, as it is the case across Francophone Africa.

6 Endogeneity of Market Penetration by French Newspaper: Instrumental Variables Strategies

Several econometric problems may arise from estimating the previous models. The consistency and the accuracy of inference regarding our Fixed Effect estimators requires that a strict exogeneity assumption be satisfied between the explanatory variables and the error term u_{ct}:

$$\mathrm{E}(u_{ct}|x_c, \alpha_c = 0) \tag{8}$$

where $x_c = \{x_{c,s} \forall c = 1, ..., T\}$ denotes our complete set of explanatory variables at all points in time (future and past).

However, our main identification problem is that the explanatory variable of interest, market penetration of French news MP_{ct}, is assumed to be endogenous. Because causality may run in both directions – from french penetration in new francophone markets to the characteristics of domestics newspapers in these new markets, and vice versa – these regressors may be correlated with the error term. Strict exogeneity is thus violated.

Moreover, we expect our dependent variables to be partly determined by their past values. Therefore, we may want to add lagged dependent variable Y_{ct-1} as explanatory variables, which gives rise to autocorrelation. Controlling for initial conditions for years 2005-2013 does not control for serial-correlation of Y_{ct} over the time period included in the regression (2014-2017). As our panel is a proper time series, we can use estimators for dynamic panel data (DPD) models.

6.1 D-GMM and S-GMM Instrumental Variables *à la* Arellano-Blundell-Bond

To estimate our dynamic panel model with an endogenous regressor, we first rely on instrumental variables *à la* Arellano-Bond, using Generalized Method of Moments estimators (GMM). Such an approach has the advantage to mitigate the potential issue of weak instruments. The Fixed Effects IV estimators are indeed likely to be biased in the way of the OLS estimators if instruments are weak.

We first perform Difference GMM (D-GMM) due to Arellano and Bond (1991), who actually built upon earlier work by Anderson and Hsiao (1982) and Holtz-Eakin, et al. (1988). We then augment the D-GMM estimation strategy (under additional assumptions) using System GMM (S-GMM). This technic is due to Blundell and Bond (1998), using an insight from Arellano and Bover (1995).

6.1.1 D-GMM *à la* Arellano-Bond

Principle and Model Specification

With Difference GMM, we take the first difference of the linear dynamic panel regression to remove both the constant term and the individual-specific, unobserved effect. Once the individual fixed effects swept out, we have straightforward instrumental variables. The principle is that past levels of the dependent variable and of the endogenous regressor MP_{ct} are used

as instruments for the current first differences of the dependent variable. We thus construct instruments for the lagged dependent variable from the second and third lags of our matrix of y, in the form of differences (D-GMM, Arellano and Bond 1991). If our idiosynchratic error term ε is i.i.d, lags of endogenous regressors will be highly correlated with the lagged dependent variable but not with the composite error term.

Hence, using lagged levels of the endogenous regressors makes the endogenous variables predetermined and hence not correlated with the error term. This is a similar strategy to a traditional two-stage least squares identification.

A key aspect of the this Arellano-Bond estimator is thus he assumption that the necessary instruments are *internal*, which means based on lagged values of the instrumented variable(s):

$$Y_{ct} = Y_{ct-1} + \beta \, ln \, MP_{ct} + \delta \, X_{ct} + I_{ct} + u_{ct} \tag{9}$$

$$\text{where } u_{ct} = \alpha_c + \varepsilon_{ct}$$

where X_{ct} includes strictly exogenous regressors, I_{ct} is predetermined, and $lnMP_{ct}$ and Y_{ct-1} are endogenous regressors, all of which may be correlated with α_c , the unobserved individual effect. Removing α_c via fixed effects (within transform) produces an inconsistent estimator with T fixed. Instead, first-differencing the model equation yields:

$$\Delta Y_{ct} = \Delta Y_{ct-1} + \beta \, ln \, \Delta MP_{ct} + \delta \, \Delta X_{ct} + \Delta I_{ct} + \Delta u_{ct} \tag{10}$$

ΔY_{ct} is correlated with $\Delta \varepsilon_{ct}$ by construction. By first-differencing, the fixed country-specific effect (and its associated omitted-variable bias) is removed, because it does not vary with time:

$$\Delta u_{ct} = \Delta \alpha_c + \Delta \varepsilon_{ct} = \varepsilon_{ct} - \varepsilon_{ct-1} \tag{11}$$

. *Instruments*

The principle of Arellano and Bond (1991) is thus to construct estimators based on moment equations constructed from further lagged levels of Y_{ct}, of the main endogenous regressor MP_{ct} and of the first-differenced errors. Under D-GMM, lags of the dependent variable and the endogenous variable greater than or equal to 2 are valid instruments for the differenced equation. Our sets of instruments are thus:

$$\mathbf{Z}_{1c} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ Y_{c,1} & 0 & 0 \cdots & 0 \\ 0 & Y_{c,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Y_{c,T-2} \end{pmatrix} \quad \text{and} \quad \mathbf{Z}_{2c} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ MP_{c,1} & 0 & 0 \cdots & 0 \\ 0 & MP_{c,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & MP_{c,T-2} \end{pmatrix}$$

This constructs a set of instruments from the second lag of Y and MP, one instrument pertaining to each time period. The first row corresponds to $t = 2$, given that the first observation is lost in applying the FD transformation. The transformation automatically includes zeros in place of missing values, which prevents the loss of additional degrees of freedom. Despite this arbitrary inclusion of zeros, the columns of the resulting instrument

matrix will be orthogonal to the transformed errors.

Such a method is also useful as we can restrict the estimation sample from year 2014 (date from which we have data on french newspapers' market shares), but without restricting the sample from which lagged variables are drawn for instrument construction. Therefore, we can use lagged values of our dependent variable (newspaper characteristics) as far as 2005 in the estimation.

D-GMM (and System GMM) also require no correlation of the unobserved effects and error terms across cross-section units of our panel. A simple method generally used to account for at least some correlation is to include time dummies in the linear regression.

Estimation results

Table 7 displays the result of the D-GMM estimation.

The effect of an increase of market penetration by French newspapers on the Normalized Herfindahl-Hirschman index (HHI) for concentration of newspaper topics is still positive and statistically significant at the 1 percent level. Compared to our previous model specifications, the coefficient slightly increases: on average, across francophone African countries, an increase by one percent of market penetration by French newspapers is associated with a 2.50 percentage point increase in concentration of news topics, all other variables held constant. The absence of control for initial conditions leads here again to an upward bias of the point estimates for market penetration of french newsapers, which increases to 3.30 (main coefficients without initial conditions control are displayed in footnotes of Table 7).

Market penetration by french newspapers also has a statistically significant positive correlation with the share of articles referring to French news at the 5 percent level. An increase by one percent of market penetration leads to an average increase by 0.36 percentage point of french topics in domestic newspapers produced in African country c, on average.

The increase in the total number of visits to French newspaper websites also displays a positive and significant correlation at the 1 percent level with the monthly frequency of publication across countries. A one percent increase of french websites visits by populations across Francophone Africa leads to an average increase in publication frequency by 0.23 percentage points. This is slightly less than the point estimate with Fixed Effects.

There is thus an increase in two of these three coefficients compared to the fixed-effects estimates. The coefficient for the impact of market penetration on mean wordcount is nevertheless equal to zero and no longer significant (both with and without controls for initial conditions). Furthermore, the increase in total visits of french websites still does not seem to explain potential variations in the local news ratio of domestic newspapers produced in Francophone African countries.

Table 7: **Arellano-Bond estimators with robust standard error (SE) and controls for initial conditions**

	Indicators of newspapers content						*Indicators of newspapers format*			
	Norm. HHI index		**French news ratio**		**Local news ratio**		**Mean Wordcount**		**Freq. Publication**	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
L.Norm HHI	0.16***	(0.04)								
L2.Norm HHI	0.08	(0.06)								
L.French ratio			0.02**	(0.02)						
L2.French ratio			-0.01	(0.03)						
L.Local ratio					0.51***	(0.07)				
L2.Local ratio					-0.04	(0.08)				
L.Mean Wordcount							0.35***	(0.08)		
L2.Mean Wordcount							-0.05	(0.04)		
L.Freq. Publication									0.41***	(0.07)
L2.Freq. Publication									0.03	(0.05)
Market penetration	2.50***	(0.91)	0.36 **	(0.05)	0.98	(1.92)	-0.00	(0.03)	0.23**	(0.11)
L.Market penetration	-0.18	(1.00)	-0.10*	(0.06)	4.63	(3.03)	-0.02	(0.05)	-0.11	(0.09)
French speakers (log)	-1.08	(1.06)	0.06*	(0.04)	-0.60	(1.27)	-0.05**	(0.02)	-0.04	(0.07)
Broadband subscriptions (%)	-3.18***	(1.19)	0.28***	(0.09)	-4.41**	(2.01)	-0.12***	(0.04)	0.11	(0.09)
Internet users (%)	0.22	(0.43)	-0.05***	(0.01)	1.25***	(0.44)	0.00	(0.01)	-0.01	(0.04)
Urban population (% of total)	-2.95	(2.78)	0.50**	(0.22)	-3.49	(7.28)	-0.15	(0.14)	0.37	(0.31)
Consumer price index (2010)	-0.01	(0.08)	-0.01***	(0.00)	0.12	(0.10)	-0.00***	(0.00)	-0.01**	(0.00)
Log GDP per cap(current US$)	-2.37	(4.44)	-0.76***	(0.18)	7.38	(8.14)	0.31***	(0.11)	-0.60***	(0.19)
Scores for components 1 and 2	Yes		Yes		Yes		Yes		Yes	
Country Fixed effects	Yes		Yes		Yes		Yes		Yes	
Time Fixed effects	Yes		Yes		Yes		Yes		Yes	
N	558		558		558		558		558	

Notes: Robust standard errors in parentheses are clustered by country. The unit of observation is a country-month. All the estimations include country and time (yearly) fixed effects.Mean Wordcount, Freq. Publication and Market penetration are in log

$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Coeff. for market penetration without controls for initial conditions: 3.30*** (HHI) , 0.49*** (French ratio), 0.64 (Local ratio) , 0.00 (Wordcount), 0.18 (Frequency Publication).

Controls also include: Access to electricity (% of pop.); Mobile subscriptions (%); Days to start a business; Adult literacy rate(%), L2.Market penetration

Issues of many and weak instruments

One main concern with D-GMM and S-GMM estimators is the proliferation of instruments (the number of instruments rises with the advance of time, as more past levels of the endogenous variables are valid instruments under the assumptions). In order to prevent the creation of too many instruments, which presents a real danger of overfitting the endogenous variable, we use two common methods:

- Firstly, we limits the number of lags so that the total number of instruments does not exceed the number of panel-group identifier (countries) in our sample (otherwise the Sargan test is weak).
- Secondly, we collapse our matrix of lagged instrument so that only one instrument by panel and lag are to be used instead of an additional instrument for each panel, time, and lag.

