Introducing News Media Sentiment Analytics to Residents' Attitudes Research

Abstract

The progress in sentiment analytics and communication research provides a powerful scaffold by which to reexamine the long-debated research on residents' attitudes towards tourism. To mitigate the limitations of the classical survey-based research method, this study takes a news media sentiment analytics perspective to unveil how the residents' attitudes towards tourism evolve over time and how socioeconomic factors interact with such evolving attitudes in the context of Hong Kong. Drawn on a news dataset containing 72,755 news articles published in Chinese language newspapers, this study computes the overall news sentiments for 156 calendar months since 2003, examines the face validity and nomological validity of the results, and discusses the long-run dynamics between residents' attitudes and typical socioeconomic factors. This study adds a vital dimension to current residents' attitudes research and practices from data-scarce to data-rich studies and from static snapshots to dynamic unfolding.

Keywords

Residents' Attitudes; News Media; Sentiment Analysis; Hong Kong; Mainland Chinese Tourists

INTRODUCTION

Understanding residents' attitudes and gaining local residents' support are essential to the successful and sustainable development of tourism destinations and have long been the primary concerns for both academia and policymakers (Sharpley 2014; Snaith and Haley 1999). The past decades have seen an increase in a rich stream of research investigating residents' attitudes towards the social, economic, and environmental impacts of tourism development. Several recent reviews (e.g., García, Vázquez, and Macías 2015; Nunkoo, Smith, and Ramkissoon 2013; Nunkoo and So 2016; Sharpley 2014; Uysal et al. 2016) have reflected on the progress made in developing the concept and nomological network of residents' attitudes towards tourism. However, these reviews critically pointed out that the classical cross-sectional survey-based methods have limited this stream of study to further contribute to our knowledge and practice. Sampling issues and self-report bias aside, the survey-based method is generally not effective in capturing the dynamic changes of residents' attitudes. As a consequence, how the overall residents' attitudes towards tourism evolve over time and how socioeconomic factors interact with such evolving attitudes remain unclear.

The progress in big data analytics and communication research provides a powerful scaffold by which to tackle these issues. The big data era has ushered in a new era of empiricism, which may help produce studies with much greater breadth and depth as well as inherently longitudinal in contrast to past research (Lazer et al. 2009). Meanwhile, the agenda-setting theory claims a strong correlation between the emphases that mass media (e.g., news media) place on certain issues and the attitudes attributed to these issues by the mass audience (McCombs and Shaw 1972). Therefore, in the current study, we attempt to breathe new air to this subject using the news media sentiment analytics to examine the news sentiment as an important indicator of residents' attitudes towards a key source market in a

world-leading destination, that is, how Hong Kong (HK) residents view Mainland Chinese tourists (MCTs).

HK is one of the world's leading tourist destinations, but more than three-quarters of its visitors come from a single source market: Mainland China. From the early active expansions of individual visit scheme (IVS) to attract MCTs in 2003 to the recent residents' protests against MCTs, the rather dramatic change over the past decade provides a vivid example that necessitates a longitudinal study examining dynamics of residents' attitudes towards tourism.

The HK residents are also kept well informed by a vigorous media environment that enjoys freedom of expression. Therefore, we conjecture that the features of media coverage in HK, such as the affective polarity of news articles, will provide abundant data and unique insights for consolidating our understandings of HK residents' attitudes towards tourism. Accordingly, this study has the following specific research objectives:

- Elaborate and justify the relations between news media coverage and residents' attitudes towards tourism based on the agenda-setting theory,
- Develop a news media sentiment analysis framework to automatically analyze textual data of news articles published in Chinese-language newspapers in HK,
- Examine the HK news sentiment towards MCTs by applying the developed framework, and
- Validate the results of news sentiment and reveal the dynamics of residents' attitudes towards tourism with major socioeconomic factors.

These efforts attempt to add a new angle to the academic understanding of residents' attitudes towards tourism and create a useful tool that can be used for long-run strategic tourism planning.

LITERATURE REVIEW

Residents' Attitudes towards Tourism

Studies about residents' attitudes towards tourism have proliferated over recent decades, indicating the increasing significance of contemporary tourism for economic and social development of a destination, and the increasing importance of residents' involvement in tourism for the destination's sustainable development. The ongoing efforts to understand what residents think about and how they respond to tourism and its impacts aim to create harmonious host–guest interactions during the development of sustainable destinations.

The rationale for such an important stream of research exists in that tourism affects the social, economic and environmental aspects of a destination. On the one hand, the local community may benefit from tourism via enhanced economic activities, upgraded recreational facilities, organization of festivals, revitalization of local cultures and investments in environmental infrastructure. On the other, tourism may increase the cost of living; cause noise pollution, crowding, traffic, begging and crimes; disrupt traditional cultures and ways of life; and even lead to environmental degradation (Kim, Uysal, and Sirgy 2013). Accordingly, most studies have classified residents' attitudes into positive and negative categories. Meanwhile, much of the current research has identified, measured, and compared a diversity of variables to explain and potentially predict residents' attitudes toward tourism (e.g., Sharpley 2014). These variables can be generally categorized into destination-related factors, such as socioeconomic variables of a destination, and individual factors, such as resident demographics (Deery, Jago, and Fredline 2012; Faulkner and Tideswell 1997).

Several review papers have clearly delineated the trajectory of residents' attitudes in tourism research (e.g., García et al. 2015; Deery et al. 2012; Kim et al. 2013; Nunkoo and So 2016; Sharpley 2014; Uysal et al. 2016). It is generally agreed that significant progress has been made in terms of defining and identifying the antecedents and consequences of

residents' attitudes towards tourism. However, in terms of research methodology, current reviews critically pointed out two major limitations that may restrict the advancement of this stream of study (Sharpley 2014).

First, both residents' attitudes and tourism destinations are dynamic; they continuously adapt and transform in response to changes in terms of cost and benefit received and the challenges emerging in a competitive tourism environment. Such dynamism has been emphasized since the beginning of residents' attitudes research, which can be traced to Doxey's (1975) Irridex model. Similar conclusions can be drawn from Butler's (1980) resort life-cycle model. However, most empirical studies on residents' attitudes are cross-sectional, of which the results are relevant within a specific timeframe but become less meaningful from a historical perspective.

A few recent studies have taken a longitudinal perspective to examine residents' attitudes towards tourism, but their data were collected from a limited number of discrete time points. For example, Li et al. (2015) tracked local residents' evolving attitudes towards the 2010 Shanghai Expo using a three-wave survey approach. Bimonte and Faralla (2016) distributed a two-wave survey in a seaside resort before and during the tourist season to identify differences in residents' attitudes towards tourism impact and overall life satisfaction. Ivleves (2017) used a six-wave European Social Survey to study how residents' well-being is affected by actual tourist arrivals by. Vargas-Sánchez et al. (2014) revealed disparate seasonal residents' attitudes toward tourism in two rounds of surveys (high and low seasons) in Spain.

Second, typical studies on residents' attitudes often adopt the questionnaire-based survey method that combines a variety of question formats. These questionnaires are usually distributed by mail or the Internet or conducted through face-to-face interviews. The survey approach is effective in measuring the individual-level attitude towards tourism in specific tourist-receiving communities; however, it often suffers from sampling and non-response

biases. Additionally, it would be inappropriate to derive the collective attitudes of a host community by simply lumping individual residents' responses together (Hiller and Wanner 2011). In addition, the survey method may be useful in identifying the antecedents and consequences of residents' attitudes; however, as it is expensive and time-consuming, this method cannot provide a useful tool aiding practical tourism planning. As suggested by Uysal et al. (2016), given that the majority of the studies employed primary data, replicating or augmenting the primary-data studies with secondary data may be an ideal option.

News coverage is one of the most important secondary data sources employed by tourism and hospitality researchers (e.g., Kapuściński and Richards 2016; Liu and Pennington-Gray 2015). Unfortunately, this stream of study has been largely neglected by existing review articles. Thus, related news media research will be reviewed in the next section.

News Coverage and Residents' Attitudes

Mass media is generally believed to play a significant cognitive role allowing people to acquire meanings and values (Scheufele and Tewksbury, 2007). Particularly, the agenda-setting theory states that there is a strong correlation between the emphases that mass media place on certain issues and the attitudes attributed to these issues by the audience (McCombs and Shaw 1972), and this allow us to look at the connection between media coverage and residents' attitudes towards tourism within a host community. Among all mass media, news media have been widely examined by social science researchers (e.g., Castelltort and Mäder 2010; Kapuściński and Richards 2016; Scheufele and Tewksbury 2007; Stepchenkova and Eales 2011). Subsequent empirical studies using a variety of non-news data sources have largely validated the core findings of many studies that initially relied on news data activity (Neresini and Lorenzet 2016).

Prior research into the relationship between news media coverage and audience's attitudes is bidirectional. On the one hand, researchers found a significant effect of the news coverage on audience's attitudes (e.g., Castelltort and Mäder 2010; McCombs and Shaw 1972; Scheufele and Tewksbury 2007). This effect occurs even with rational readers who are conscious of media bias (Anderson and McLaren 2012). In the tourism and hospitality context, researchers also found a significant effect of news coverage on host/guest decisions, with implications for destination management. For instance, news media may influence tourist destination image formation and then destination choice (Stepchenkova and Eales 2011). Santos (2004) analyzed 45 travel news articles from four U.S. newspapers and identified two salient frames representing Portugal as a tourism destination. The study of Castelltort and Mäder (2010) attributed the influence of news media on destination image formation to their high credibility and ability to reach a broad mass of potential tourists. News media also play important roles in event tourism, especially for international mega-events, such as the Olympics, the World Cup, and the Expos (Getz 2008). Robertson and Rogers (2009) stated that news media may influence and change the public perception of the sociocultural effects of festivals. Kim and colleagues (2014) examined 295 news reports of the 2012 Expo in Korea and concluded that newspaper reports have a positive effect on attendance. Additionally, news coverage of hazards is often commented to be of critical importance to individuals' perceived risks associated with destination choice. Guided by framing theory and the existing knowledge of perceived risks in tourism, Kapuściński and Richards (2016) employed an experimental design to explore the influence of different media frames concerning hazards on tourists' judgment of risks. Several other studies, such as those of Brown (2015), Liu and Pennington-Gray (2015), and Stepchenkova and Eales (2011), have also reported that negative news coverage could deter potential travelers from visiting a destination.

