Attention to Bitcoin

Amirhossein Sadoghi [*]
ESC Rennes School of Business

February 15, 2020

Please do not circulate or distribute this draft of the paper

Abstract

In this paper, we study the cause of the burst of the Bitcoin bubble rooted in the impacts of attention to news media on Bitcoin. We analyse textual data taken from news articles on Bitcoin and investigate the predictive and causal power of extracted information to model the dynamics of Bitcoin prices. In doing so, we apply the Latent Dirichlet Allocation (LDA) model to classify news article into some topics and measure the unusualness of each topic in a daily basis. We implement a regression discontinuity design to make an inference on how news media about Bitcoin has shaped changing dynamics in Bitcoin prices. From this quasi-natural experiment, we find that early on, the flow of information on blockchain and Fintech shifted the attention of traders to an unregulated market like Bitcoin, consequently expanding the bubble. During the bubble phase, uncertainty surrounding the security of this digital currency resulted in the bursting of the bubble.

Keywords: Cryptocurrency, Bitcoin crash, Bubble, Discontinuity Regression Design, Text mining, LDA, Natural Language Processing (NLP).
JEL classification: D80, D84, G14, C55.

[*] Corresponding Author, E.mail: amirhossein.sadoghi@rennes-sb.com . We are grateful for the comments of participants of the 13th German Probability and Statistics Days Freiburger Stochastik-Tage, February, 2018.

.

1 Introduction

After the dot-com crash, the market observed several ”internet stock” failures. At the end of 2017, the majority of cryptocurrencies like Bitcoin suffered the same fate. The resilience of this bubble stems from the attention of market participants governed by media. At the time, the flow information about Bitcoin influenced individual incentives to time the market, ultimately resulting in the persistence of bubbles over a considerable period. The growth and bursting of a bubble of this type of market denotes the investors' inability to coordinate their trading strategies. Generally, some events related to the global economy or to a specific market can generate informational flow. This flow has significant impacts dependent on the main messages carried and can result in the growth or bursting of a bubble. Preferably, market participants wish to depart this market prior to a bubble crash. Rational arbitrageurs may know that a market will finally break down but in the meantime would like to ride the bubble as long as it is growing. In between, the media act as a vehicle to direct the attention of investors. The limited attention or (selective) inattention of investors with existing the searching cost lead to rational arbitrageurs stay in the market for a long time. However, in this limited-attention environment, the media which convey specific information might attract the attention of rational arbitrageurs and facilitate synchronized decisions on entering or exiting such a market. The main question concerns how media exposure with regard to its intrinsic informational content can influence investors' decisions to enter or exit the market.

This study uses an empirical design to address above questions regarding the causes of the growth and bursting of the Bitcoin bubble. To address these challenges, we develop a new framework based on a set of Regression Discontinuity (RD) experiments associated with attention to news information. We measure attention to news media characterised by entropy or unusualness (Glasserman and Mamaysky, 2017) in informational flows and analyse its predictive power in affecting prices. We show that mass media news focused on different subjects influences the behaviours of market participants. In doing so, we classify news items on Bitcoin into a group of subjects. We then investigate the effects of each group of news subjects to identify discontinuity in the dynamics of Bitcoin prices. The main goal of this research is to shed new light on recent debates on the Bitcoin bubble through an examination of the effects of attention to Bitcoin governed by the media.

Generally speaking, any major event creates a cascade of events and generates an informational flow. This flow can consist of several messages corresponding to different subjects ans carries some primary messages as well as noise. As a result, major concepts might not be interpreted precisely and it is necessary to extract information from the flow. With helping of topic modelling approach, we decompose informational flows into classes of informational flows for each subject. Each component of this informational flow can be assigned to different classes of subjects based on a probabilistic model. The amount of information given on a subject and the degree to which uncertainty on each topic can be reduced is measured with applying entropy theory.

For this study, we run a quasi-natural experiment to understand the impacts of each topic of informational flow on the dynamics of Bitcoin prices under a regression discontinuity framework. Our empirical methodology relies on comparing Bitcoin prices set around the bubble event as well as before and after the event. We specify three exogenous cut-off points. Our regression discontinuity methodology focuses only on price information disseminated around cut-off points to ensure that we only measure the effects of the informational flows of each news topic. It is worthwhile to note key underlying assumptions behind our approach to deal with several endogeneity problems might be critical issues for the analysis of cause and effect relationships. The results of the RD design technique can be biased by endogeneity issues like omitting variables or reverse causality. In the experiment design section, validation subsection 6.1, we address these issues and explain our strategy to deal with endogeneity problems. We first verify that informational flows are continuous around cut-off points, and we then examine how these flows behave around the known cut-off points. This design enables us to make causative inferences about the effects of market attention on digital currencies. The key purpose of our empirical design is to verify the smoothness of decomposed news topics across the three defined cut-off points on the price of Bitcoin.