The instruments matrices thus become :

$$\mathbf{Z}_{1c} = \begin{pmatrix} 0 \\ Y_{c,1} \\ \vdots \\ Y_{c,T-2} \end{pmatrix} \quad \text{and} \quad \mathbf{Z}_{2c} = \begin{pmatrix} 0 \\ MP_{c,1} \\ \vdots \\ MP_{c,T-2} \end{pmatrix}$$

If the model was exactly identified, it would be impossible to detect invalid instruments. However, as this type of GMM model is overidentified, we can use a test statistic for the joint validity of the moment conditions (identifying restrictions). Given this surfeit, we run several common overidentification tests. In particular, we fail to reject the null hypothesis under the Hansen's J-statistic test for heteroskedastic errors, which indicates that our instruments are valid. As neither our many y nor our market penetration variable are extremely persistent, we indeed expected the lagged levels of Y and MP not to be weak instruments for Y and MP respectively in D-GMM.

We also run the Arellano-Bond test for zero autocorrelation in first-differenced errors (one of the main assumption of the Arellano-Bond estimator is that there is no autocorrelation of the residuals). By construction, the residuals of the differenced equation should possess serial correlation, but if the assumption of serial independence in the original errors is warranted, the differenced residuals should not exhibit significant AR(2) behavior. Under our set-up, we reject no autocorrelation of order 1 and cannot reject no autocorrelation of order 2. Consequently, there is evidence that both these assumptions are satisfied, so that second lags of endogenous variables are appropriate instruments for their current values. Table 7 indeed shows that all dependent variables of endogenous newspapers characteristics are significantly correlated with their first lagged value, bu that second lags display no correlation.

This D-GMM estimator has however has potential weaknesses, because the lagged levels are often rather poor instruments for first-differenced variables (especially for variables that are close to a random walk).

6.1.2 S-GMM *à la* Blundell-Bond with Forward Orthogonal Deviations

Principles and Instruments

We thus use in a second step System GMM (S-GMM). System GMM augments D-GMM by adding an assumption which generates an additional set of moment conditions to leverage. It requires that lagged *changes* in the dependent variable are valid instruments for the*level* of the lagged dependent variable in the level equation. It thus uses the levels equation (Equation (9) without differences) to obtain a system of two equations: one in differences and one in levels. We obtain additional instruments, as the variables in levels are instrumented with their own first differences.

Under this assumption, S-GMM achieves a greater efficiency than F-GMM. Moreover, it allows to estimate the effects of time-invariant regressors since it uses the level version of the dynamic panel model in addition to the differenced version. Lags of the change in the dependent variable greater than or equal to 1 are now valid instruments for the level equation.

We rely on the version of Blundell-Bond system GMM with Forward Orthogonal Deviations (FOD), as proposed by Arellano and Bover (1995). Here the average of all available *future* observations for each variable is subtracted from the current-value variable rather than the once-lagged variable, while still removing the individual-specific, unobserved effect.

Using forward orthogonal deviations instead of simple S-GMM is especially useful as our panel dataset is unbalanced. Indeed, it will preserve degrees-of-freedom, which are lost with lots of first-differencing. The main disadvantage of the first-difference GMM transformation is indeed that it magnifies gaps in unbalanced panels. Any missing observation for a given y_{ct} will imply that both Δy_{ct} and Δy_{ct-1} will be dropped in the regression. We can indeed observe in Table 7 that the number of monthly observations is falling compared to previous estimations.

This is solved with forward orthogonal deviations. While the F-GMM drops the first observation for each individual in the panel, the FOD transformation drops the last observation for each individual and has the advantage of being computable for all periods excepted the last one, even for unbalanced panels. It also allows for autocorrelated errors.

Estimation results

Compared to D-GMM, we can see on Table 8 that S-GMM increases the value of coefficients for the normalized Herfindahl-Hirschman index and the Frequency of publication. A one percent increase of market penetration by french newspapers is now correlated with a 3.13 percentage points increase of the HHI index for topics concentration. Frequency of publication increases by 0.26 percent. Both results are significant at the 1 percent level.

The impact of market penetration by french newspapers is also associated with an increase of the ratio of news related to France in domestic African digitial newspapers by 0.11 percentage points. This coefficients witnesses a slight decline compared to the D-GMM instrumental variables estimates.

Table 8: **Blundell-Bond system GMM with forward orthogonal deviations and controls for initial conditions**

	Indicators of newspapers content						*Indicators of newspapers format*			
	Norm. HHI index		**French news ratio**		**Local news ratio**		**Mean Wordcount**		**Freq. Publication**	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
L.Norm HHI	0.17***	(0.04)								
L2.Norm HHI	0.07	(0.06)								
L.French ratio			-0.01	(0.03)						
L2.French ratio			-0.05*	(0.03)						
L.Local ratio					0.58***	(0.08)				
L2.Local ratio					0.02	(0.08)				
L.Mean Wordcount							0.44***	(0.07)		
L2.Mean Wordcount							0.02	(0.05)		
L.Freq. Publication									0.46***	(0.06)
L2.Freq. Publication									0.06	(0.05)
Market penetration	3.13***	(0.81)	0.11**	(0.06)	-0.43	(2.17)	-0.00	(0.03)	0.26***	(0.09)
L.Market penetration	0.18	(1.19)	-0.00	(0.05)	2.64*	(1.51)	-0.04	(0.04)	0.00	(0.11)
French speakers (log)	-1.04	(0.90)	-0.12	(0.08)	0.65	(1.16)	0.03	(0.04)	-0.13	(0.11)
Broadband subscriptions (%)	1.00*	(0.53)	0.11	(0.08)	-1.92	(2.06)	-0.01	(0.03)	0.02	(0.09)
Internet users (%)	-0.04	(0.11)	0.02**	(0.01)	-0.32	(0.23)	-0.00	(0.00)	0.00	(0.01)
Urban population (% of total)	0.09	(0.09)	0.02*	(0.01)	-0.03	(0.16)	-0.01**	(0.00)	0.01	(0.01)
Consumer price index (2010)	-0.08**	(0.03)	-0.01**	(0.00)	0.07	(0.08)	0.00	(0.00)	-0.01	(0.00)
Log GDP per cap (current USD)	-2.66	(2.27)	-0.74**	(0.33)	8.76	(10.00)	0.19**	(0.08)	-0.42**	(0.21)
Scores for components 1 and 2	Yes		Yes		Yes		Yes		Yes	
Country Fixed effects	Yes		Yes		Yes		Yes		Yes	
Time Fixed effects	Yes		Yes		Yes		Yes		Yes	
N	628		614		614		628		628	

Notes: Robust standard errors in parentheses are clustered by country. The unit of observation is a country-month. All the estimations include country and time (yearly) fixed effects. Mean Wordcount, Freq. Publication and Market penetration are in log.
$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Controls also include: Access to electricity (% of pop.); Mobile subscriptions (%); Days to start a business; Adult literacy rate(%), L2.Market penetration
Coeff. for market penetration without controls for initial conditions: 3.52*** (0.82)(HHI) , 0.46*** (0.17)(French ratio), 1.93 (2.07) (Local ratio), 0.02 (0.04)(Wordcount), 0.19 (0.13) (Frequency Publication)

This second Generalized Method of Moments estimations still does not allow to attribute any effects of an increase in market penetration by French newspapers on the ratio of local news produced by domestic newspapers in country c, neither on the mean wordcount per articles per month.

Similarly, forgetting to include controls for initial conditions on the domestic online newspaper markets lead to a slight overestimation of the impact of the entry of french firms on the matrix of endogenous newspaper characteristics.

Limitations

As System GMM uses more instruments than Difference GMM, it is worth noting that S-GMM may not be appropriate for a dataset using a small number of countries (Sargan test may be weak when there are more instruments than countries).

Another inconvenient with this kind of Dynamic Panel Data estimators (both System and Difference GMM), is that they are highly sensitive to the particular specification of the model and its instruments.

Eventually, these GMM instrumental approaches often allow to obtain slightly more robust estimates of the impact of the endogenous regressor than under Fixed Effects alone, but they are generally not the most powerful instruments one can find.

6.2 Instrumental Variables Estimates Based on French Newspaper Articles

6.2.1 Instrument Specification

Using a Generalized Method of Moments *à la* Arellano-Bond was a safe method to generate instruments and to address the issue of endogeneity of our main regressor without introducing a new and potentially weak external instrument. With weak instruments (instruments which have a low correlation with the endogenous explanatory variable), the Fixed Effects IV estimators are indeed likely to be biased in the way of the OLS estimators. This could result in a larger variance in the coefficient, and severe finite-sample bias: "the cure can be worse than the disease" (Bound, Jaeger, Baker, 1993/1995). Nevertheless, because these GMM instruments are internal, we can see that point estimates only slightly vary compared to the OLS and Fixed Effects estimates.

We introduce in this section a new, external instrument, using characteristics of French newspaper articles. All previous regression models indeed *exclude* qualitative data collected for French newspapers. Our regression designs use circulation data for French newspapers as a main regressor (based on total visits of french newspaper websites by populations in African countries), but our set of dependent variables and all remaining covariates are restricted to francophone African digital newspapers. We nevertheless collected 96 180 articles from French online newspapers targeting francophone Africa over the whole period of the study, that we mainly used for descriptive statistics and comparison purposes. We can then make use of this external data in our baseline regressions.

We use in this section the share of French newspaper articles that mention a given African

country c from our sample of Francophone countries over the share of all remaining French articles, as an instrument for market penetration of french news. The instrument is computed as follow.

1. For each month, we compute the total number of French newspaper articles, excluded from our initial regression sample, which have for main subject a given Francophone African country c.
2. We compute for each month the total of all articles available from these French online newspapers.
3. We eventually compute, for each month and for each country c, the share of French newspaper articles which concerns this given country c.

Equation 12 describes our identification equation:

$$Y_{ct} = \beta \ln \widehat{MP_{ct}} + \delta X_{ct} + I_{ct} + \alpha_c + \gamma_t + \varepsilon_{ct} \tag{12}$$

where the first stage equation is :

$$\ln MP_{ct} = \theta References_{ct} + \delta X_{ct} + I_{ct} + \alpha_c + \gamma_t + e_{ct} \tag{13}$$

All variables are the same as previously defined, $\widehat{MP_{ct}}$ is the estimated value of MP_{ct} derived from the first-stage equation, and $References_{ct}$ is our instrument, which is a variable of the share of French newspaper articles that refers to a given Francophone African country c.

6.2.2 Tests of IV Consistency

Correlation with the Endogenous Regressor

As French digital newspapers do not physically export their articles to each country, but only add new articles on their websites targeting Africa, any article is uniformly provided to all Francophone speakers over all African countries targeted in our study. They are not restricted in exporting their production by tariffs, legal procedures to enter a country, or other physical barriers to trade. They could thus produce articles about any Francophone African country, and any of these articles would immediately be available online for each citizen with access to internet in Africa and in the world.