On the other hand, news coverage can also be influenced by, and accordingly, represent attitudes. Surveys of media professionals and ethnographic studies of newsrooms have suggested that journalists must consider the audience in their assessments of ‘‘newsworthiness’’ and that they *purposively* choose which events to cover and highlight with audience preferences in mind (Lester 1980; Molotch and Lester 1974). Recent communication research has also argued that different types of events may affect the direction of such influence. Public issues like tourism development are more likely to follow the audience-driven framework in which news media is influenced by community concerns (Uscinski 2009).

Tourism and hospitality researchers have also studied attitudes using news content analysis that may be explicitly or implicitly based on the above relationship. For example, Wu et al. (2012) employed the news frame analysis to identify salient public debate topics about the “Golden Week” holiday policy in China. Hwang et al. (2012) analyzed regional newspapers to determine the sequence and effects of local community collective actions to tourism development. Peel and Steen (2007) analyzed news coverage on international backpackers in Australia to understand their interactions with local residents. Liu and Pennington-Gray (2015) conducted a framing analysis of news coverage to understand public concerns of bed bug crisis in the U.S. travel industry.

Sentiment Analytics in Tourism

Sentiment analytics generally refers to the use of analytical tools and frameworks to computationally detect people’s opinions, attitudes, and emotions toward objects (Kirilenko et al. 2018; Xiang et al. 2017; Fu et al. 2019). A sentiment is mathematically defined as a positive or negative attitude about an object or an aspect of the object from opinion holders (Liu 2012). This has been regarded as an effective and efficient way to gain the affective

attribute and evaluative tone of textual data (Cambria 2016; Kirilenko et al. 2018). More and more tourism researchers and practitioners have recognized the importance of sentiment analytics to the increasingly digitalized tourism industry. Its applications include processing exponential volume of tourism data, real-time capturing of tourists' and residents' attitudes, and monitoring tourism service quality and recovery (Alaei, Becken, and Stantic 2019).

Technically speaking, sentiment analytics represents a polarity (i.e., positive and negative) classification problem (Fu et al. 2019). Existing approaches to such classification fall into two categories: semantic-oriented and machine-learning approaches. The former mainly relies on predefined lexicon with semantic orientations of words to determine a document's overall polarity. The latter trains machine-learning algorithms, such as Support Vector Machines (SVM) and Naïve Bayes (NB), by a corpus of affectively annotated texts to predict the polarity of a new document.

Recent reviews (Alaei, Becken, and Stantic 2019; Kirilenko et al. 2018) have shown that sentiment analytics studies in tourism have generally focused on user-generated contents (UGC), including social media posts and online reviews. For example, Philander and Zhong (2016) employed a lexicon-based method with a general sentiment dictionary to analyze twitters' sentiment towards hotels. Duan et al. (2016) applied a machine learning approach to calculate online hotel review sentiment for service quality.

However, news media, a major agent through which residents acquire local tourism information, have been largely neglected by researchers. As aforementioned, accumulated evidence has shown a strong correlation between news coverage and residents' attitudes—one that reinforces each other reciprocally (Uscinski 2009). Therefore, we believe in the importance of taking the news media sentiment analytics perspective to gain insights into resident attitudes' towards tourism.

The Research Context

This study focuses on examining HK residents' attitudes, through news sentiment, towards tourism development associated with MCTs. This particular context is chosen for the following reasons. First, MCTs are the mainstay of HK's burgeoning tourism industry and are an important market for the retail trade. From another point of view, for a city accommodating 7.4 million residents in its 1,100 km^2 land, huge tourist arrivals have brought considerable social challenges that go beyond a tourism issue and cannot be easily offset by the economic benefits. Local support is critical to the sustainable development of tourism, and it is impossible to nurture Hong Kongers' support without understanding their sentiment.

Second, the scale and growth rate of the entry of MCTs to HK, in retrospect, make it an important case to study. Since the IVS launched in 2003, MCT arrivals had an over five-fold increase from 8.5 million in 2003 to 44.4 million in 2017. The IVS allows Mainland Chinese residents of major cities to visit HK on an individual basis without having to join a tour group as previously required. The adoption of Multiple Entry Individual Visit Endorsements (M-Permit) for Shenzhen residents in 2009 exacerbated the phenomenon, because M-Permit holders can now visit HK multiple times with a single endorsement. The size of the city and its population as opposed to the number of MCT arrivals, along with the economic linkage and resource dependence between HK and Mainland China, could intensify the tension between hosts and guests (Chen, Hsu, and Li 2018). Accordingly, we may observe sharp changes in residents' attitudes along with the influx of tourists.

Finally, the availability of world-class telecommunications and a keen interest in HK's affairs have attracted many news agencies to establish regional headquarters or representative offices in HK. HK enjoys a highly developed media environment and level of freedom of expression (HKSAR Government 2016). Despite the growing prevalence of social and online media, news media remain popular in such a competitive media market. According to the

report of Independent Press Standards Organization media survey, HK's yearly newspaper readership remains at 5.4 million in recent years (HKSAR Government 2016; Siu 2017). Thus, host–tourist interactions have been well covered by the news media. In sum, the particular social, economic, and media environment creates an interesting backdrop and a suitable case for this study on residents' attitudes towards tourism.

A NEWS MEDIA SENTIMENT ANALYSIS FRAMEWORK

Our objective is to computationally analyze the textual content of news articles and detect HK residents' attitudes towards MCTs. To this end, we developed an automated news media sentiment analysis framework to analyze local news articles published in major Chinese newspapers in HK. According to the prior literature (Duan et al. 2016; Fu et al. 2019), we operationalized the news media sentiment analysis as the task of analyzing the sentiment of textual data from news articles and classifying these articles into positive and negative sentiment categories.

Although sentiment analysis has been regarded as an important research venue in tourism applications, it still faces several challenges when used in tourism and hospitality research (Kirilenko et al. 2018; Leung et al. 2013). First, sentiment analysis is highly dependent on research domains, and its classification accuracy is still limited for tourism and hospitality research. Second, most of the sentiment classification studies in tourism and hospitality concentrate on UGCs, of which opinions are typically expressed directly and freely. However, journalists intend to maintain their "objectivity" and to express opinions by reporting both favorable and unfavorable facts. Such implicit tendencies are more difficult to detect by using existing sentiment analysis models. In addition, the wording and phrasing of news articles are quite different from UGCs in many aspects, such as word selection, paragraph format, and article length. Finally, the use of both the written form of local Cantonese dialect and the official national language (i.e., Mandarin) in major Chinese newspapers in HK exacerbates the difficulties encountered in processing news articles in sentiment analysis.

Therefore, we adapted a four-stage machine-learning based news sentiment analysis framework to deal with these challenges. As shown in Figure 1, the four major stages are

news collection and cleaning, model building and evaluation, news sentiment estimation, and data analysis and visualization.

==========

Please insert Figure 1 about here.

==========

(1) News collection and cleaning. The major task of this stage is to prepare datasets for the news sentiment analysis. Online news about the topics of interest are collected, cleaned, and stored into local storage. Further, two datasets are generated: a benchmark dataset and a full news dataset. The benchmark dataset, which is a random sample of full news dataset, is used to build and evaluate sentiment classification models in the second stage. Meanwhile, the full news dataset is used to estimate the sentiment polarity for each news article in the third stage.

(2) Model building and evaluation. This stage involves an iterative process to fine-tune the set of parameters and the sentiment classification model for news articles based on the benchmark dataset. The benchmark dataset is annotated with positive and negative sentiment by human experts. After that, the text features are extracted from this dataset, weighted into term vectors, and inputted into the analysis models. The machine-learning based sentiment analysis generally requires this benchmark dataset to be further partitioned into a training set and a test set. A training process is conducted on the training set to tune the parameters of the sentiment classification model and to formulate the final model for news sentiment estimation. The testing process then employs the trained classifiers and validates its predictive performance on the testing set. Both the training and testing are conducted iteratively until an acceptable performance is achieved.

(3) News sentiment estimation. Once the classification model is evaluated and fine-tuned, this stage applies the model to the full news dataset to estimate the sentiment polarity for each news article. The generated polarity is used in the next stage.

(4) Data analysis and visualization. This process computes news sentiment scores for different time intervals, conducts descriptive and inferential data analysis, and visualizes the evolving pattern of news sentiment towards MCTs. In this study, we proposed an indicator of *Sentiment* to measure overall news sentiment based on the sentiment polarity of each news article. *Sentiment* is obtained using Formula (1) expressed as

$$Sentiment_t = \frac{P_t}{P_t + N_t}\ , \tag{1}$$

where $Sentiment_t$ denotes the overall news sentiment regarding a particular period t, and $P_t + N_t$ represents the total number of news articles published within the specific period, among which P_t denotes the number of positive news and N_t denotes the number of negative news. The news sentiment ranges between 0 and 1, where 0 represents the absolutely negative residents' attitude towards tourism and 1 represents the absolutely positive attitude.

In this study, we focused on binary sentiment polarities due to the following theoretical and technical considerations. Theoretically, although news articles are supposed to be "objective," communication theories argue that journalists resort to expression of their positive or negative opinions by embedding statements in more complex discourses, purposefully omitting some facts and highlighting others, or quoting the preferred persons to say what journalists want to say (Entman 1993; Van De Kauter, Breesch, and Hoste 2015; Scheufele and Tewksbury 2007). In addition, prior tourism impact studies aimed to capture both the positive and negative effects of tourism; thus, they tend to classify residents' attitudes into positive or negative (e.g., Kim, Uysal, and Sirgy 2013; Sharpley 2014; Uysal et al. 2016). Technically, the attitude of neutrality is vague, difficult to identify, and varies among readers, thereby creating many technical challenges to efficiently and accurately

identify neutral opinions (Duan et al. 2016; Cambria 2016). Therefore, many tourism sentiment analytics studies adopt binary sentiment polarities (e.g., Kang, Yoo, and Han 2012; Wang et al. 2013; Schuckert, Liu, and Law 2015; Ye, Zhang, and Law 2009).

EMPIRICAL STUDY

Application and Specifications

An empirical study was conducted by applying the abovementioned news media sentiment analysis framework. The objective was to analyze the polarity of each news article and to compute the overall news sentiment regarding HK residents' attitudes towards tourism from 2003 to 2015 since the IVS was launched which led to an unprecedented increase of MCTs.