However, we first describe in more detail how we collect unstructured textual information and our initial preparations for the primary analysis. We have employed several strategies to select relevant news items about Bitcoin. To avoid endogeneity (simultaneity) problems in the estimation of the causal inferences, we exclude news items about Bitcoin Price. Each news article examined includes information on different subjects and can be classified into different categories. These classes or categories can represent different concepts known as topics. A topic is ”a bag word”. We interpret each topic explored in the news articles with regard to the meaning of its collection of words. News items arrive continuously, and to ensure consistency with market data, we create aggregated daily informational flows.

We then estimate attention to news media with measuring the entropy of the informational flows. The entropy of information flows of news topics is used as a running variable in RD design models. In fact, in a highly volatile environment like the Bitcoin market, with increased variance in price, attention to an unusual event might be decreased. This inattention of the market is mainly rooted in the existence of barriers and constraints in the receiving and processing of information of an event. Consequently, responses to the shocks of an event are not as accurate as they were before. Indeed, it is not realistic to assume that all market participants receive the same signal about the market or that they interpret the same signals in the same way. As a result, investors may predict the likelihood of high volatility in the future and may adjust their actions to mitigate the consequences. Hence, we also consider a fuzzy RD design approach which can identify not fully treated observations.

Our study framework is consistent with models on the rational inattention behaviours of market participants proposed by Sims (2003) and particularly for the bubble phase (Giglio et al., 2016; Moinas and Pouget, 2013; Lei et al., 2001). Beyond this theoretical link to the rational inattention model, we present an empirical relationship. Following by (Glasserman and Mamaysky, 2017), we measure attention to news media with calculating the entropy or unusualness of informational flows. Glasserman and Mamaysky (2017) employ the entropy of terms in news articles as predictors to forecast the market volatility. In contrast, we classify terms into topics and use the entropy of topics as variables. It helps to interpret a term correctly with different meanings associated with other terms in a topic.

As one takeaway from this study, we find that attention to news about blockchain technologies and Fintech systematically increases the price of Bitcoin. During the bubble phase, investors enjoyed riding the bubble until they received negative signals about the security of this digital currency. After the bubble phase, the media still had an influence on the market but was publishing news items about the regulation of this market, as central banks entered this market to regulate it. From a methodological perspective, we present a link between the areas of econometrics and natural language processing. Beyond these findings, we show ways to construct daily time series of soft information from textual data that can be used as a supplement to the hard information of market data.

The remainder of this paper is organized as follows. Section 2 describes our contributions to several strands of the literature in empirically revealing mechanisms behind causes of the bursting of the Bitcoin bubble as a result of the news media. Section 3 provides a short overview on Bitcoin and on its main functions. Section 4 provides an overview of the methodology used. The source of data used and ways to extract information from the text are described in section 5. Section 6 describes the experimental design of our model. Section 7 presents our results on the identification of the effects of news media on the dynamics of Bitcoin prices. Final remarks are given in section 9.

2 Literature Review

Our work is related to several bodies of literature on finance, machine learning. This study is mainly related to the role of media and the implications of attention in the financial market. We study the impacts of attention to the news media on the Bitcoin market.

The effect of the news on financial markets such as the stock market, currency market, and equity market is well understood (Tetlock, 2007). The responses of markets can be diverse in correspondence with news sentiment (Boyd et al., 2005). The finance literature has primarily focused on implications of attention for stock prices and volume (see e.g., (Hirshleifer et al., 2009; Menzly and Ozbas, 2010; Cohen and Frazzini, 2008; Odean, 1999, 1998; Engelberg and Gao, 2011). This literature mostly focus on the attention of individual investors; Ben-Rephael et al. (2017) examine the implication of institutional attention. In our study, we extend this literature by analyzing the attention of investor with regarding the content of news articles and corresponding responses of the markets.

Our central contributions are related to the cryptocurrency market. Corbet et al. (2019) provide a systematic assessment of the empirical literature on this market and on the development of cryptocurrency as a financial asset. Bordo and Levin (2017) show that how digital currency is controlled by central banks can ensure the stability of financial systems and facilitate monetary policy. Similarly, Fung et al. (2018) concludes that digital currencies cannot be entirely safe without government or central bank intervention. In our study, we show how news on Bitcoin with a primary concept focusing on central banks can influence the attention of market participants.