However, one can assume that the proportion of articles related to a given country c is correlated with the relative proportion of consumers from country c visiting the website. One can easily expect that population from Togo is not as interested about news concerning Madagascar that citizens of Madagascar. Let's suppose that a major event is happening in Madagascar over one or several months, as a domestic election, the vote of an important law, or a domestic football tournament. In the meantime, no such major events are happening in Togo. The event in Madagascar is heavily relayed in domestic newspapers, but French newspapers targeting Africa should also write relatively more about Madagascar than bout Togo over this time period. Population in Togo, located in West Africa, is not really concerned about political elections in Madagascar. As a low proportion of articles produced by french newspapers concerns Togo for these months, people from Togo should get less interested and the total number of visits from Togo should decrease. Hence, we expect a high and positive

correlation between the total number of visits from population in a given country c over a month, and the total number of articles produced by French websites and targeting this country c.

We indeed find a high correlation between these two variables, and a basic OLS regression without controls of market penetration by french newspapers in country c on the total number of references to country c gives a positive and statistically significant estimate at the 1 percent level. A one percent increase in the total number of reference to country c in a month is associated to a 33 percent increase in total visits of French newspapers by population for country c, all other things constant.

Non-Correlation with the Dependent Variables

Moreover, if the relative share of French articles mentioning country c should affect total visits by population of this country c , it is assumed to be exogenous for the matrix of newspaper characteristics in a given country c. Indeed, the matrix of outcome variables captures the characteristics of domestic digital newspapers for a given country c across the 21 Francophone countries in the sample. These domestic characteristics should not be affected by the relative proportion of French newspaper articles mentioning their country.

The correlation matrix between this instrument and our different dependent variables shows indeed no correlation. We also run OLS regressions to test the impact of the instrument on the five outcome variables, and find no causal relationships, at the exception of the frequency of publication per month. Table 9 shows that all estimates are statistically non-significant (we obtain very-high p-value, between 0.210 and 0.865), excepted for the frequency of publication.

Table 9: **OLS regressions of instrument with outcome variables**

	Norm. HHI index	**French news ratio**	**Local news ratio**	**Mean Wordcount**	**Frequency Publication**
References by	0.02	-0.02	0.30	0.01	0.19***
French newspapers (log)	(0.29)	(0.04)	(0.86)	(0.01)	(0.06)
Constant	17.38***	0.58***	61.75***	6.12***	5.31***
	(0.48)	(0.06)	(1.43)	(0.02)	(0.10)
Observations	645	645	645	645	645
Adjusted R^2	-0.002	-0.001	-0.002	-0.001	0.014

Standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The conditions required for the IV estimator to be consistent (uncorrelated with the error term, correlated with our potentially endogenous regressors) are thus satisfied for four outcome variables out of five, so that beta estimates resulting from the IV regression should be consistent estimates of this causal impact for these variables.

One could however expect that the number of publications per month in a country c be correlated with the number of reference to this country c by French newspapers. Indeed, any major event as the one we mentioned above is likely to lead to an increase in the total number of articles released by domestic newspapers. Similarly, French newspapers are more than likely to mention this event. Hence, our instrument is logically correlated with this particular dependent variable, and thus not valid for it.

Tests of Endogeneity with all Outcome Variables of Newspaper Characteristics

We made the preliminary hypothesis that market penetration by French newspapers is endogenous for the domestic newspaper characteristics across Francophone African countries. Our main concern is indeed a potential reverse causality. As previous findings seem to indicate that our instrument respects the conditions for the consistency of the IV estimators for all dependent variables *excepted one*, we test more formally endogeneity. We run several Durbin–Wu–Hausman tests for the endogeneity of regressors in order to determine whether market penetration is believed to be endogenous for all variables of newspaper characteristics.

The test consist in regressing our endogenous variable (total number of visits of French newspapers) on the instrument and the controls. Then, we store residuals of the regression, before regressing each of our dependent variable separately on the endogenous regressor, the residuals stored, and the controls. We eventually run the test on the stored residuals.

We find that OLS estimates are *not* consistent for our three indicators of newspaper content: the Normalized Herfindahl-Hirschman index, the Local news ratio and the French news ratio. For these variables, we strongly reject the null hypothesis of the exogeneity of the regressor. On the contrary, we fail to reject the null hypothesis of exogeneity for our two indicators of newspaper format: mean wordcount and frequency of publication. This indicates OLS is consistent for these two outcome variables.

To conclude, the instrument thus respect the exogeneity conditions for the three dependent variables for which we have endogeneity issues.

(i) We ruled out any direct effect of the instrument on the dependent variable, as the domestic characteristics of digital newsapers in a given African country c should not be affected by the relative proportion of French newspaper articles mentioning this country rather than another Francophone African country c'.

(ii) We can also rule out any reverse effect of these three dependent variables on the instruments. Neither the diversity of topics treated by newspapers in a given African country, nor their share of local and French news is likely to influence the total number of references to this particular country in French newspaper targeting francophone Africa. Firstly because, as explained above, most French newspaper websites decided to enter the francophone African market uniformly, and provide online articles with the aim to reach the largest possible population of Francophone speakers in Africa. Secondly, because most of them logically follow the production habits of their parent company, and produce articles with the same standards as the ones released for France. We can evenetually expect that a journalist working for a French newspaper will produce articles about Morocco or Niger in a given month because a particular event happened in Morroco or Niger, and not due to the particular characteristics of newspapers in these countries. We thus do not expect reverse causality issues between our dependent variables and the excluded instrument.

(iii) Eventually, as explained above, we find a significant influence of the instrument on the endogenous regressor.

6.2.3 Estimation Results

Results from instrumental variable regressions for our five outcome variables are displayed in Table 10. We also provide estimates for the two indicators of newspaper's format for which market penetration by french newspapers is believed to be exogenous, to compare the bias with OLS and FE estimates.

Confirmation of the Instrument's Validity and Strength

First-stage regression provides as expected a positive coefficient for the impact of the total number of references to country c on the total number of visits of French websites per citizens from this country c. This point estimates is significant at the 1 percent level. It proves the instrument is highly correlated with the endogenous regressor even after controlling for the exogenous regressors. The Angrist-Pischke first-stage F-statistics also confirms the additional instrument has a significant explanatory power (p-value equal to zero and F statistics significantly bigger than 10).

The Durbin-Wu-Hausman test provided with each full IV estimation confirms again that market penetration is endogenous for our three indicators of newspapers contents, but not for the two indicators of newspaper formats.

Several second-stage test statistics provide weak-instrument robust inference for testing the significance of the endogenous regressors in the equation being estimated. In particular, the two weak identification tests strongly reject the weakness of our instrument for the three dependent variables for which we had endogeneity issues. We obtain both a Cragg-Donald Wald F statistic and a Kleibergen-Paap rk Wald F statistic[26] superior to the threshold of ten (between 16 and 27 for our different variables). Our F statistics is also well above the different Stock-Yogo weak ID test critical values.

As our model is just-identified, the Hansen-J test of overidentifying instruments and the the Sanderson-Windmeijer (SW) first-stage chi-squared test of under-identification are not relevant. Both are indeed strongly rejected.

These different tests thus confirm that our excluded exogenous variable is a valid instrument for the three dependent variables of newspaper content, sufficiently correlated with the included endogenous regressors but uncorrelated with the error term. Moreover, this instrument is not weak. If the instrument was only weakly correlated with the endogenous regressors, the usual 2SLS estimators would be biased toward the OLS estimator, and inference based on the standard errors reported by instrument variable regression can be severely misleading. This is what seems to happen for our two outcome variables of newspapers format, which display non-significant IV estimates contrary to OLS.

Analysis of Results for the Indicators of Newspapers Content

The three endogenous indicators of newspapers content all display significant estimates at the 5 or 1 percent level following a one percent increase in market penetration by French newspapers. We find substantial differences between IV estimates and OLS and FE estimates.

[26] Kleibergen-Paap rk Wald F is the relevant test in our case as we used clustered-robust standard errors at the country level

Table 10 shows that the IV point estimate for the Normalized HHI index is positive and statistically significant at the 1 percent level. A one percent increases in market penetration of French newspapers is associated with a 10.65 percentage point increases in the index of topics concentration. This suggests the OLS estimate was severely biased downward (1.42 for OLS with Fixed Effects instead of 10.65; 2.39 for simple multivariate OLS without country and times dummies, see Table 4). We can conclude that the entry of the French digital newspapers in Francophone Africa is associated with an average decrease of the diversity of topics treated by domestic newspapers across countries. Although we cannot conclude in terms of absolute market shares[27], audience gains by French websites are strongly associated with a decreasing diversity of topics treated monthly by each country, on average.

The most important change compared to the OLS estimates concerns the Local News ratio. The IV point estimates is strongly positive and significant at the 1 percent level. A one percent increase in market penetration by French newspapers is associated with a 31.83 percentage point increases in the local news ratio, all other things constant. This means that an increase by one percent of the total number of monthly visits of French newspaper websites by population from country c is expected to increase the share of local news produced by domestic newspapers by almost 32 percentage points, on average. Here again, the OLS estimate suffered from a strong downward bias. Moreover, neither Multivariate OLS, nor Fixed Effects Tobit Estimates were significant. The OLS regression without country and time dummies gives a non-significant coefficient of -1.74 (Table 4), which decreases further to -0.64 when adding country and time dummies. Fixed Effects Tobit estimate is also of -0.59 percent. The only significant coefficient at the 1 percent level was obtained when running Multivariate OLS regression with country and time fixed effects, but without controls (a one percent increase in market penetration was correlated ith a 6.17 percentage point increase in the ratio of local news, see Table E.1 in Annexes).

Eventually, Columns 3 and 4 show the entry of French online newspapers leads to a slight and negative decrease in the French news ratio. The lack of controls for initial conditions results in a non-significant coefficients and high robust standard errors. Controlling for the past characteristics of the Francophone African newspaper markets from 2005 to 2013 leads to a significant coefficient at the 5 percent level. A one percent increase in the total number of monthly visits of French websites is associated with a 0.46 percentage point decrease in the total share of domestic news talking about France. This point estimate seems small compared to the coefficients of the HHI and the Local News ratio, but it is rather high compared to the small share of news related to France in francophone African newspapers. The average share of news about France in domestic newspapers is indeed 0.474 percent across our panel, with high variations across countries (standard deviation of 0.91). This monthly share varies between 0 and 11 percent, and the median share is only of 0.185 percent (see Table A.1). A 0.46 percentage point decreases thus implies that the share of articles dedicated to France drops nearly to zero on average across our panel.

[27]While the total readerships of French newspapers increases, we could also have a parallel increase of the total number of media across Francophone African markets. We could also observe an homogenous or more than proportional increase of the audience for domestic digital newspapers. As we do not have exhaustive data about the state of the Francophone Africa landscape, we do not talk about market share of French newspapers.