In Stage 1 of the news media sentiment analysis framework, we selected the news source from the WiseNews database. This covers 19 major Chinese-language newspapers published in HK, including *Apple Daily, Hong Kong Commercial Daily, Ming Pao Daily, Oriental Daily News, Sing Tao Daily, Ta Kung Pao, Wen Wei Po*, and others. The WiseNews database (http://wisenews.wisers.net) is the most comprehensive HK news database providing accurate news content from verifiable and recognized sources. According to a Hong Kong Special Administrative Region Government report (HKSAR Government 2016), a total of 19 Chinese-language dailies cover local news in 2016. Thus, our work included *all*, instead of *a sample of*, related news coverage in the proposed sentiment analysis.

We carefully selected Chinese keywords (e.g., "Mainland Chinese," "tourist," "tourism," and "travel") to search for news articles related to residents' attitudes towards MCTs. WiseNews database is a commercial product that provides comprehensive and precise news information search supported by intelligent algorithms and expert-tailored linguistic taxonomy. We also estimated the false positives of our specific search by randomly selecting 1,500 retrieved news articles and re-checked by human raters. The rate of false positives was 8.6%. This indicates a high precision of our search considering that the average false positive rate of general searching engines, such as Google and Yahoo, is about 24% (Kumar and Prakash 2009).

The returned news articles in the WiseNews database were in the format of HTML. We

then programmed using Python to clean these tags and saved the textual body and metadata (i.e., title, date, newspaper) of a piece of news into local storage. Following the suggestions of Moznette and Rarick (1968), the full news dataset should eliminate articles shorter than 100 words and articles purely reporting statistical data (e.g., stock quotes). As such, we had a dataset containing 72,755 news articles. The average length of these articles is 862.07 words with a minimum of 100 and a maximum of 16,125.

Stages 2 and 3 of the proposed news sentiment analysis framework were mainly implemented by RapidMiner Studio. RapidMiner Studio is one of the top open source data science platforms and provides advanced analytics capabilities by combining analysis operators and setting their parameters. Operators are the building blocks of RapidMiner Studio to conduct professional data analysis, including basic data processing and classification. Due to the limited Chinese capabilities of RapidMiner Studio, in this study, we developed a Python program based on a widely-used Chinese segmentation tool named Jieba (https://github.com/fxsjy/jieba) along with a manually refined lexicon containing both simplified and traditional Chinese to process text features of news articles. The process included the removal of stopping words and segmentation of Chinese characters.

The predictive performance of the sentiment classification model was evaluated against a balanced human-annotated random sample of the full news dataset (i.e., the benchmark dataset). We adopted a prior benchmark annotation procedure (Fu et al. 2019) to train human raters and annotate the sentiment polarity of news articles. Communication theories argue that journalists express implicit sentiment towards a specific target by selecting, omitting, and emphasizing some specific aspects of the target, as well as elaborately organizing them (e.g., Scheufele and Tewksbury 2007). In the context of identifying HK residents' attitudes towards MCTs, a piece of news may present many facts pertinent to the increase of inbound tourists to HK, whereas the focus(es) could be about only one or a couple of subjects (e.g.,

environmental problems, traffic jams, or inconveniences for the local residents). Therefore, the sentiment polarity of this piece of news was annotated as negative although it included objective narratives. Following Fu et al. (2019) and Stepchenkova and Eales (2011), we used control coding to minimize annotator bias. Two native HK residents who graduated from a tourism management program were employed to independently annotate these articles into positive and negative polarities. Cohen's kappa (Cohen 1960) was measured as 0.870, constituting a high level of agreement between the two annotators (Landis and Koch 1977). After that, the differences of their annotation were resolved through follow-up discussions.

Typical classifiers, such as SVM and NB, assume a balanced dataset whose categories (e.g., positive and negative categories) are evenly distributed (Lee and Lee 2012). In particular, for machine learning-based sentiment classifications, a balanced training dataset is preferable for building better classifiers and achieving higher accuracy (Zhang et al. 2016; Moraes, Valiati, and Gavião Neto 2013). Therefore, we used stratified sampling and randomly selected 1,000 articles (i.e., 500 positives and 500 negatives) from the annotated samples to build the benchmark dataset.

Next, we evaluated the predictive performance of popular classifiers (i.e., SVM and NB) suggested by prior research (Liu et al. 2013; Ye, Zhang, and Law 2009) against a set of widely-used metrics including accuracy and F-measure (Liu et al. 2013; Ye, Zhang, and Law 2009) with a 10-fold cross-validation procedure (Liu 2012). The results show that the SVM classifier (Accuracy = 0.913; F-measure = 0.914) outperformed the NB classifier (Accuracy = 0.839, F-measure = 0.840) in our specific classification task. Therefore, in this study, we selected SVM as the sentiment classification model.

According to Kirilenko et al. (2019), we also calculated Cohen's kappa between the machine-predicted and human-rated results as 0.818, suggesting a high level of agreement between machines and humans. Kappa is regarded as a robust and conservative measure than

the percentage agreement (i.e., the accuracy metric) because the former considers the agreement occurring by chance (Stepchenkova and Eales 2011). This allowed us to implement Stage 3 to estimate news sentiment for each article in the full news dataset and to compute the overall news sentiment regarding a particular period using Formula (1). Finally, in Stage 4, we exported the estimated sentiment polarity of each news article into analytical software for further analysis and visualization.

Overall Results

In this section, we reported the overall results of the evolving pattern of news sentiment towards MCTs for the studied 156 calendar months. Table 1 lists the positive and negative news frequency distribution by year, and Figure 2 illustrates the evolving pattern that reflects the fluctuation of HK residents' attitudes towards tourism over time. The average monthly news sentiment is 0.703, indicating a general positive attitude towards MCTs during the 13-year period. The overall news sentiment peaked at 0.978 in June 2003 and bottomed at 0.304 in September 2012.

==========

Please insert Table 1 about here.

==========

==========

Please insert Figure 2 about here.

==========

Robustness Check

The competitive media industry of HK is widely recognized to enjoy freedom of expression, with a few newspapers demonstrating strong, even partisan views on China-related issues, including MCTs. Therefore, we conducted a robustness check to examine the influences of the political stance of newspapers in HK on the overall evolving pattern of news sentiment

towards MCTs.

Newspapers in HK can generally be classified into three categories (e.g., Chan and Lee 2007; Jiang 2013; Wikipedia 2018): leftist, centrist, and rightist. Most newspapers are generally political neutral (i.e., centrist). Among those with a particular political stance, *Ta Kung Pao, Wen Wei Po,* and *Hong Kong Commercial Daily* are generally leftists and *Apple Daily*, *Sharp Daily,* and *Ming Pao* are generally rightists. We then aggregated the news sentiment scores across the three categories. Figure 3 illustrates that the general evolving patterns of news sentiment across the three newspaper categories are quite similar. The conclusion is further supported by the significantly strong Pearson correlations across the three series of news sentiment (i.e., $r_{\text{All-Leftist}} = 0.897$, $r_{\text{All-Rightist}} = 0.895$, and $r_{\text{All-Centrist}} = 0.986$). Although we observed mean differences of news sentiment across the three categories, the political stance of newspapers had little influence on the overall evolving pattern. Nevertheless, we believe this deserves further in-depth investigations in future research.

==========

Please insert Figure 3 about here.

==========

In addition, news articles are widely dispersed among different newspapers. The top newspapers in terms of article volume are *Sing Tao Daily* (n = 7,387), *Wen Wei Po* (n = 7,298), *Hong Kong Economic Times* (n = 6,905), *Ming Pao* (n = 6,137), *Ta Kung Pao* (n = 6,003), and *Apply Daily* (n = 5,471). Therefore, our sentiment results are unlikely to be swayed due to the dominant number of news from one or few newspaper(s) with biased views. Hence, in the following sections, we mainly employed the overall news sentiment in Figure 2 to further examine its face validity and nomological validity with socioeconomic factors.

VALIDATION AND DISCUSSION

Face Validity with Socioeconomic Events

To develop the face validity of our news sentiment and put things into context, we reviewed major socioeconomic events, which are presumably related to MCTs and might influence HK residents' attitudes. Table 2 compiles selected key events in chronological order. It seems that these key events generally match the inflection points of the overall news sentiment illustrated in Figure 2, thus demonstrating the validity of our analysis results.

==========

Please insert Table 2 about here.

==========

For example, in 2003, to recover from the economic downturn caused by the Severe Acute Respiratory Syndrome (SARS) epidemic, the HK SAR Government launched the IVS. Before that, most Mainland Chinese could only travel to HK under business visas or by joining group tours. The IVS was first implemented only in the four cities of Guangdong Province, but even then, the MCT arrivals by IVS still contributed greatly to HK's economic recovery. The high sentiment score in 2003 (M = 0.854, SD = 0.077) reflects the very positive attitudes of HK residents towards MCTs.

From July 2003 to January 2007, the IVS gradually expanded from the initial 4 cities to 49 cities in the Mainland, which remained so until now. The annual MCT arrivals grew explosively from 8.46 million in 2003 to 13.60 million in 2006. As such expansions have created more frequent interactions between MCTs and local residents, they led to the emergence of associated problems, such as crowding in transportation junctions, attractions and shopping malls; the lack of urban public resources; public security concerns; and cultural conflicts. An obvious decline of sentiment score can be observed in Figure 2: from 2003 (M = 0.854, SD = 0.077) to 2006 (M = 0.648, SD = 0.038).

From 2007 to 2009, under the dual influence of the CEPA amendment and the global financial crisis, a continuous increase of the sentiment score emerged, reflecting a gradual positive change in HK residents' attitudes towards MCTs. In 2009, the HK SAR Government issued the M-Permit for Shenzhen residents to stimulate their visits. Subsequently, the surge of MCTs helped create employment opportunities and combat the influence of the 2008 global financial crisis (Legislative Council Secretariat 2014).

While benefiting the HK economy, the IVS, especially the M-permit, led to an unprecedented increase of MCTs and caused social tensions in HK. The number of MCTs soared from 17.96 million in 2009 to 40.75 million in 2013. Different from the early stage of IVS, the high demand of MCTs and parallel trading activities led to the increased homogenization of shopping malls, with shops selling general household goods for locals being displaced by luxury stores. The shortage of some daily products, especially baby formula, was widely reported in news and created many social protests in HK. Figure 2 shows that the sentiment score peaked in early 2010 ($M = 0.793$, $SD = 0.074$) and then declined until 2013 ($M = 0.503$, $SD = 0.103$) when the HK SAR Government issued a restriction on the quantity of baby formula for persons departing from HK.