Another strand of the literature focuses on the economic and monetary aspects of digital currencies. Böhme et al. (2015) studies the possibility of disrupting payment or monetary systems, Gandal et al. (2018) analyse price manipulation in the Bitcoin ecosystem, and a study by Huberman et al. (2017); Easley et al. (2019) focuses on transaction fees of Bitcoin and its infrastructure level in an equilibrium model. Monetary characteristics of digital currencies are scrutinized in studies like Weber (2016), who compares the fixed supply of digital currencies (Bitcoin standard) to the Gold Standard, and Schilling and Uhlig (2018), who propose a model of Bitcoin as a digital currency versus other traditional currencies. Auer and Claessens (2018) analyse the impact of regulating cryptocurrencies on valuations and trading volumes. An empirical study by Agarwal (2015) claims that the implementation of digital currencies can lead to the development of a negative interest rate policy. In our study, we observe the regulatory and monetary aspects of digital currency attracted the attention of investors, mainly after the bubble phase.

The blockchain is the innovative technological platform behind Bitcoin. The literature on the economics of blockchain technology remains less developed. Yermack (2017) explores the potential corporate governance implications of blockchain technology. Abadi and Brunnermeier (2018) compare distributed ledger technology (DLT) to a traditionally centralized intermediary and conclude that DLT cannot ensure the correctness, decentralization, or cost efficiency of a system at the same time. Cong and He (2019) consider how the decentralization and effectiveness of a blockchain can reshape business models and competition between industries. Chen et al. (2019) provide an evidence on value of FinTech innovation, notably blockchain innovations yield substantial value to innovators and a faster and more flexible settlement system (Chiu and Koeppl, 2019). With our research, we show that attention to the Bitcoin market at the start of the bubble phase corresponds to news articles discussing the blockchain.

We also contribute to the literature on the understanding of the underline mechanism of the digital asset bubble. According to rational bubble theory, as long as an agent believes about a bubble return, which leads to additional expected bubble returns, stay in the market. This associates with economically rational inattention behavior of the agent. Several theoretical models about the formation of bubbles and crashes have suggested in literature such as Giglio et al. (2016); Moinas and Pouget (2013); Lei et al. (2001). Rational bubble theory (Diba and Grossman, 1988) is an important conceptualization of bubbles and the building block of the empirical literature. [1] Our study empirically shows that Bitcoin market participants until receiving negative signals about the security of the digital currency stayed in the market. Our results are consistent with the finding of Foley et al. (2019), which indicate that the illegal share of bitcoin activity, roughly one-quarter, diminishes interest in Bitcoin.

[1] We can list some empirical studies of bubble in financial market as follows: (Brooks and Katsaris, 2003; Charemza and Deadman, 1995; Assaf, 2018; Drescher and Herz, 2016; Chang et al., 2016; Engsted, 2016).

Our identification methodology is related to a growing literature in the social sciences examining the exogenous effects on financial markets. We use RD design to measure the causal effects of media news on the Bitcoin market. The RD design methodology focuses exclusively on informational flows around defined cut-off points and helps to identify effects locally. We use the robust R package developed by Calonico et al. (2014), who introduces robust data-driven statistical inference in RD designs. Calonico et al. (2014) propose the use of robust bias-corrected confidence intervals by employing data-driven optimal bandwidths. In the present work, we obtain optimal bandwidths based on the Calonico et al. (2015) procedure, though in the robustness section, we select a set of bandwidths to examine the robustness of our results. For the visualization of discontinuity in the model, we use the procedure described in Calonico et al. (2015). Calonico et al. (2018) analyse the impact of bias correction on confidence intervals in terms of selecting kernel densities and the degree of local polynomial regression. We report the results of the bias-correction method and the conventional RD design. In the robustness section, we study the sensitivity of our results to kernel densities and the degree of local polynomial regression observed. We also apply a local randomization method (Cattaneo et al., 2015) as a compliment and a robustness check of conventional RD inference methodologies.