Several potential explanations can be drawn from for these coefficients. When looking at the descriptive statistics of the characteristics of French newspaper dedicated to Francophone Africa (i.e. exclusively the "African" version of each french newspaper), we find a very low share of local news: 24 percent in average, against 50.32 percent in average among domestic francophone African articles (see Table A.2).

The share of news dedicated to France is however lower in French articles than in Francophone African articles (0.34 percent against 0.47 percent), but this is easily explained by the fact the African version of French newspaper websites are supposed to exclusively target Africa and francophone Africa. A small proportion of articles dedicated to France seems nevertheless to be published on these websites.

We could thus think that Francophone African newspapers increase the proportion of local news in response to the very low proportion of French articles targeting very local topics. Many critics made by Francophone African citizens on the websites of French newspapers indeed underline the partiality of journalistic contents, which are perceived as unrepresentative of the subtleties of each country. Local news produced by French newspapers are regularly criticized as a weak copy-paste of international programs, with very few substantial qualitative analysis and not particularly inventive topics. The relatively small place accorded to local news in French newspapers could thus explain the strong causal impact of their entry on the ratio of local news domestically produced. It could also explain the small decrease in the ratio of news related to France in domestic newspapers. Eventually, it could explain the 10 percentage point increase in the concentration of articles topics. Figures B.1 and B.2 display a relatively stable increase of topics related to *domestic politics* or *society/community and lifestyle* and a relative decrease in topics referring to *international relations and global issues*, or to *corporate and industrial news*. It could thus be that the entry of foreign media players resulted in narrowing the focus of Francophone African newspapers on more domestic news, in reaction to the relative homogenization of content in foreign media.

Analysis of Results for the Indicators of Newspapers Format

In opposition to the three endogenous indicators of newspaper content, the two indicators of newspaper format for which market penetration was tested as exogenous indeed display no significant instrumental variable estimates. Moreover, the beta coefficients for market penetration do not show substantial variations in amplitude compared to the multivariate OLS case. This seems to confirm that OLS is a consistent estimate for these two outcome variables, and that any potential impact of French newspapers entry on the format of online domestic newspapers is weak. We remind that based on OLS estimates with country and time fixed effects, mean monthly wordcount decreases by 0.05 percent and the frequency of publication increases by 0.27 percent, both at the 5 percent significance level, following a one percent increase in visits of french newspaper websites. The Durbin-Wu-Hausman tests of endogeneity of the full IV specification indicates indeed again that OLS seems to provide consistent estimates of the causal impact for these two variables.

Table 10: **Instrumental Variable Estimates for Endogenous Regressors, monthly**

	Indicators of newspapers content						*Indicators of newspapers format*			
	Normalized HHI		**French News Ratio**		**Local News Ratio**		**Freq. Publication**		**Mean Wordcount**	
	without i.c	i.c	without i.c	i.c	without i.c	i.c	without i.c	i.c	without i.c	i.c
Market penetration (log)	15.76***	10.65***	-0.61	-0.46**	29.71***	31.83***	-0.06	0.52	0.08	0.11
	(5.40)	(3.88)	(0.48)	(0.22)	(10.95)	(11.15)	(0.43)	(0.32)	(0.13)	(0.14)
French speakers (log)	-2.76	-1.24	0.20	0.23**	-3.80	-4.39	0.38**	0.24*	-0.23***	-0.23***
	(2.43)	(1.65)	(0.21)	(0.10)	(4.83)	(4.98)	(0.17)	(0.12)	(0.05)	(0.05)
Log GDP pc(current USD)	12.91**	7.13	-0.61	-1.24***	45.38***	47.42***	-1.25**	-0.74*	0.34**	0.36**
	(6.08)	(4.64)	(0.60)	(0.35)	(14.81)	(15.40)	(0.55)	(0.43)	(0.14)	(0.14)
Days to start a business	0.27	-1.48***	0.06	0.00	-4.17***	-3.50***	-0.32***	-0.15***	0.03**	0.03**
	(0.54)	(0.50)	(0.08)	(0.03)	(1.31)	(1.31)	(0.07)	(0.04)	(0.01)	(0.01)
Broadband subscriptions (%)	6.00**	1.62	-0.16	0.08	1.44	3.30	-0.24	0.28*	-0.15**	-0.12**
	(2.76)	(1.97)	(0.22)	(0.14)	(6.24)	(6.34)	(0.20)	(0.16)	(0.06)	(0.06)
Internet users (%)	-0.74	-0.51*	0.02	-0.05***	0.24	0.12	-0.07**	-0.11***	0.04***	0.03***
	(0.45)	(0.29)	(0.03)	(0.02)	(0.88)	(0.88)	(0.03)	(0.02)	(0.01)	(0.01)
Mobile subscriptions (%)	0.39***	0.25**	-0.02	0.01	-0.20	-0.13	0.01	0.03**	0.00	0.01
	(0.15)	(0.12)	(0.02)	(0.01)	(0.37)	(0.36)	(0.02)	(0.01)	(0.01)	(0.01)
Scores for components 1 & 2	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	645	645	645	645	645	645	645	645		
adj. R^2	-0.382	0.460	0.150	0.874	-0.541	-0.604	0.222	0.665	0.148	0.168

Notes: Robust standard errors in parentheses are clustered by country. The unit of observation is a country-month. All the estimations include country and time (yearly) fixed effects. Normalized HHI is an index from 0 to 100. Mean Wordcount, Freq. Publication and Market penetration are in log.
$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Columns *i.c* provide point estimates and robust standard errors with additional controls for initial conditions; without i.c do not include control for initial conditions.
Controls also include: Access to electricity (% of pop.); Adult literacy rate, pop. 15+ (%); Urban population (% of total); CPI (2010 = 100)

7 Concluding Remarks and Directions for Future Work

Our empirical analysis offered a first overview of the impact of the entry of foreign media firms on the characteristics of the domestic information products available to consumers in the digital newspaper market. Based on the estimation of the impact of market penetration by French newspaper firms in Francophone Africa, we find a significant increase of the concentration of digital newspaper topics over the period 2014-2017 (and thus an average decrease of the diversity of subjects treated by Francophone African newspapers), an important increase of their share of local news, and a small decrease of the total number of articles related to France. This results are obtained while controlling for the initial conditions of these indicators on the digital newspaper markets for each African country, over the period 2005-2013. Our evidences suggest that the entry of French digital newspapers, which produce a relatively low share of local news and whose websites are generally targeting uniformly all Francophone African countries, could lead to a decrease of the relative variety of news produced across Francophone Africa. A possible scenario is a reallocation of newspaper topics between French newspapers and domestic newspapers, characterized by a strong and significant increase of the number of domestic articles focused on local news, and by French newspapers that would relay more general information, indifferently related to all African countries.

We also find a slight but significant impact of the penetration of French digital newspapers on two indicators of newspaper format, namely a small increase of the average frequency of publication of Francophone African newspapers, and a very small decrease of the mean wordcount per articles. The specific format of the news displayed on the web, generally significantly shorter than printed articles and with a higher publication rate, could explain the weakness of the effect.

This research opens the way to many possible extensions.

A first extension would be to build and to estimate a structural model for explaining the qualitative transformation of digital newspaper characteristics following foreign entry. Several theoretical extension of this work could be considered. The stylized facts of this empirical analysis could serve as a basis to derive parameters for an adapted version of the Melitz (2003) model. This model would preserve the main features of Melitz (2003) and allow for endogenous markups, as in the dynamic industry model of trade reallocation of Melitz and Ottaviano (2005), by giving up CES preferences. Such features could be achieved by introducing consumers with Cobb-douglas preferences and a discrete choice framework (e.g., Anderson, de Palma and Thisse, 1992; Feenstra, 2004). Anderson et al. (1992) indeed show that if goods are consumed in continuous units, a discrete choice model generates the same market demand functions as a representative consumer with CES preferences. This adapted Melitz model could be augmented in several ways. Firstly, total circulation could be larger than the number of households, due to the fact that online newspapers are public goods (non-exclusive, non-rival products). Secondly, utility would be "quality"-augmented to control for endogenous product characteristics. Thirdly, we could assume news are subject to cultural discount (as in Waterman, 2005). It would eventually be interesting to extend the research by a welfare analysis, and assess the welfare consequences of free entry on the online newspaper market (based on insights by Berry and Waldfogel, 1999).

The second and most straightforward extension would be to carry out an event analysis, at the country-level. Our results are indeed subjects to important variations across countries and over time. Firstly, for some countries, the entry of French firms is significantly associated with a increase of the concentration of newspapers topics by month, that goes in hand with a strong increase of the share of news dedicated to local topics. But some other countries seem to display the inverse trend. Although we controlled for many variables related to demographics, internet access, political systems and measures of freedom of the press, it would be interesting to carry out a country-per-country analysis and look more precisely at the impact of each of these variables. Secondly, each country internally displays substantial day-to-day variations in the values of our five indicators, with peaks and falls at some given dates. The most discussed categories of news topics also strongly vary. Carrying on a precise event analysis for each country, based on the frequency of the apparition of topics over time, could allow to shed lights on what drives these variations. It would be especially interesting to determine whether main historical events find repercussions in the press, and influences the characteristics of the media produced, beyond the entry of foreign players. The Arab Spring who started in 2011 is for instance likely to impact the reception and consumption of French online articles by domestic populations. It is also more than likely to influence the share of digital articles directly mentioning North African countries over that period.

A third extension would be to enlarge the sample in the long-run. Firstly, in terms of coverage for Francophone Africa, by crawling and storing more articles from a more exhaustive panel of newspaper websites. Secondly, by extending the scope of the analysis to anglophone countries, or even to local African languages. Many media companies from the United Kingdom are active in Africa, as the *BBC, CNBC* or *The Guardian*. In the case of English speakers, the impact may however be more difficult to asses, as the sample of exporting countries is extremely large. Capturing flows of news from every foreign English newspapers would surely be a challenge, but an interesting one. Eventually, new players on the African media markets are not only French or English speaking countries, and do not produce news in French and English exclusively. Chinese firms, as *CCTV*, started investing massively in the market, producing news in English but also in Swahili,. CCTV moreover plans to extend its production to French articles in the short-run. *Anadolu* from Turkey, and *Al-Jazeera*, also turned to the African media markets in the last couple of years.