From 2013 to 2015, while the conflicts between local residents and MCTs continued to grow, media reports of hostility towards MCTs and the political tension in HK led potential MCTs to consider alternative travel destinations. Consequently, the drop in MCT flows during the May–June holidays in 2014 caused significant retail sales slump. The HK community began to debate on the advantages and disadvantages of IVS, especially the M-permit, on major news media. As a result of the public debate, the Beijing Central Government decided to replace the M-permit with a one-trip-per-week limit upon the request of the HK SAR Government. The new measure was expected to balance HK's tourism capacity and its visitor arrivals and to curb parallel trading activities by those who enter HK

multiple times a week or a day. During this period, another gradual increase of the sentiment score can be observed from 2013 (M = 0.503, SD = 0.103) to 2015 (M = 0.657, SD = 0.073).

Nomological Validity with Major Socioeconomic Factors

As indicated in prior review studies (e.g., Sharpley 2014), although many destination communities enjoy the economic benefits brought about by tourism development, they also suffer from the negative social consequences of the development. Accordingly, most studies adopted socioeconomic factors to explain residents' attitude towards tourism (e.g., Deery, Jago, & Fredline 2005; Deery et al. 2012; Harrill 2004; Sharpley 2014). Among these factors, tourist arrival has been regarded as a critical determinant of social impacts (e.g., Butler 1980; Deery, Jago, and Fredline 2012; Doxey 1975; Ivlevs 2017; Sharpley 2014), whereas employment opportunity has been identified as a salient economic benefit of tourism (e.g., García, Vázquez, and Macías 2015; Sharpley 2014). Therefore, we hypothesized that *the overall news sentiment is significantly correlated with these socioeconomic factors*.

To test the above hypothesis, we selected MCT density and unemployment rate as variables of interest after considering the context of HK and the availability of HK statistical data. MCT density (*Density*) is an indicator of MCT arrivals relative to the HK population within the same time period. Compared with the absolute number of tourist arrivals, tourist density is a more robust measure of tourism development level (Deery, Jago, and Fredline 2005; Ivlevs 2017). Unemployment rate (*Unemp*) refers to the proportion of unemployed persons in the labor force of HK.

These variables were sourced from the HK Census and Statistical Bureau. Descriptive statistics and correlations of these monthly variables are reported in Table 3. Significant correlations are found between news sentiment score (*Sentiment*) and the two socioeconomic factors: *Density* (r = –0.625, $p < 0.001$), and *Unemp* (r = 0.570, $p < 0.001$). Such a result provides good support to our hypothesis and establishes the overall nomological validity of our

results. The validated news sentiment can help in further revealing the dynamics between residents' attitudes and socioeconomic factors.

==========

Please insert Table 3 about here.

==========

Dynamics of Residents' Attitudes

The dynamics of residents' attitudes have been theoretically investigated by previous tourism studies. Among them, Doxey's Irridex model (Doxey 1975) has been widely cited as the theoretical foundation for subsequent resident's attitudes research (Sharpley 2014). The Irridex model considers the level of tourism development and suggests that an increase in the number of tourists would result in negative changes in resident's attitudes over time. Along with the development of a destination for tourism purposes, the residents' attitudes may evolve from initial euphoria and subsequent apathy towards final annoyance and antagonism in extreme cases.

However, the Irridex model has been criticized because it has rarely been directly tested or verified by empirical research findings. The news sentiment in our study enables an empirical test of the Irridex model from a longitudinal perspective. In the current context, we are particularly interested in discussing the dynamics of residents' attitudes in terms of long-run causality and equilibrium among three variables *Sentiment*, *Density*, and *Unemp*.

Long-run causality helps determine the direction of causality between residents' attitudes and socioeconomic factors, and long-run equilibrium is useful in confirming whether residents' attitudes drift upwards and downwards together with these factors. To discuss these dynamics, we employed Stata to conduct econometric analysis following prior research (Oh 2005; Lee and Brahmasrene 2013). The three variables are transformed through the use of natural logarithm to ease the interpretation of coefficients. Coefficients in the log function are

interpreted as elasticities, which illustrate a percentage change in a dependent variable given a 1% change in an independent variable.

(1) Long-run Causality

The Granger causality test via Vector Autoregression (VAR) modeling (Sims 1980) is used to discuss whether MCT density influences HK residents' attitudes or the converse. It enables such analysis by measuring the ability to predict the future values of a variable using prior values of another variable (Granger 1988). The Granger causality test requires stationary variables. The non-stationary series has a unit root and may cause spurious regression results. In this study, we applied Augmented Dickey-Fuller (ADF) tests (Dickey, Bell, and Miller 1986) to examine the non-stationarity. Results show that all variables *Sentiment* (ADT statistic = –2.28, $p > 0.1$), *Density* (ADT statistic = –2.48, $p > 0.1$) and *Unemp* (ADT statistic = –2.31, $p > 0.1$) are not stationary. However, the first differences of *Sentiment* (ADT statistic = –7.06, $p < 0.01$), *Density* (ADT statistic = –7.73, $p < 0.01$) and *Unemp* (ADT statistic = –6.13, $p < 0.01$) are all stationary at the 1 percent significance level[*].

Therefore, we used their first differences to construct VAR models for Granger causality tests (Oh 2005). The optimal lag 6 for the VAR model was determined by several information criteria including Akaike Information Criteria, Schwartz Bayesian Information Criteria, and Likelihood Ratio. Granger causality tests show that *Density* (F-statistic $_{\text{Density -> Sentiment}}$ = 19.76, $p<0.01$) and *Unemp* (F-statistic $_{\text{Unemp -> Sentiment}}$ = 14.26, $p<0.01$) are the Granger-cause of the *Sentiment* while their reverse is not the case (F-statistic $_{\text{Sentiment -> Density}}$ = 8.98, $p = 0.17$; F-statistic $_{\text{Sentiment –> Unemp}}$ = 8.82, $p = 0.18$), indicating that *Density* and *Unemp* lead to changes in *Sentiment*. Therefore, we are confident that, in the HK community, a one-way causality exists from MCT density and unemployment rate to news sentiment in the long run,

[*] For the ADT tests, the null hypothesis is about non-stationarity and the probability values are determined by comparing to the critical values suggested by MacKinnon (1996).

supporting the Irridex model.

(2) Long-run Equilibrium

The long-run equilibrium was examined through cointegration analysis and vector error correction models (VECM) (Lee and Brahmasrene 2013). According to Granger (1981), a linear combination of several non-stationary variables may be stationary. If such a linear combination exists, these non-stationary variables are claimed to be cointegrated, thereby suggesting a long-run equilibrium that prevents these variables from drifting away from each other. In the present study, we adopted the Johansen (1991) test with the constant and time trend to examine cointegration among *Sentiment*, *Density*, and *Unemp*. Table 4 lists the results of Johnsen test with trace statistics and max statistics, which consistently indicate that the three variables have cointegrated relations and there exists one cointegrating vector. The results of the cointegration analysis imply the presence of a long-run equilibrium among *Sentiment*, *Density*, and *Unemp,* although they may deviate from their long-run trend in the short run.

==========

Please insert Table 4 about here.

==========

According to Engle and Granger (1987), if non-stationary variables are cointegrated, we can formulate a VECM with a one-period correction to account for a long-run relationship. Therefore, we constructed four different models to study the long-run equilibrium for overall news sentiment and sentiments of the leftist, centrist, and rightist newspapers, respectively. Table 5 illustrates the coefficient estimations (i.e., the elasticities) of the cointegrating equations that represent long-run equilibriums among these variables. All coefficients are significant at the 0.01 level.

==========

Please insert Table 5 about here.

==========

According to Table 5, we can construct an Equation (2) to reflect such long-term equilibrium among overall news sentiment, MCT density, and unemployment rate.

$$\ln Sentiment_t = -1.87 \ln Density_t + 0.61 \ln Unemployment_t + 0.02t - 2.56 \quad (2)$$

Based on this long-run equilibrium in Equation (2), a 1% increase in MCT Density leads to a 1.87% decrease in news sentiment, suggesting a support to the Irridex model in the long run. Figure 4 depicts a clear negative relationship between news sentiment and MCT density in HK during a 13-year period.

==========

Please insert Figure 4 about here.

==========

Equation (2) also suggests that a 1% increase in unemployment rate yields a 0.61% increase in overall news sentiment. This finding is consistent with previous studies, which generally agreed that the creation of jobs and employment opportunities could lead to positive residents' attitudes towards tourism (Ivlevs 2017; Sharpley 2014). Figure 5 shows the strong positive relation between news sentiment and the unemployment rate in HK. During the periods with a higher unemployment rate, such as the 2003 SARS epidemic and the 2008 global financial crisis, local residents showed more positive views of MCTs, which is reflected in higher sentiment scores.

==========

Please insert Figure 5 about here.

==========

We also observed that some positive relations between news sentiment and MCT density during certain periods, such as the global financial crisis (January 2008 to January 2010). Such a phenomenon could be explained by the combined influences of MCT density and unemployment rate together. Specifically, during that period, the increased news sentiment with MCT density may be attributed to the fact that the positive influences of the increasing unemployment rate might outweigh the negative influences of the increasing MCT density. Accordingly, the identified equilibrium could resolve some contradictory results reported by previous cross-sectional studies.

Finally, Table 5 further shows that long-run equilibriums for different newspaper categories generally follow the patterns identified for the overall news sentiment: *Sentiment* is negatively influenced by *Density* and positively influenced by *Unemp*. However, the magnitudes of these influences across three newspaper categories differ. Compared with rightist ($\beta_{Density} = -1.99$) and centrist ($\beta_{Density} = -1.99$) newspapers, the sentiment of leftist ($\beta_{Density} = -1.34$) newspapers decreased the least with the increase of MCT density. Meanwhile, compared with leftist ($\beta_{Unemp} = 0.41$) newspapers, the sentiment of rightist ($\beta_{Unemp} = 1.12$) and centrist ($\beta_{Unemp} = 0.62$) newspapers grew much faster with the increase of unemployment rate. As such, we may conjecture that if we assume centrist newspapers have a relatively neutral standpoint, in the long run, leftist newspapers tend to mitigate the negative influence of MCT density, while rightist newspapers are inclined to focus more on the positive influence of unemployment rate.