References

- Anderson, Simon, and Jean Jaskold Gabszewicz (2005). "The Media and Advertising: A Tale of Two-Sided Markets," *CEPR Discussion Papers* 5223.
- Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. R*eview of Economic Studies* 58: 277–297.
- Arellano, M., and O. Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68: 29–51.
- Berry, Steven and Joel Waldfogel (2001), "Do mergers increase product variety? Evidence from radio broadcasting." *The Quarterly Journal of Economics*, 116, 1009–1025.
- Berry, S.T., Haile, P., 2010. "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers". *Cowles Foundation, Yale University.* Discussion Paper 1718.
- Blundell, R., and S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87: 115–143.
- Campbell, Jeffrey R. and Hugo A. Hopenhayn. "Market Size Matters," *Journal of Industrial Economics*, 2005, v53(1,Mar), 1-25.
- Chu, Chenghuan Sean (2010), "The effect of satellite entry on cable television prices and product quality." *RAND Journal of Economics*, 41, 730–764.
- Disdier, A-C., K. Head and T. Mayer (2009), "Exposure to Foreign Media and Changes in Cultural Traits: Evidence from Naming Patterns in France". *Journal of International Economics* 80, 226-238.
- Disdier, AC., S. Tai, L. Fontagne, L. and T. Mayer (2010), "Bilateral Trade of Cultural Goods" *Review of World Economy* 145, 575-595.
- Dixit, Avinash and Joseph Stiglitz (1977), "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67:297—308.
- Fan, Y., "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market", *American Economic Review*, 103(5), August 2013, 1598-1628
- Feenstra, R. C., 2004. Advanced International Trade. Theory and Evidence. Princeton: *Princeton University Press.*
- Gabszewicz et Anderson, 2005, "The Media and advertising : a tale of two-sided markets", *CORE Discussion paper* 2005/88.
- Gentzkow, Matthew and Jesse Shapiro (2010), "What drives media slant? Evidence from U.S. daily newspapers." *Econometrica*, 78, 35–71.
- George, L., Waldfogel, J., 2003. "Who affects whom in Daily Newspaper Markets?" *Journal of Polital Economy* 111, 765–768.
- Hanson, G.H., and Xiang, C., 2004. The Home Market Effect and Bilateral Trade Patterns. *American Economic Review* 94(4): 1108-1129.

- Hanson, Gordon H. and Chong Xiang. 2008. "Testing the Melitz Model of Trade: An Application to US Motion Picture Exports." *NBER Working Paper* 14461.
- Hausman, J.A., 1996. "Valuation of new goods under perfect and imperfect competition". In: Bresnahan, T., Gordon, R. (Eds.), *The Economics of New Goods*. In: *Studies of Income and Wealth*, vol. 58. National Bureau of Economic Research, Chicago, IL.
- Head, K., Mayer,T. and Ries, J. , 2010, "The erosion of colonial trade linkages after independence" *Journal of International Economics*, 81(1):1-14. (formerly CEPII discussion paper 2008-27)
- Honoré, Bo E. (1992): "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, Vol. 60, No. 3, Pages 533-565.
- Hopenhayn, Hugo (1992), "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60: 621-653.
- Hsiang Iris, George Sylvie (2001) "The Medium is Global, the Content is Not: The Role of Geography in Online Newspaper Markets", *Journal of Media Economics*, 14:4, 231-248,
- Lacy, S., Picard, R. G. (1990). "Interactive monopoly power in the daily newspaper industry". *The Journal of Media Economics*, 3(2), 27–38.
- Marvasti, Akbar (1994) "International Trade in Cultural Goods: A Cross-Sectional Analysis". *Journal of Cultural Economics* 18: 135–148.
- Marvasti, A. and E.R. Canterbury (2005), "Cultural and Other barriers to Motion Pictures Trade". *Economic Inquiry*, 43(1), 39-54.
- Masood Maria . New Evidence on Development and Cultural Trade: Diversification, Reconcentration and Domination. *Preliminary work*. 2013.
- Mayer, T. and Zignago, S. (2011). Notes on CEPIIs distances measures: The GeoDist database. *Working Paper* 2011-25. CEPII.
- Melitz, Marc J. (2003), "The Impact of Trade on Intra- Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71:6, pp. 1695-1725
- Melitz, Marc and Gianmarco Ottaviano (2008), "Market Size, Trade and Productivity", *Review of Economic Studies*, Blackwell Publishing, vol. 75(1), pages 295-316, 01.
- Ottaviano, G. I. P., Tabuchi, T. and Thisse, J.-F. (2002), "Agglomeration and Trade Revisited", *International Economic Review*, 43, 409–436.
- Rauch, James E., and Trindade, Vitor. 2009. "Neckties in the Tropics: A Model of International Trade and Cultural Diversity." *Canadian Journal of Economics*, 42: 809-843.
- Rochet, Charles, and Jean Tirole (2003). "Platform Competition in Two-Sided Markets" *Journal of the European Economic Association*, Vol. 1, No. 4, pp. 990-1029
- Roodman, D. M. 2009. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71: 135–158.

- Roodman, D. 2009. How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata. *Stata Journal* 9(1): 86-136.
- Rosse, James N. (1967), "Daily newspapers, monopolistic competition, and economies of scale." *American Economic Review*, 57, 522–533.
- Rosse, James N. (1970), "Estimating cost function parameters without using cost data: Illustrated methodology." *Econometrica*, 38, 256–275.
- Shiller B., Waldfogel J. and Ryan J. (2017), "Will Ad Blocking Break the Internet?" *NBER Working Paper* No. 23058. Issued in January 2017. NBER Program(s): IO.
- Spence, M., "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies* XLIII, (1976), 217-236.
- Spence, M. and Bruce Owen (1977) "Television Programming, Monopolistic Competition, and Welfare," *Quarterly Journal of Economics*, Vol. 91, No. 1, pp. 103-26.
- Waldfogel, J., 2003. Preference externalities: an empirical study of who benefits whom in differentiated- product markets. *RAND J. Econ.* 34, 557–568.
- Wildman, Steven and Siwek, Stephen (1988) International Trade in Films and Television. *American Enterprise Institute/Ballinger Publications, Cambridge, Mass.*
- Wooldridge, J.M. 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: *MIT Press.*

Appendix

Appendix A Summary Statistics

Table A.1: **Summary Statistics of Endogenous Newspapers Characteristics, panel**

	Panel level	Mean	Sd	Min	Max	Obs
Concentration of news topics (norm. HHI index)	overall	20.13747	8.988764	11.54127	100	1024
	between	.	3.963683	14.72494	28.49235	21
	within	.	8.36389	6.001267	102.8065	51.2
Local news ratio (monthly)	overall	50.45381	28.01225	0	100	1025
	between	.	22.58776	14.45198	88.49975	21
	within	.	15.94878	-27.99806	105.4749	51.25
French news ratio (monthly)	overall	.4735824	.9085319	0	11	1025
	between	.	.4019836	.0890232	1.439624	21
	within	.	.8385325	-.6677709	10.75463	51.25
Mean wordcount (monthly)	overall	403.2892	167.7959	74.72	1333.4	1025
	between	.	119.6218	253.5255	754.9741	21
	within	.	121.1688	-8.266172	1151.917	51.25
Frequency of publication (monthly)	overall	553.1524	739.5291	1	5521	1025
	between	.	500.5393	46.25	2130.281	21
	within	.	536.0138	-1143.866	4805.646	51.25

Table A.2: **Summary Statistics of French Newspapers Characteristics**

	Mean	Median	Sd	Min	Max	Obs
Concentration of news topics (norm. HHI index)	17.96	16.84	3.88	13	38	139
Local news ratio (monthly)	24.17	24.50	5.23	5	36	139
French news ratio (monthly)	0.34	0.20	0.73	0	8	139
Mean wordcount (monthly)	436.34	421.63	107.74	273	780	139
Frequency of publication (monthly)	226.42	168.00	144.76	5	602	139

Table A.3: **Summary Statistics of the main regressors and country demographics**

	Panel level	Mean	Sd	Min	Max	Obs
Market penetration (Total visits french newspapers, log total)	overall	13.33926	1.097362	9.458449	16.22746	798
Market penetration (Total visits french newspapers, absolute values)	overall	1095794	1336464	12816	1.12e+07	798
French speakers (log)	overall	15.29558	.8983795	13.41354	17.43115	1333
Adult literacy rate, both sexes (%)	overall	60.49127	19.56704	15.4567	86.94787	1333
Education, lower secondary	overall	27.37487	15.17393	2.03596	47.30499	606
Education, upper secondary	overall	13.32373	8.099138	.49505	25.41652	542
Population median age	overall	20.91978	4.399079	15	30.3	1333
Log GDP per capita (current US$)	overall	7.170628	.9917896	4.947448	9.35272	1333
Urban population (% of total)	overall	47.11184	17.61019	9.375	87.156	1333
Consumer price index (2010 = 100)	overall	114.6995	22.38034	42.23554	186.8617	1333
GINI index (World Bank)	overall	41.63281	5.860921	31.45	56.24	1113
Log weighted distance with France (pop-wt, km)	overall	8.131755	.5612249	7.117863	9.057404	1333
Days to start a business	overall	20.04801	19.65685	4	158	1333
Fixed broadband subscriptions (%)	overall	1.058845	1.499145	.0006665	5.566853	1333
Internet users (%)	overall	16.00534	15.54146	.2213414	57.08	1333
Mobile cellular subscriptions (%)	overall	81.42835	38.54379	1.969013	171.375	1333
Access to electricity (% of population)	overall	53.76309	36.76041	3.5	100	1333

Table A.4: **Descriptive Statistics of Internet penetration per Cross-Sectional Unit (from 2015)**

Country	Broadband subscription (%)	Internet users (%)	Mobiles subscription (%)	Access to electricity	Trade in services(%)
Central Africa					
Cameroon	0.07	20.68	71.85	53.70	15.83
Central Afr. Rep.	0.01	4.56	20.45	10.80	.
Congo republic	0.01	7.62	111.66	41.60	45.83
Congo (RDC)	0.00	3.80	52.99	16.40	6.24
Guinea	0.01	4.70	87.17	26.20	12.84
Chad	0.08	2.70	40.17	6.40	.
East Africa					
Burundi	0.03	4.87	46.22	6.50	13.30
Madagascar	0.07	4.17	46.02	15.40	24.51
Rwanda	0.17	18.00	70.48	18.00	20.20
West Africa					
Benin	0.67	6.79	85.64	38.40	14.05
Burkina faso	0.04	11.39	80.64	13.10	14.37
Ivory Coast	0.52	21.00	119.31	55.80	13.29
Gabon	0.63	23.50	168.92	89.30	12.47
Mali	0.02	10.34	139.61	25.60	20.18
Niger	0.06	2.22	46.50	14.40	14.69
Senegal	0.67	21.69	99.95	56.50	17.12
Togo	0.92	7.12	64.95	31.46	20.41
North Africa					
Algeria	5.57	38.20	113.03	100.00	8.71
Egypt	4.52	35.90	110.99	100.00	13.07
Morocco	3.38	57.08	126.87	100.00	22.45
Tunisia	4.34	48.52	129.93	100.00	17.09

Table A.5: **Descriptive Statistics of Freedom Indexes per Cross-Sectional Unit**

Country	Days to start a business	Press freedom index	Property right index	Ethics and corruption	Government efficiency	Legal right index
Central Africa						
Cameroon	17.84	41.59	3.92	2.84	3.45	6.00
Congo republic	65.83	36.73	.	.	.	.
Congo (RDC)	29.69	52.67	.	.	.	.
Central Afr. Rep.	22.09	36.12	.	.	.	.
Guinea	13.51	33.15	2.49	2.15	2.77	6.00
Chad	60.10	39.66	2.91	2.22	2.80	6.00
East Africa						
Burundi	5.68	55.78	2.72	2.46	2.89	2.00
Madagascar	13.36	26.71	3.04	2.51	2.76	1.00
Rwanda	7.22	54.11	5.41	5.53	5.39	11.00
West Africa						
Benin	16.22	30.32	3.83	2.75	3.33	6.00
Burkina faso	14.12	23.85	.	.	.	.
Ivory Coast	12.04	30.42	4.14	3.82	4.18	6.00
Gabon	51.07	34.83	3.95	3.23	3.39	6.00
Mali	10.96	38.27	3.69	3.09	3.56	6.00
Niger	16.23	27.21	.	.	.	.
Senegal	8.18	26.72	4.08	3.45	3.99	6.00
Togo	19.76	30.75	.	.	.	.
North Africa						
Algeria	23.11	42.83	3.59	3.01	3.27	2.00
Egypt	8.86	55.78	3.71	3.46	3.26	2.00
Morocco	11.47	42.41	4.47	3.82	3.74	2.00
Tunisia	11.00	32.22	.	.	.	.