CONCLUSIONS AND IMPLICATIONS

This study employed the news media sentiment analytics to reveal the evolving patterns of HK residents' attitudes towards tourism and to validate the results with typical socioeconomic factors. Our study shows that the overall news sentiment fluctuates between 0.30 and 0.98, with an average of 0.70, thereby demonstrating generally positive residents' attitudes towards tourism in HK. The inflection points of the overall news sentiment match the key socioeconomic events, and the overall news sentiment score is significantly correlated with typical socioeconomic factors, including MCT density and unemployment rate, which supports the face and nomological validities of our results. In addition, our study results reveal the long-run dynamics among news sentiment, MCT density, and unemployment rate in terms of a one-way causality and a long-run equilibrium. This study entails important academic contributions and managerial implications.

Academic Contributions

The research method employed in this study made an important contribution by adding a new dimension of residents' attitudes. The underlying big data analytics offers the possibility to addressing several major limitations of survey-based research.

First, big data is dynamic in nature and big data analytics inherently allows longitudinal studies, in contrast to the conventional cross-sectional survey approach (Lazer et al. 2009). Such characteristic lends our study nicely to directly resolve the long debate on the Irridex model in a mainstream tourism destination. Findings of this study empirically confirmed the long-run causality and equilibrium between residents' attitudes and socioeconomic factors in HK, thus providing strong evidence to support the Irridex model.

Second, big data ushers in a new model of empiricist epistemology, which is more open to using a hybrid combination of deductive and inductive approaches to advance the

understanding of a phenomenon by coping with the large volume of data (Kitchin 2014). In this work, we deductively proposed our news sentiment measure of residents' attitudes based on agenda-setting theory. We then inductively examined the evolving patterns of news sentiment and its long-run dynamics with typical socioeconomic factors. As a result, the analysis framework with the news sentiment measure provides a conceptually important and complementary perspective to existing research on residents' attitudes. The framework has the potential to be generalizable to other media-active contexts, where discussions on tourism impacts are woven into social dialogues, via traditional and potentially social media channels. Such efforts can also facilitate and connect discussions on tourism issues with broader socioeconomic phenomena.

Third, big data is characterized as seeking to be exhaustive and fine-grained in scope. In this study, we utilized all, instead of a sample of, relevant Chinese news coverage in HK to explore the evolving pattern of news sentiment. The dataset was intended to be exhaustive. The resulting sentiment was fine-grained to each piece of news, enabling a variety of advanced data analyses according to different time intervals. Accordingly, it is now possible to subsequently convert them into valuable research tools for tourism researchers.

Managerial Implications

The findings suggest a number of managerial implications. First, the unveiled evolving pattern of news sentiment provides useful information for tourism policymakers to a better understanding of the host community's attitudes, as well as to identify policy impacts on local residents. The study provides useful baseline data on HK residents' attitudes for future comparison.

Second, unlike prior survey-based research, the developed sentiment analysis framework and the news sentiment measure can be implemented as a practical social listening tool by

destination management organizations (DMOs), such as the HK Tourism Commission, to periodically monitor the dynamics of residents' attitudes towards tourism. They are also relevant to the current global discussion on the phenomenon of "overtourism" as similar dissatisfaction of local communities with foreign tourists and tourism development, in general, has been documented across many popular destinations, such as Barcelona, Berlin, and Venice (Ivlevs 2017; Pidd 2011; Ross 2015).

Third, as we revealed the long-run equilibrium among residents' attitudes, tourist density, and unemployment rate, tourism policymakers and practitioners should monitor these factors and develop appropriate strategies accordingly. Our study results also indicate that tourism policies governing MCTs have coincided with each inflection point of news sentiment, demonstrating the potential pivotal influence of policy-related issues on residents' attitudes, which deserves more attention from tourism government officers and practitioners.

Finally, our study emphasizes the close nexus between residents' attitudes and news coverage. Tourism practitioners and policymakers should pay more attention to news media and monitor the media discourse to understand residents' attitudes towards tourism. Doing so will help policymakers acquire an in-depth understanding of public concerns regarding tourism development. Moreover, such action can benefit the practitioners in establishing public relation campaigns that cope with specific tourism events in advance and gain community support with a long-term perspective.

Limitations and Future Research

This study attempts to add a vital dimension to current residents' attitudes research from a big data analytics perspective in general and the news media sentiment analytics in particular. However, it still suffers several limitations. First, our study focuses on exploring the evolving patterns of residents' attitudes by employing the news media sentiment analytics. While the

method excels at accurately unveiling dynamic nature of a phenomenon, it fails to explain the nomological network of residents' attitudes, an important subject of prior research. Therefore, combining big data analytics with classical tourism research methods could be a promising direction to disclose comprehensive knowledge of residents' attitudes towards tourism. Furthermore, the current framework only considers positive and negative sentiments; the extension of binary polarities to continuous valences including the neutral sentiment may further enhance the validity of news media analytics for residents' attitudes. In addition, in this study, we focused on long-run causality and equilibrium between residents' attitudes and socioeconomic factors; their short-run dynamics with the inclusion of more control variables also deserve future investigations.

Second, in this study, we only focused on news media that have long been regarded as one of the most prominent discursive sites that allow researchers to understand framing and its underlying mechanism (D'Angelo and Kuypers 2009). In light of the rapid growth of social media in recent years, researchers have increasingly employed social media sentiment analysis in addressing tourism and hospitality problems (Duan et al. 2016; Geetha, Singha, and Sinha 2017; Luo and Zhai 2017). The consolidation of news coverage with user-generated content in social media would form a promising "big data" research direction towards further understanding residents' attitudes towards tourism.

REFERENCES

Alaei, Ali Reza, Susanne Becken, and Bela Stantic. 2019. “Sentiment Analysis in Tourism: Capitalizing on Big Data.” *Journal of Travel Research* 58 (2): 75–191. doi:10.1177/0047287517747753.

Anderson, Simon P., and John McLaren. 2012. “Media Mergers and Media Bias with Rational Consumers.” *Journal of the European Economic Association* 10 (4): 831–59. doi:10.1111/j.1542-4774.2012.01069.x.

Bimonte, Salvatore, and Valeria Faralla. 2016. “Does Residents’ Perceived Life Satisfaction Vary with Tourist Season? A Two-Step Survey in a Mediterranean Destination.” *Tourism Management* 55: 199–208. doi:10.1016/j.tourman.2016.02.011.

Brown, Carol B. 2015. “Tourism, Crime and Risk Perception: An Examination of Broadcast Media’s Framing of Negative Aruban Sentiment in the Natalee Holloway Case and Its Impact on Tourism Demand.” *Tourism Management Perspectives* 16: 266–77. doi:10.1016/j.tmp.2014.12.001.

Butler, R.W. 1980. “The Concept of a Tourist Area Cycle of Evolution: Implications for Management of Resources.” *Canadian Geographer* 24 (1): 5–12. doi:10.1111/j.1541-0064.1980.tb00970.x.

Cambria, Erik. 2016. “Affective Computing and Sentiment Analysis.” *IEEE Intelligent Systems* 31 (2): 102–7. doi:10.1109/MIS.2016.31.

Castelltort, Magí, and Gabriela Mäder. 2010. “Press Media Coverage Effects on Destinations – A Monetary Public Value (MPV) Analysis.” *Tourism Management* 31 (6): 724–38. doi:10.1016/j.tourman.2009.06.007.

Chan, Joseph M., and Francis L. F. Lee. 2007. “Media and Politics in Post-Handover Hong Kong: An Introduction.” *Asian Journal of Communication* 17 (2): 127–33. doi:10.1080/01292980701306456.

Chen, Nan, Cathy H. C. Hsu, and Xiang (Robert) Li. 2018. “Feeling Superior or Deprived?

Attitudes and Underlying Mentalities of Residents towards Mainland Chinese Tourists." *Tourism Management* 66 (June): 94–107. doi:10.1016/j.tourman.2017.11.007.

Cohen, Judy. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20: 37–46.

D'Angelo, Paul, and Jim A. Kuypers. 2009. *Doing News Framing Analysis: Empirical and Theoretical Perspectives*. New York: Routledge.

Deery, Margaret, Leo Jago, and Liz Fredline. 2005. "A Framework for the Development of Social and Socioeconomic Indicators for Sustainable Tourism in Communities." *Tourism Review International* 9 (1): 69–77. doi:10.3727/154427205774791780.

———. 2012. "Rethinking Social Impacts of Tourism Research: A New Research Agenda." *Tourism Management* 33 (1): 64–73. doi:10.1016/j.tourman.2011.01.026.

Dickey, David A., William R. Bell, and Robert B. Miller. 1986. "Unit Roots in Time Series Models: Tests and Implications." *The American Statistician* 40 (1): 12–26. doi:10.1080/00031305.1986.10475349.

Doxey, G.V. 1975. "A Causation Theory of Visitor-Resident Irritants: Methodology and Study Inferences." In *Proceedings of the Sixth Annual Conference of the Travel Research Association*, 195–98. San Diego, CA: Travel and Tourism Research Association.

Duan, Wenjing, Yang Yu, Qing Cao, and Stuart Levy. 2016. "Exploring the Impact of Social Media on Hotel Service Performance: A Sentimental Analysis Approach." *Cornell Hospitality Quarterly* 57 (3): 282–96.

Engle, Robert F., and Clive W J Granger. 1987. "Cointegration and Error Correction Representation, Estimation and Testing." *Econometrica* 55: 251–76.

Entman, Robert M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43 (4): 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x.

Faulkner, Bill, and Carmen Tideswell. 1997. “A Framework for Monitoring Community Impacts of Tourism.” *Journal of Sustainable Tourism* 5 (1): 3–28.

Fu, Yu, Jin-Xing Hao, Xiang (Robert) Li, and Cathy H.C. Hsu. 2019. “Predictive Accuracy of Sentiment Analytics for Tourism: A Metalearning Perspective on Chinese Travel News.” *Journal of Travel Research* 58 (4): 666–79. doi:10.1177/0047287518772361.

García, Fernando Almeida, Antonia Balbuena Vázquez, and Rafael Cortés Macías. 2015. “Resident’s Attitudes towards the Impacts of Tourism.” *Tourism Management Perspectives* 13: 33–40.

Geetha, M., Pratap Singha, and Sumedha Sinha. 2017. “Relationship between Customer Sentiment and Online Customer Ratings for Hotels - An Empirical Analysis.” *Tourism Management* 61: 43–54. doi:10.1016/j.tourman.2016.12.022.

Getz, Donald. 2008. “Event Tourism: Definition, Evolution, and Research.” *Tourism Management* 29 (3): 403–28. doi:10.1016/j.tourman.2007.07.017.

Granger, Clive W J. 1981. “Some Properties of Time Series Data and Their Use in Econometric Model Specification.” *Journal of Econometrics* 16: 121–30.

———. 1988. “Some Recent Developments in a Concept of Causality.” *Journal of Econometrics* 39: 199–211.

Harrill, Rich. 2004. “Residents’ Attitudes toward Tourism Development: A Literature Review with Implications for Tourism Planning.” *Journal of Planning Literature* 18 (3): 251–66. doi:10.1177/0885412203260306.

Hiller, Harry H., and Richard A. Wanner. 2011. “Public Opinion in Host Olympic Cities: The Case of the 2010 Vancouver Winter Games.” *Sociology* 45 (5): 883–99. doi:10.1177/0038038511413414.

HKSAR Government. 2016. “Hong Kong: The Facts - The Media.” Government Report. Hong Kong. http://www.gov.hk/en/about/abouthk/factsheets/docs/media.pdf.

Hwang, Doohyun, William P. Stewart, and Dong-wan Ko. 2012. “Community Behavior and Sustainable Rural Tourism Development.” *Journal of Travel Research* 51 (3): 328–41.

Ivlevs, Artjoms. 2017. “Happy Hosts? International Tourist Arrivals and Residents’ Subjective Well-Being in Europe.” *Journal of Travel Research* 56 (5): 599–612. doi:10.1177/0047287516662353.

Jiang, Qiaolei. 2013. “Celebrity Athletes, Soft Power and National Identity: Hong Kong Newspaper Coverage of the Olympic Champions of Beijing 2008 and London 2012.” *Mass Communication and Society* 16 (6): 888–909. doi:10.1080/15205436.2013.789528.

Johansen, S. 1991. “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models.” *Econometrica* 59: 1551–80.

Kang, Hanhoon, Seong Joon Yoo, and Dongil Han. 2012. “Senti-Lexicon and Improved Naïve Bayes Algorithms for Sentiment Analysis of Restaurant Reviews.” *Expert Systems with Applications* 39 (5): 6000–6010. doi:10.1016/j.eswa.2011.11.107.

Kapuściński, Grzegorz, and Barry Richards. 2016. “News Framing Effects on Destination Risk Perception.” *Tourism Management* 57: 234–44. doi:10.1016/j.tourman.2016.06.017.

Kim, Kyungmi, Muzaffer Uysal, and M. Joseph Sirgy. 2013. “How Does Tourism in a Community Impact the Quality of Life of Community Residents?” *Tourism Management* 36 (June): 527–40. doi:10.1016/j.tourman.2012.09.005.

Kim, Tae-Kyun, Choong-Ki Lee, James W. Mjelde, and Hye-Mi Lee. 2014. “The Carryover Effect of Newspaper Reports on a Mega Event: Ex Post Analysis of the 2012 Expo Yeosu Korea.” *Asia Pacific Journal of Tourism Research* 19 (9): 1009–22. doi:10.1080/10941665.2013.833126.

Kirilenko, Andrei P., Svetlana O. Stepchenkova, Hany Kim, and Xiang (Robert) Li. 2018.

"Automated Sentiment Analysis in Tourism: Comparison of Approaches." *Journal of Travel Research* 57 (8): 1012–25. doi:10.1177/0047287517729757.

Kitchin, Rob. 2014. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1 (1): 1–12. doi:10.1177/2053951714528481.

Kumar, B T Sampath, and J N Prakash. 2009. "Precision and Relative Recall of Search Engines: A Comparative Study of Google and Yahoo." *Singapore Journal of Library & Information Management* 38 (1): 124–37.

Landis, J. R., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33: 159–74. doi:10.2307/2529310.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. "Computational Social Science." *Science* 323 (5915): 721–23. doi:10.1126/science.1167742.

Lee, Chou-Yuan, and Zne-Jung Lee. 2012. "A Novel Algorithm Applied to Classify Unbalanced Data." *Applied Soft Computing* 12 (8): 2481–85. doi:10.1016/j.asoc.2012.03.051.

Lee, Jung Wan, and Tantatape Brahmasrene. 2013. "Investigating the Influence of Tourism on Economic Growth and Carbon Emissions: Evidence from Panel Analysis of the European Union." *Tourism Management* 38: 69–76. doi:10.1016/j.tourman.2013.02.016.

Legislative Council Secretariat. 2014. "Individual Visit Scheme." 6. Research Brief. Hong Kong: Legislative Council.

Lester, Marilyn. 1980. "Generating Newsworthiness: The Interpretive Construction of Public Events." *American Sociological Review* 45 (6): 984–94.

Leung, Daniel, Rob Law, Hubert van Hoof, and Dimitrios Buhalis. 2013. "Social Media in Tourism and Hospitality: A Literature Review." *Journal of Travel & Tourism*

Marketing 30 (1–2): 3–22.

Li, Xiang (Robert), Cathy H. C. Hsu, and Laura J. Lawton. 2015. "Understanding Residents' Perception Changes toward a Mega-Event through a Dual-Theory Lens." *Journal of Travel Research* 54 (3): 396–410. doi:10.1177/0047287513517422.

Liu, Bing. 2012. "Sentiment Analysis and Opinion Mining." *Synthesis Lectures on Human Language Technologies* 5 (1): 1–167. doi:10.2200/S00416ED1V01Y201204HLT016.

Liu, Bingjie, and Lori Pennington-Gray. 2015. "Bed Bugs Bite the Hospitality Industry? A Framing Analysis of Bed Bug News Coverage." *Tourism Management* 48: 33–42. doi:10.1016/j.tourman.2014.10.020.

Liu, Shaowu, Rob Law, Jia Rong, Gang Li, and John Hall. 2013. "Analyzing Changes in Hotel Customers' Expectations by Trip Mode." *International Journal of Hospitality Management* 34 (1): 359–71. doi:10.1016/j.ijhm.2012.11.011.

Luo, Qiuju, and Xueting Zhai. 2017. "'I Will Never Go to Hong Kong Again!' How the Secondary Crisis Communication of 'Occupy Central' on Weibo Shifted to a Tourism Boycott." *Tourism Management* 62: 159–72. doi:10.1016/j.tourman.2017.04.007.

Mackinnon, J.G. 1996. "Numerical Distribution Functions for Unit Root and Cointegration Tests." *Journal of Applied Econometrics* 11: 601–18.

McCombs, Maxwell E., and Donald L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *Public Opinion Quarterly* 36 (2): 176–87. doi:10.1086/267990.

Molotch, Harvey, and Marilyn Lester. 1974. "News as Purposive Behavior: On the Strategic Use of Routine Events, Accidents, and Scandals." *American Sociological Review* 39 (1): 101–12. doi:10.2307/2094279.

Moraes, Rodrigo, João Francisco Valiati, and Wilson P. Gavião Neto. 2013. "Document-Level Sentiment Classification: An Empirical Comparison between SVM and ANN." *Expert Systems with Applications* 40 (2): 621–33. doi:10.1016/j.eswa.2012.07.059.

Moznette, James, and Galen Rarick. 1968. “Which Are More Readable: Editorials or News Stories?” *Journalism Quarterly* 45: 319–21. doi:10.1177/107769906804500214.

Neresini, Federico, and Andrea Lorenzet. 2016. “Can Media Monitoring Be a Proxy for Public Opinion about Technoscientific Controversies? The Case of the Italian Public Debate on Nuclear Power.” *Public Understanding of Science* 25 (2): 171–85. doi:10.1177/0963662514551506.

Nunkoo, Robin, Stephen L. J. Smith, and Haywantee Ramkissoon. 2013. “Residents’ Attitudes to Tourism: A Longitudinal Study of 140 Articles from 1984 to 2010.” *Journal of Sustainable Tourism* 21 (1): 5–25. doi:10.1080/09669582.2012.673621.

Nunkoo, Robin, and Kevin Kam Fung So. 2016. “Residents’ Support for Tourism: Testing Alternative Structural Models.” *Journal of Travel Research* 55 (7): 847–61. doi:10.1177/0047287515592972.

Oh, Chi-Ok. 2005. “The Contribution of Tourism Development to Economic Growth in the Korean Economy.” *Tourism Management* 26 (1): 39–44. doi:10.1016/j.tourman.2003.09.014.

Peel, Victoria, and Adam Steen. 2007. “Victims, Hooligans and Cash-Cows: Media Representations of the International Backpacker in Australia.” *Tourism Management* 28 (4): 1057–67. doi:10.1016/j.tourman.2006.08.012.

Philander, Kahlil, and Yun Ying Zhong. 2016. “Twitter Sentiment Analysis: Capturing Sentiment from Integrated Resort Tweets.” *International Journal of Hospitality Management* 55: 16–24. doi:10.1016/j.ijhm.2016.02.001.

Pidd, H. 2011. “Without Tourists, Berlin Is Stuffed. But Try Telling That to the Angry Natives.” *The Guardian*, May 9. http://www.theguardian.com/commentisfree/2011/may/09/berliners-angry-over-tourists.

Robertson, Martin, and Phil Rogers. 2009. “Festivals, Cooperative Stakeholders and the Role of the Media: A Case Analysis of Newspaper Media.” *Scandinavian Journal of Hospitality and Tourism* 9 (2–3): 206–24. doi:10.1080/15022250903217019.

Ross, W. 2015. “The Death of Venice: Corrupt Officials, Mass Tourism and Soaring Property Prices Have Stifled Life in the City.” *The Independent*, May 14. http://www.independent.co.uk/news/world/europe/the-death-of-venice-corrupt-officials-mass-tourism-and-soaring-property-prices-have-stifled-life-in-10251434.html.

Santos, Carla Almeida. 2004. “FRAMING PORTUGAL: Representational Dynamics.” *Annals of Tourism Research* 31 (1): 122–38. doi:10.1016/j.annals.2003.08.005.

Scheufele, Dietram A., and David Tewksbury. 2007. “Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models.” *Journal of Communication* 57 (1): 9–20. doi:10.1111/j.0021-9916.2007.00326.x.