Table A.6: **Descriptive Statistics of Demographics per Cross-Sectional Unit**

Country	French speakers (%)	Literacy adults (%)	Education low. sec.	Education upp. sec.	Median age	GPD pc (log)	Urban pop. (% total)	CPI (2010 ref.)	GINI index
Central Africa									
Cameroon	40.00	74.55	36.23	18.17	18.28	7.14	53.53	109.17	46.34
Congo	58.00	79.31	.	.	18.80	7.77	64.75	111.45	48.91
Central Afr. Rep.	40.00	36.75	.	.	19.68	5.84	39.68	149.61	56.24
Congo (RDC)	47.00	75.96	42.45	19.85	17.38	6.03	41.78	122.20	42.10
Guinea	24.00	29.81	.	.	18.59	6.23	36.52	164.73	34.09
Chad	13.00	39.53	.	.	15.78	6.76	22.33	111.12	43.32
East Africa									
Burundi	8.00	85.62	.	.	17.61	5.56	11.66	140.55	33.36
Madagascar	20.00	64.64	.	.	18.38	6.02	34.22	128.80	42.35
Rwanda	6.00	70.48	12.18	8.42	18.18	6.48	27.42	115.79	50.56
West Africa									
Benin	35.00	38.03	.	.	18.38	6.68	43.37	107.39	43.44
Burkina faso	22.00	35.61	8.34	2.81	17.09	6.42	28.64	105.93	35.73
Ivory Coast	35.23	42.92	.	.	18.99	7.24	53.12	108.27	43.18
Gabon	61.00	83.09	.	.	20.69	9.14	86.70	106.74	42.18
Mali	17.00	32.67	11.43	6.20	16.31	6.61	38.76	107.46	33.26
Niger	13.00	18.76	.	.	15.01	5.92	18.41	103.95	34.14
Senegal	29.00	50.96	12.87	7.41	18.06	6.90	43.01	103.48	40.28
Togo	39.00	65.43	19.07	9.22	18.89	6.34	39.24	107.34	45.88
North Africa									
Algeria	33.00	77.75	43.61	22.58	26.72	8.40	67.77	104.68	.
Egypt	3.00	74.70	.	.	25.11	8.06	43.07	132.01	.
Morocco	31.00	64.34	.	.	26.80	7.93	57.94	100.02	40.72
Tunisia	54.00	80.65	39.44	.	30.21	8.31	66.55	118.12	35.84

Table A.7: **Market Penetration by French Newspapers, per Cross-Sectional Unit over the period 2014- February 2017**

Country	Total visits	Total unique users	Total page views	Population
Central Africa				
Cameroon	31.43	107.82	12.79	23.26
Congo republic	22.56	79.55	9.73	4.62
Congo (RDC)	201.94	571.01	98.91	77.92
Central Afr. Rep.	37.85	112.76	18.70	5.01
Guinea	33.91	137.32	13.70	8.68
Chad	6.78	19.16	3.20	13.13
East Africa				
Burundi	10.36	29.90	4.12	11.03
Madagascar	7.99	26.04	3.73	24.09
Rwanda	6.63	16.27	2.69	11.92
West Africa				
Benin	14.05	45.86	5.50	10.74
Burkina faso	31.57	95.60	10.84	18.24
Ivory Coast	63.41	205.93	26.00	22.69
Gabon	13.50	42.92	5.09	1.45
Mali	18.50	62.31	7.75	17.26
Niger	16.00	58.43	5.38	19.20
Senegal	54.38	189.06	18.65	14.09
Togo	9.76	32.44	3.48	7.32
North Africa				
Algeria	73.24	219.73	46.20	36.62
Egypt	11.06	21.23	7.36	88.10
Morocco	113.82	369.43	49.18	32.38
Tunisia	66.29	173.53	43.12	11.01

Note: All figures are in million.

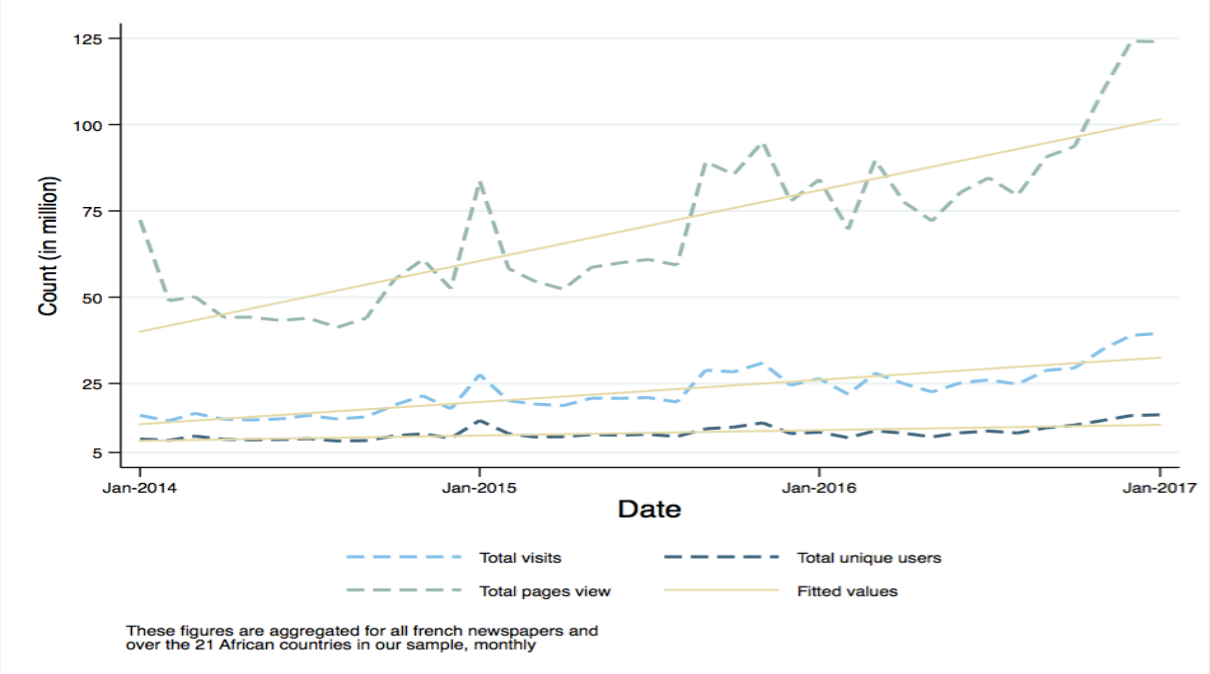

Figure A.1: Market penetration of digital french newspapers in francophone Africa
Aggregation over all francophone African countries, 2014-2017

Appendix B Overview of the distribution of newspaper topics per country

	Burundi		Benin		Burkina		Central Afr.		Ivory Coast		Cameroon		Congo	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Arts/entertainment	36	5	602	6	2,013	4	90	1	3,663	3	3,613	4	2,105	10
Corporate/ industrial news	54	7	997	10	4,104	9	261	4	10,838	10	8,499	10	1,416	7
Crime/legal action	37	5	713	7	3,632	8	600	8	8,184	8	7,242	8	972	5
Domestic politics	236	31	3,681	38	17,411	36	2,517	35	33,667	31	22,221	26	4,778	24
Financial and economic news	22	3	191	2	981	2	97	1	5,335	5	2,470	3	507	3
Health	43	6	207	2	1,320	3	173	2	3,043	3	2,155	3	860	4
International relations/ global issues	101	13	612	6	4,202	9	764	11	10,045	9	7,079	8	2,031	10
Risk news and disasters	30	4	442	5	3,649	8	1,576	22	8,034	7	6,745	8	1,342	7
Sciences/humanities and technologies	15	2	40	0	423	1	14	0	983	1	436	1	100	0
Society/community and lifestyle	137	18	1,244	13	5,419	11	876	12	15,013	14	10,598	12	2,513	12
Sports	39	5	922	10	5,075	11	205	3	9,524	9	14,204	17	3,647	18

	Algeria		Egypt		Gabon		Guinea		Morocco		Madagascar		Mali	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Arts/entertainment	11,870	10	343	4	269	3	20	2	3,432	7	10	5	4,440	2
Corporate/ industrial news	13,622	12	599	6	1,370	16	76	7	8,941	18	37	17	11,609	6
Crime/legal action	6,000	5	279	3	418	5	107	10	3,022	6	21	10	22,715	12
Domestic politics	22,328	19	3,122	32	2,625	30	492	47	10,974	22	30	14	60,340	32
Financial and economic news	3,767	3	450	5	646	7	21	2	2,349	5	19	9	3,535	2
Health	2,357	2	55	1	238	3	17	2	1,049	2	5	2	3,965	2
International relations/ global issues	8,500	7	989	10	908	10	50	5	4,948	10	16	7	19,555	10
Risk news and disasters	11,685	10	1,957	20	408	5	71	7	5,131	10	19	9	29,038	15
Sciences/humanities and technologies	674	1	23	0	28	0	5	0	271	1	3	1	1,199	1
Society/community and lifestyle	12,251	11	965	10	829	9	153	15	6,287	13	40	18	20,959	11
Sports	22,505	19	996	10	996	11	42	4	3,578	7	19	9	10,955	6

	Niger		Congo (Rep.Dem)		Rwanda		Senegal		Chad		Togo		Tunisia	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Arts/entertainment	41	3	1,967	2	62	2	2,613	3	206	3	1,648	3	6,849	6
Corporate/ industrial news	91	8	7,984	9	177	4	9,309	10	556	8	5,143	9	13,282	12
Crime/legal action	37	3	8,715	9	652	16	6,417	7	762	11	4,198	7	8,228	7
Domestic politics	541	46	24,860	27	1,099	27	26,292	28	2,067	30	18,097	32	24,916	22
Financial and economic news	25	2	2,191	2	48	1	3,900	4	171	2	1,514	3	4,944	4
Health	20	2	3,612	4	60	1	2,934	3	171	2	2,235	4	1,117	
International relations/ global issues	170	14	9,467	10	327	8	9,749	10	744	11	6,694	12	7,111	6
Risk news and disasters	38	3	11,140	12	959	24	7,804	8	957	14	4,284	8	12,868	12
Sciences/humanities and technologies	2	0	273	0	5	0	256	0	26	0	155	0	470	
Society/community and lifestyle	129	11	8,793	9	385	10	10,012	11	1,000	15	9,029	16	9,761	9
Sports	83	7	14,522	16	231	6	14,567	16	219	3	3,875	7	21,411	19

	France	
	Freq.	%
Arts/entertainment	2,538	7
Corporate/ industrial news	3,939	11
Crime/legal action	2,359	7
Domestic politics	9,522	26
Financial and economic news	728	2
Health	948	3
International relations / global issues	2,333	6
Risk news and disasters	6,430	18
Sciences/humanities and technologies	170	0
Society/community and lifestyle	6,089	17
Sports	1,221	3

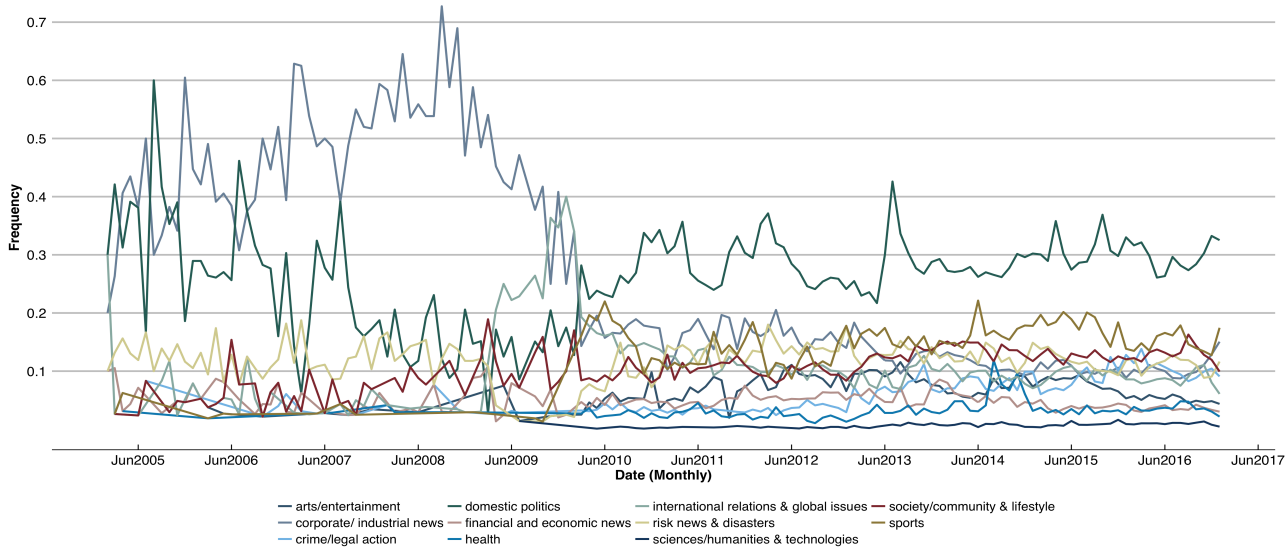

Figure B.1: Relative Frequency of topics by month
Average over all francophone African countries, 2005-2017

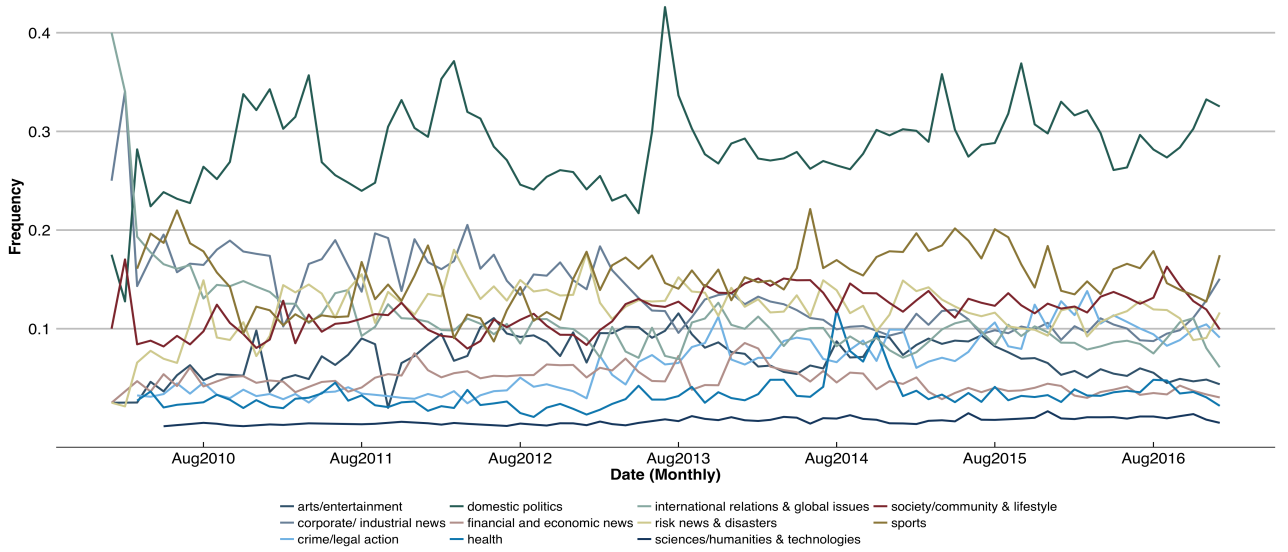

Figure B.2: Relative Frequency of topics by month
Average over all francophone African countries, 2010-2017

Table C.1: **Newspapers and total number of articles analyzed per country**

Region	Country	Iso codes	Newspapers	Total Articles	Journalists
Central Africa	Central African Rep.	CAF	Journal de Bangui.com	6322	1
	Cameroon	CMR	Camereco.com Cameroon Tribune Cameroon-Info.net Journal du Cameroun.com	17014	0
	Congo	COG	Journal de Brazza.com La Semaine Africaine Les Depeches de Brazzaville	20305	471
	Guinea	GIN	Guineenews Africaguinee.com Guinee7.com Aminata Guinee58.com Journal de Conakry.com Guinee Matin	4793	1
	Congo (RDC)	RDC	Afrikarabia Agence Congolaise de Presse GROUPELAVENIR.CD Journal de Kin.com L'Avenir LAvenir Le Soft International Radio Okapi	24338	0
	Chad	TCD	Alwihda Journal du Tchad.com	8441	3
East Africa	Burundi	BDI	Le Renouveau	941	3
	Madagascar	MDG	L'Express de Madagascar	590	1
	Rwanda	RWA	Rwanda News Agency	3505	45

Appendix C Overview of the sample of newspapers used for each country

Region	Country	Iso codes	Newspapers	Total Articles	Journalists
West Africa	Benin	BEN	Agence Benin Presse Fraternite Journal du Benin.com La Nouvelle Tribune actubenin.com Agence d'Informations du Burkina	3832	49
	Burkina faso	BFA	Burkina 24 Journal du Faso.com L'Observateur Paalga Le Pays LeFaso.net Sidwaya Sidwaya Quotidien @bidj@n.net	11562	0
	Ivory Coast	CIV	Agence Ivoirienne de Presse Connection Ivoirienne Fraternite Matin Ivoireco.com Journal d'Abidjan.com Le Patriote Lintelligent dAbidjan RTI	26599	0
	Gabon	GAB	Agence Gabonaise de Presse Gabon Eco Gabon Economie Gabon Review Gabonews Infos Plus Gabon Agence Gabonaise de Presse Journal du Gabon.com L'Union	8311	4
	Mali	MLI	Journal du Mali.com LEssor Mali Actu MaliActu.net MaliWeb Le Sahel	30759	0
	Niger	NER	Journal du Niger.com Le Sahel Tamtaminfo ACTUNET.sn	1059	3
	Senegal	SEN	ACTUSEN.com APANEWS Agence de Presse Senegalaise DAKARACTU.com Gouvernement du Senegal Journal du Senegal.com LEUK SENEGAL Le Journal du Pays Le Quotidien Le Soleil Ouestaf.com Rewmi.com SeneNews.com Sud Quotidien Walfadjri 5-Plus Dimanche	134206	38
	Togo	TGO	Journal du Togo.com Lome Infos Radio Lome Savoir News Togoactualite.com aLome.com aLome.com Republic of Togo	15670	0

Region	Country	Iso codes	Newspapers	Total Articles	Journalists
North Africa	Algeria	DZA	Algerie-Focus L'expression Algerie Presse Service El Moudjahid Reporters Le Quotidien d'Oran Reflexion El Mihwar TSA La Nouvelle Republique Reporters Le Jeune Independant Seybouse Times Liberte El Moudjahid El Watan Horizons Maghreb Confidentiel	156280	327
	Egypt	EGY	Le Progres Egyptien Watani	9422	22
	Morocco	MAR	L'Economiste Al Bayane H24info Yabiladi Liberation La vie eco Medias24 La Nouvelle Tribune Le Matin Lemag.ma Maghreb Confidentiel eMarrakech	71509	74
	Tunisia	TUN	African Manager Le Renouveau Realites Leconomiste Maghrebin Gnet L'Expression La Presse de Tunisie Le Temps Maghreb Confidentiel Webmanagercenter.com	179376	288
Europe	France	FRA	AfrikiPresse Slate Afrique Jeune Afrique La Croix (Urbi & Orbi Africa) Le Monde Afrique RFI France24 Le Huffington Post Le Point Afrique	96180	2305