Schuckert, Markus, Xianwei Liu, and Rob Law. 2015. “Hospitality and Tourism Online Reviews: Recent Trends and Future Directions.” *Journal of Travel & Tourism Marketing* 32 (5): 608–21. doi:10.1080/10548408.2014.933154.

Sharpley, Richard. 2014. “Host Perceptions of Tourism: A Review of the Research.” *Tourism Management* 42 (June): 37–49. doi:10.1016/j.tourman.2013.10.007.

Sims, C. 1980. “Macroeconomics and Reality.” *Econometrica* 48: 1–48.

Siu, Sai-wo. 2017. “Newspaper Titles Retain Value.” *The Standard*, December 11. http://www.thestandard.com.hk/section-news.php?id=190547.

Snaith, Tim, and Art Haley. 1999. “Residents’ Opinions of Tourism Development in the Historic City of York, England.” *Tourism Management* 20 (5): 595–603. doi:10.1016/S0261-5177(99)00030-8.

Stepchenkova, Svetlana, and James S. Eales. 2011. “Destination Image as Quantified Media

Messages: The Effect of News on Tourism Demand." *Journal of Travel Research* 50 (2): 198–212. doi:10.1177/0047287510362780.

Uscinski, Joseph E. 2009. "When Does the Public's Issue Agenda Affect the Media's Issue Agenda (and Vice-Versa)? Developing a Framework for Media-Public Influence." *Social Science Quarterly* 90 (4): 796–815. doi:10.1111/j.1540-6237.2009.00663.x.

Uysal, Muzaffer, M. Joseph Sirgy, Eunju Woo, and Hyelin (Lina) Kim. 2016. "Quality of Life (QOL) and Well-Being Research in Tourism." *Tourism Management* 53: 244–61. doi:10.1016/j.tourman.2015.07.013.

Van De Kauter, Marjan, Diane Breesch, and Véronique Hoste. 2015. "Fine-Grained Analysis of Explicit and Implicit Sentiment in Financial News Articles." *Expert Systems with Applications* 42: 4999–5010. doi:10.1016/j.eswa.2015.02.007.

Vargas-Sánchez, Alfonso, Nuria Porras-Bueno, and María de los Ángeles Plaza-Mejía. 2014. "Residents' Attitude to Tourism and Seasonality." *Journal of Travel Research* 53 (5): 581–96. doi:10.1177/0047287513506295.

Wang, Hongwei, Pei Yin, Jiani Yao, and James N. K. Liu. 2013. "Text Feature Selection for Sentiment Classification of Chinese Online Reviews." *Journal of Experimental & Theoretical Artificial Intelligence* 25 (4): 425–39. doi:10.1080/0952813X.2012.721139.

Wikipedia. 2018. "List of Newspapers in Hong Kong." In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_newspapers_in_Hong_Kong.

Wu, Bihu, Lan Xue, Alastair M. Morrison, and Xi Y. Leung. 2012. "Frame Analysis on Golden Week Policy Reform in China." *Annals of Tourism Research* 39 (2): 842–62. doi:10.1016/j.annals.2011.10.002.

Ye, Qiang, Ziqiong Zhang, and Rob Law. 2009. "Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches." *Expert Systems*

with Applications 36: 6527–35. doi:10.1016/j.eswa.2008.07.035.

Xiang, Zheng, Qianzhou Du, Yufeng Ma, and Weiguo Fan. 2017. “A Comparative Analysis of Major Online Review Platforms : Implications for Social Media Analytics in Hospitality and Tourism.” *Tourism Management* 58: 51–65. doi:10.1016/j.tourman.2016.10.001.

Zhang, Yan, Yong Zhang, Jennifer Xu, Chunxiao Xing, and Hsinchun Chen. 2016. “Sentiment Analysis on Chinese Health Forums: A Preliminary Study of Different Language Models.” In *International Conference on Smart Health*, 68–81. Springer, Cham. doi:10.1007/978-3-319-29175-8_7.

TABLES

Table 1. Positive and Negative News Frequency Distribution by Year

Year	Negative	Positive	Total
2003	1382	5345	6727
2004	1702	6451	8153
2005	1030	2727	3757
2006	2684	5053	7737
2007	544	2042	2586
2008	481	1420	1901
2009	465	2190	2655
2010	697	2733	3430
2011	2125	4484	6609
2012	3134	3030	6164
2013	3585	3202	6787
2014	2709	4049	6758
2015	3339	6152	9491

Table 2. Major Socioeconomic Events related to MCTs

Period	Events	Descriptions
2002-2003	Severe Acute Respiratory Syndrome (SARS) epidemic	This serious infectious disease that has caused damage worldwide greatly influenced Hong Kong as well.
July 2003	Individual Visit Scheme (IVS)	The agreement between the HKSAR Government and the Central Government aimed to remove the tourist quota and encourage MCTs to visit Hong Kong.
Oct 2006	IVS expansion	The Central Government announced the expansion of IVS to 49 cities in Mainland China starting January 1, 2007.
July 2007	Amendment of the Closer Economic Partnership Arrangement (CEPA)	As part of the celebration of the 10th anniversary of Hong Kong's return to China, CEPA was expanded to further liberalize trade in services for Hong Kong with relaxed tourism market access conditions from one (i.e., Guangdong) to 9 provinces.
2008-2010	Global financial crisis	The worldwide financial crisis greatly impacted the economy of Hong Kong.
July 2009	Multiple-entry Individual Visit Endorsements (M-permit) for Shenzhen residents	The M-permit was launched on April 1, 2009, for permanent residents of Shenzhen and then extended to non-permanent residents of Shenzhen on Jan 1, 2010.
Mar 2013	Restrictions on quantity of powdered formula for persons departing Hong Kong	The HKSAR Government issued the import and export regulation to restrict the purchases of infant milk formula by MCTs, so as to protect the local supply to meet the demands of local residents.
April 2015	One trip per week for Shenzhen residents	The M-permit was limited to one trip per week to balance Hong Kong's tourism capacity and its visitor arrivals.

Table 3. Descriptive Statistics and Correlations of Variables

Variable	N	Mean	SD	Sentiment	Density	Unemp
Sentiment	156	0.703	0.128	-		
Density	156	0.286	0.153	*–0.625***	-	
Unemp	156	0.045	0.147	*0.570***	*–0.710***	-

Notes: *** $p < 0.01$; Sentiment: News sentiment (monthly); Density: MCT density (monthly); Unemp: Unemployment rate (monthly)

Table 4. Results of the Johnsen Test for Cointegration

Maximum rank (No. of cointegrating equations)	Trace statistic	%5 Critical value for the trace statistic	Max statistic	%5 Critical value for the max statistic
0	47.25**	34.55	38.04**	23.78
1	9.21	18.17	5.71	16.87
2	3.50	3.74	3.0	3.74

Notes: ** $p < 0.05$; Optimal lag is set to 6 according to the VAR modeling. Probability values for rejection of the null hypothesis of no integration are estimated based on MacKinnon et al. (1999).

Table 5. Coefficient Estimation of VECMs

Variable	Coefficient of Cointegrating Equation (Sentiment as the dependent variable)			
	Overall	Leftist	Centrist	Rightist
Density	–1.87***	–1.34***	–1.99***	–1.99***
Unemp	0.61***	0.41***	0.62***	1.12***
Time trend	0.02***	0.02***	0.02***	0.02***
Constant	–2.56	–1.87	–2.82	–1.56
R^2	0.40***	0.44***	0.41***	0.40***
F-statistic	89.09	103.79	92.36	88.67

Notes: *** $p < 0.01$; Sentiment: News sentiment; Density: MCT density; Unemp: Unemployment rate