Appendix D Evolution of newspapers characteristics in Francophone Africa

Average over all francophone African countries, 2005-2017

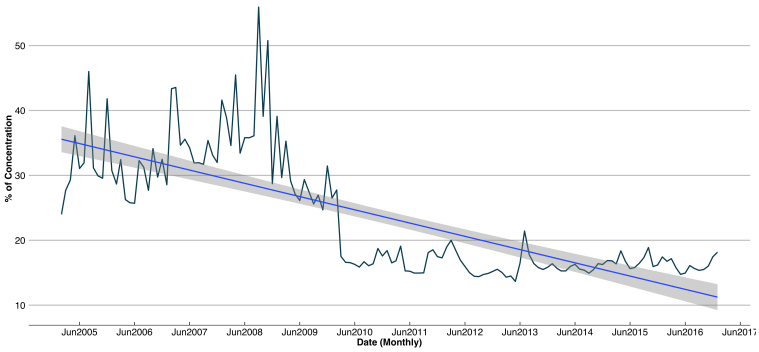

Figure D.1: Herfindahl HIrschman index, monthly

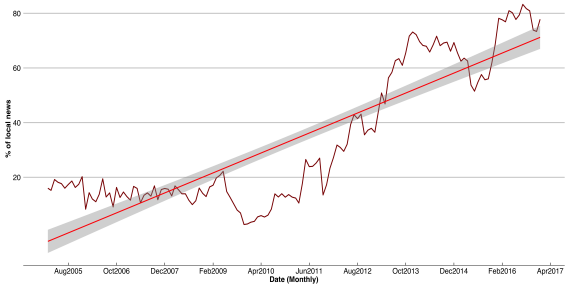

Figure D.2: Local news ratio, monthly

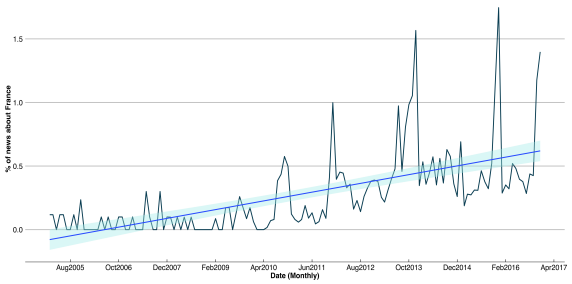

Figure D.3: French news ratio, monthly

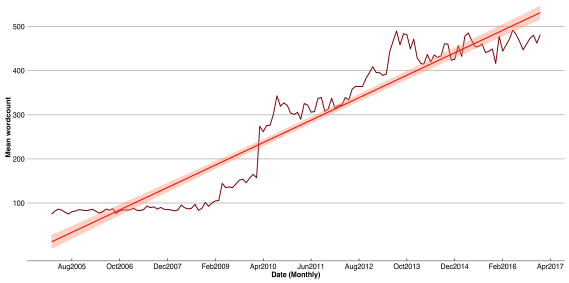

Figure D.4: Mean wordcount, monthly

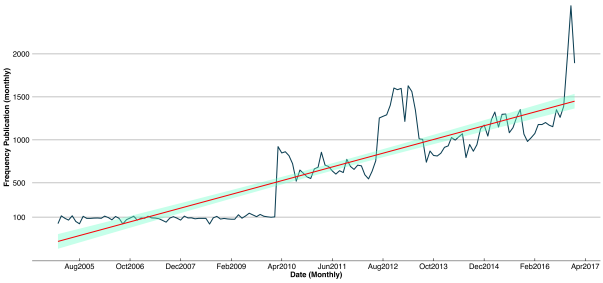

Figure D.5: Frequency of Publication, monthly

Note: The whole period 2005-2017 is not representative of the sample used in this paper, as we only have data for the period 2005-2010 for a couple of countries from North Africa. Zooming over the period 2010-2017 (see thereafter) is more representative of the overall trend.

Average over all francophone African countries, zoom over the period 2010-2017

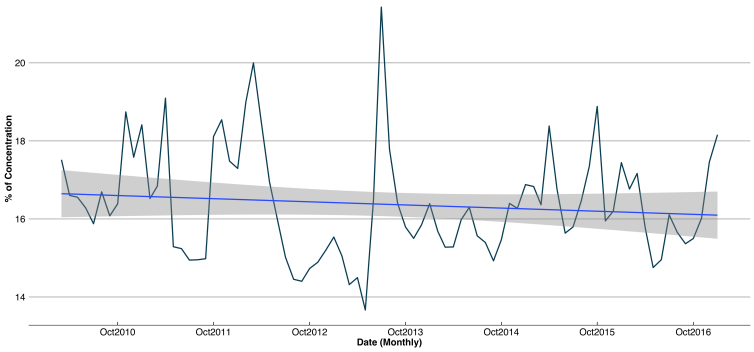

Figure D.6: Herfindahl HIrschman index, monthly

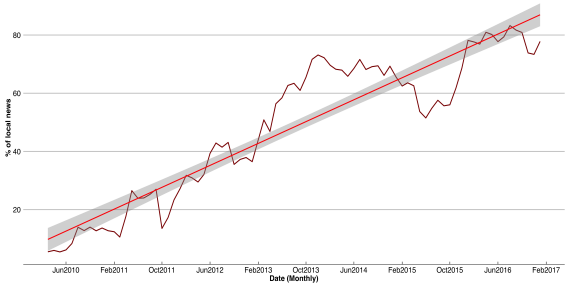

Figure D.7: Local news ratio, monthly

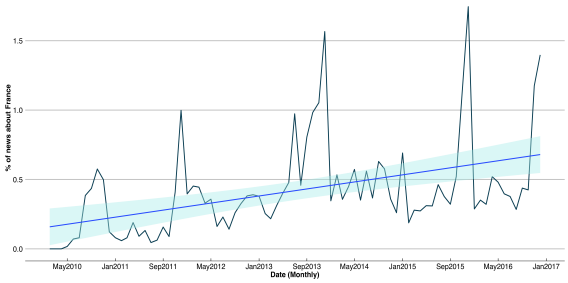

Figure D.8: French news ratio, monthly

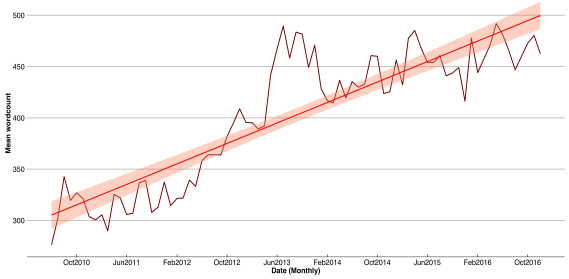

Figure D.9: Mean wordcount, monthly

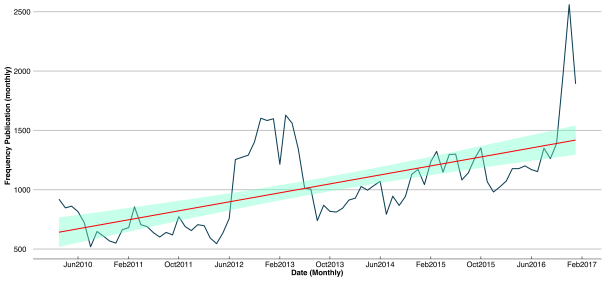

Figure D.10: Frequency of Publication, monthly

Appendix E Regression Tables without controls

Table E.1: **Multivariate Multiple OLS Regression Estimates, monthly**

	Normalized HHI	French News ratio	Local News ratio	Wordcount	Publication Freq.
Market penetration (log total)	1.32*** (0.14)	-0.01 (0.02)	6.17*** (0.24)	0.51*** (0.00)	0.27*** (0.01)
Country Fixed effects	Yes	Yes	Yes	Yes	Yes
Time Fixed effects	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.872	0.512	0.967	0.998	0.984
Observations	645				

Robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table E.2: **Fixed Effect Estimates, monthly**

	(1) Mean Wordcount (log)		(2) Frequency of publication (log)	
Market peneration (log total)	-0.04**	(0.02)	0.28***	(0.09)
Country Fixed effects	Yes		Yes	
Time Fixed effects	Yes		Yes	
Observations	645		645	
adj. R^2	0.029		0.144	

Robust tandard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table E.3: **Fixed Effects Tobit Estimates, monthly**

	(1) Normalized HHI		(2) Local News Ratio		(3) French News Ratio	
Market penetration (log total)	0.97	(0.66)	0.33	(1.35)	0.28***	(0.09)
Country Fixed effects	Yes		Yes		Yes	
Time Fixed effects	Yes		Yes		Yes	
sigma_u	3.54***	(0.77)	18.51***	(2.90)	0.48***	(0.10)
sigma_e	6.96***	(0.20)	12.12***	(0.34)	0.91***	(0.03)
Observations	645		645		645	

R obusttandard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table E.4: **Arellano-Bond Difference GMM Estimators, monthly**

	(1) Norm HHI		(2) French ratio		(3) Local ratio		(4) Mean Wordcount		(5) Freq. Publication	
L.Norm HHI	0.32***	(0.05)								
L2.Norm HHI	0.08**	(0.04)								
L.French ratio			-0.06	(0.13)						
L2.French ratio			-0.02	(0.10)						
L.Local ratio					0.56***	(0.06)				
L2.Local ratio					0.02	(0.06)				
L.Mean Wordcount							0.42***	(0.07)		
L2.Mean Wordcount							-0.02	(0.05)		
L.Freq. Publication									0.68***	(0.07)
L2.Freq. Publication									0.01	(0.05)
Market penetration	3.34***	(0.98)	0.48***	(0.17)	0.56	(1.60)	-0.00	(0.03)	0.18	(0.13)
L.Market penetration	-0.79	(1.24)	-0.21	(0.21)	4.86	(3.33)	-0.02	(0.05)	-0.13	(0.12)
L2.Market penetration	0.16	(0.89)	0.10	(0.18)	-4.47	(3.91)	-0.04	(0.02)	0.13	(0.10)
Country Fixed effects	Yes		Yes		Yes		Yes		Yes	
Time Fixed effects	Yes		Yes		Yes		Yes		Yes	
N	558		558		558		558		558	

Robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
All variables are for monthly data. Mean Wordcount, Freq. Publication and Market penetration are in log.

Table E.5: **Blundell-Bond System GMM with Forward Orthogonal Deviations, monthly**

	(1) Norm HHI		(2) French ratio		(3) Local ratio		(4) Mean Wordcount		(5) Freq. Publication	
L.Norm HHI	0.35***	(0.05)								
L2.Norm HHI	0.10**	(0.05)								
L.French ratio			0.02	(0.12)						
L2.French ratio			-0.04	(0.06)						
L.Local ratio					0.62***	(0.05)				
L2.Local ratio					0.05	(0.05)				
L.Mean Wordcount							0.48***	(0.05)		
L2.Mean Wordcount							0.01	(0.05)		
L.Freq. Publication									0.76***	(0.07)
L2.Freq. Publication									0.06	(0.06)
Market penetration	3.29***	(0.92)	0.46***	(0.17)	2.36	(2.31)	0.03	(0.04)	0.19	(0.13)
L.Market penetration	-1.09	(1.37)	-0.22	(0.21)	3.53	(2.53)	-0.04	(0.06)	-0.11	(0.14)
L2.Market penetration	-0.83	(0.72)	-0.19	(0.21)	-5.13	(4.44)	-0.06**	(0.03)	0.07	(0.08)
Country Fixed effects	Yes		Yes		Yes		Yes		Yes	
Time Fixed effects	Yes		Yes		Yes		Yes		Yes	
N	628		614		614		628		628	

Robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
All variables are for monthly data. Mean Wordcount, Freq. Publication and Market penetration are in log.