FIGURES

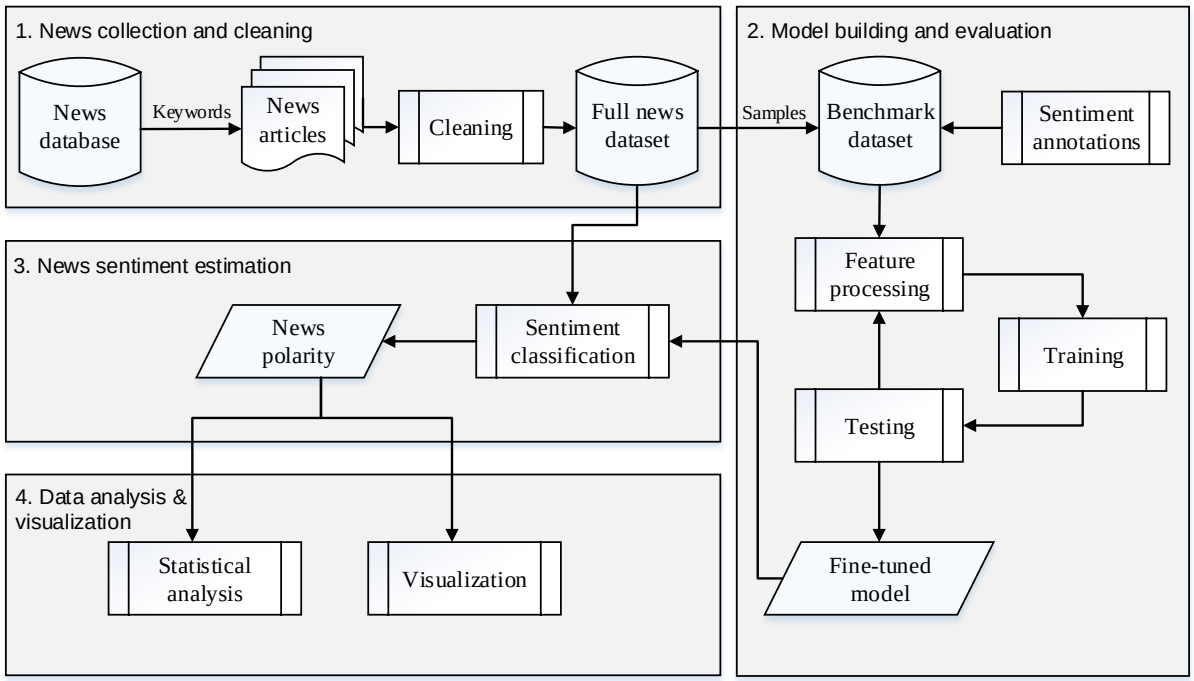

Figure 1. The news media sentiment analysis framework

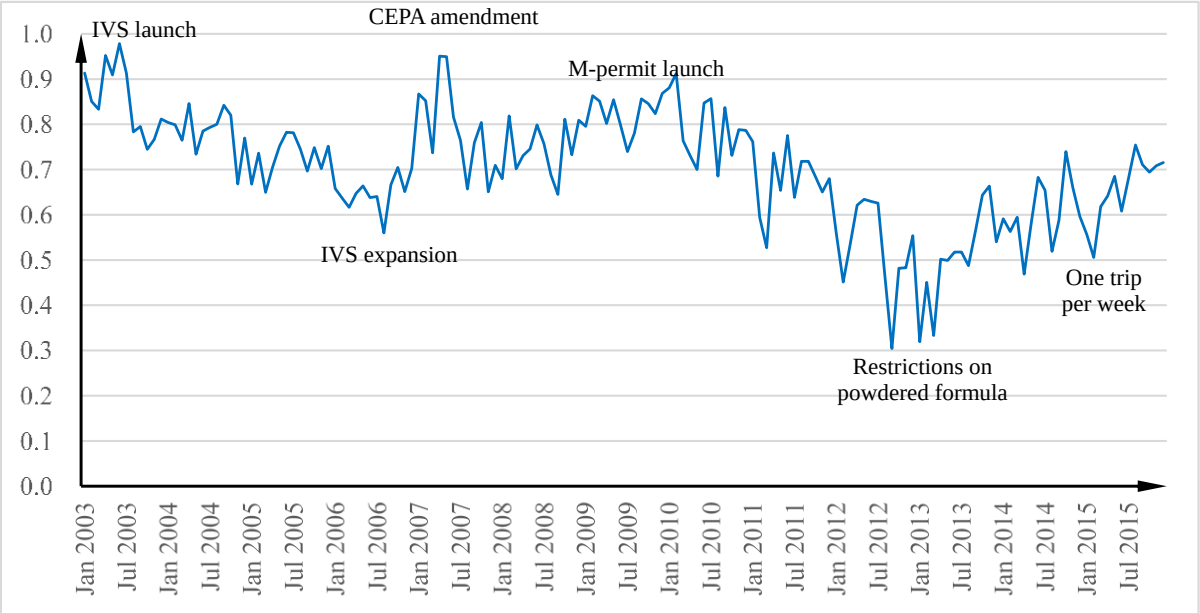

Figure 2. The evolving pattern of HK news sentiment with the announcement of major government policies related to MCTs

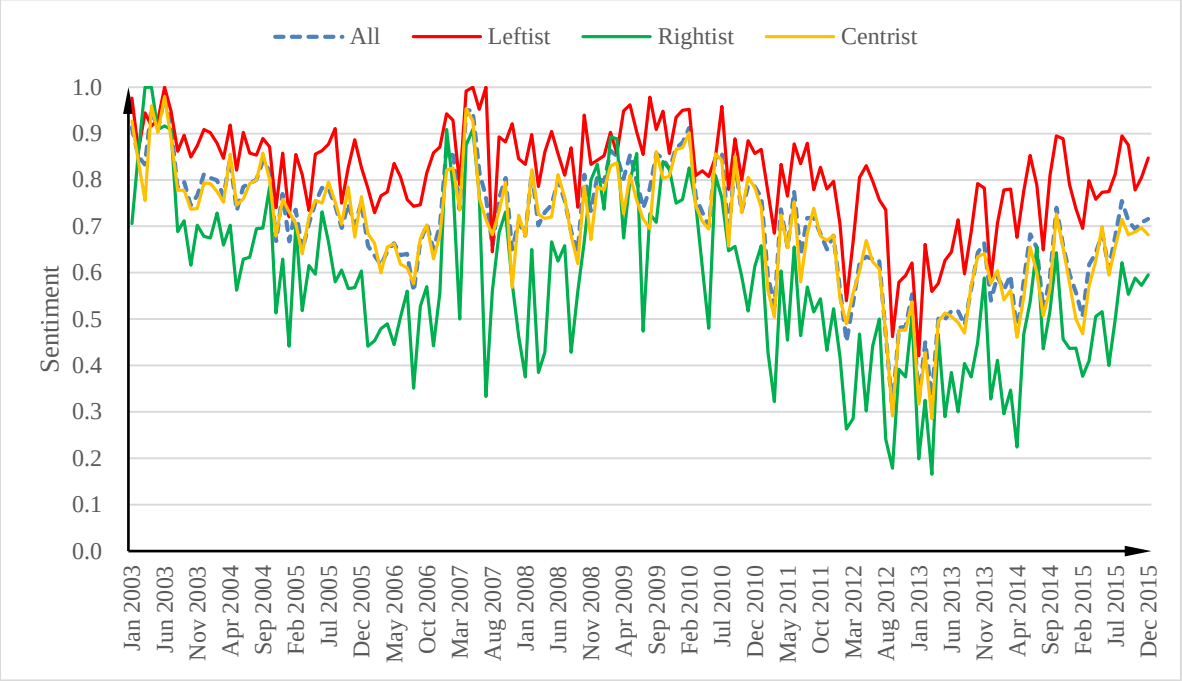

Figure 3. The evolving pattern of HK news sentiment across newspaper categories

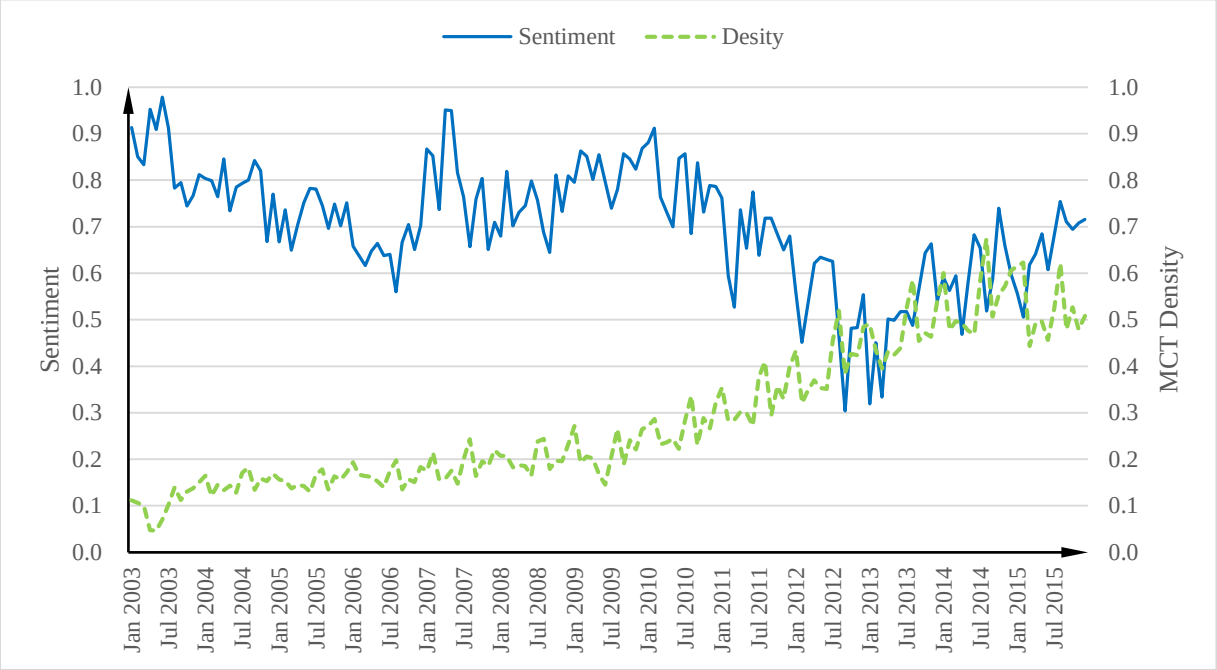

Figure 4. HK news sentiment with MCT density

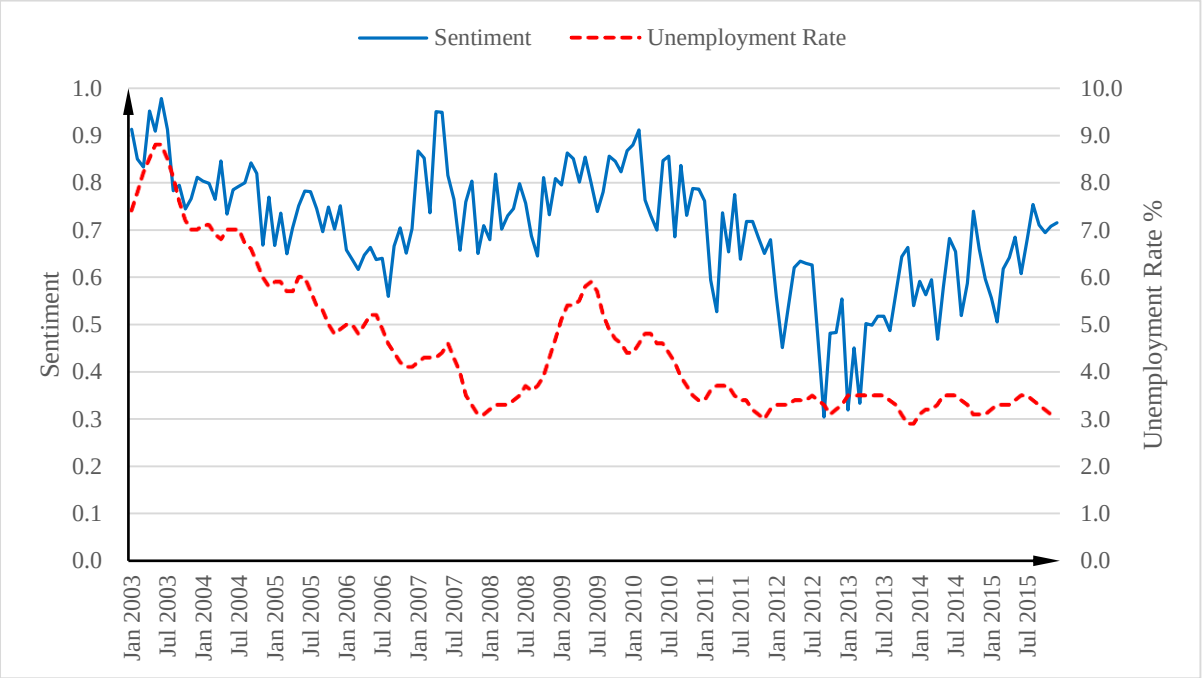

Figure 5. HK news sentiment with unemployment rate