We also make a methodological contribution to the literature on machine learning related to economics. In recent years, research on the application of machine learning in economics has become very popular (Athey, 2018). There has also been a new tendency to use applications of big data in economics (Athey, 2017). Gentzkow et al. (2017) provide a survey of these studies. We apply the natural language process (NLP) method to extract information to create explanatory variables from terms and major topics similar to Manela and Moreira (2017). In the present paper, we apply the LDA method to high-frequency textual data and extract relevant information to Bitcoin price. Another similar research to our paper is a study by Thorsrud (2018), which constructs a daily business cycle index on GDP growth from textual information of business newspaper. He uses the LDA method to decompose contents of news articles into topics and predicts future revisions of GDP in a dynamic factor model. This technique also is applied by Hansen et al. (2017) to transcripts of Federal Open Market Committee (FOMC)'s meetings to identify the transparency of central bank design and the deliberation of monetary policy. The LDA (Blei et al., 2003) method is a machine learning algorithm for probabilistic topic modeling. The LDA method uses unlabeled text data and finds links between terms, documents, and topics without prior information about the economic meaning of words. The conventional methods in economic studies use given particular word lists to identify the links between the above elements. Also, we can mention some other literature which applies LDA in economic studies such as (Budak et al., 2014; Nimark et al., 2016; Bandiera et al., 2017; Mueller and Rauh, 2018). Mostly, these studies employ this method to extract topics as significant concepts of economic text. The present paper goes one step further and constructs daily soft information index from the results of this method to compliment the hard information on market price.

3 Background Information

In this section, we review the history and fundamental functions of Bitcoin. Nakamoto (2008) first introduced Bitcoin in 2008. When Bitcoin could first be used to trade a product in 2009, 10.000 Bitcoins could be used to purchase two pizzas. Up until 2012, Bitcoin was traded at a rather low rate. Interest in Bitcoin grew when the digital currency started to be traded on an online exchange platform called Mt.Gox. During the Cyprus banking crisis at the end of 2013, the price increased quickly to $200, and many investors tried to enter the market, as the government does not control this market. In the following years, Bitcoin attracted more attention from major players within the market and creates large, recurrent arbitrage opportunities across exchanges, and within countries (Makarov and Schoar, 2019). There was significant and increasing arbitrage possibilities during bubble time between global Bitcoin markets. Fuelled by high levels of media attention, the price of Bitcoin increased immensely and reached a peak of nearly $20,000 in mid-December 2017. Since then, the rate has declined continuously.

Crypto-currency may be the most influential invention to revolutionize the financial system after the invention of the Internet. It supports inexpensive and fast transactions without involving a third-party intermediary or the control of governments or central banks. The technology behind Bitcoin solved a double-spending problem and was the first digital currency that could be spent multiple times, and therefore, an intermediary was always required. Bitcoin uses a peer-to-peer distributed time stamp server to produce computational evidence of the sequential order of transactions (Nakamoto, 2008). Once a transaction is confirmed, it sits in a block with other transactions, and this block will add to the group of blocks known as the blockchain (Böhme et al., 2015). The blockchain covers all transactions made in sequential order. The blockchain can be fundamentally viewed as a distributed ledger. This means that every participant of this market can update the ledger. An up-to-date copy of the entire ledger is stored by each participant (BIS, 2018). A new transaction is examined in contradiction with transactions listed in the blockchain to guarantee that a similar Bitcoin has not previously been spent. In the blockchain, each transaction is verified via asymmetric encryption. This form of encryption uses two types of keys for a single Bitcoin: a public key and a private key. The public key is available to the public and is used as the basis for a Bitcoin address. On the condition that a public key cannot be associated with the owner's identity, the Bitcoin remains unidentified. Nevertheless, it is not difficult to create a link between a public key and a person.

Bitcoin network is needed to take care of authenticating and recording each transaction. This can be done by a small group of participant miners. A miner confirms transactions and generates new blocks with solving complex mathematical problems and (s)he can earn a new mined Bitcoin and a transaction fee. Then, the miner needs to publish the new block, and it required to be verified (”proof-of-work”). After the ”proof-of-work” is published, other participants should confirm it, as a result the new block can be added to the blockchain system.

There are different understandings about Bitcoin concerning its functions. Commonly, money has three primary functions (BIS, 2018). First, money is a unit of account, and it means that it is used to measure the value of goods or services. Second, money can be a medium of exchange, and it means that a vendor can accept it as a means of expense.

Third, money serves as a store of value, and it means that it enables market participants to transfer acquiring authority of goods or services over time. Bitcoin can be considered as money with a limitations of supply elasticity and not controlled or regulated by any authorities such as central banks or governmental institutions. The last functionality is essential since for a currency to be served as a medium of exchange or a unit of account. However, it first requires to be acknowledged as a store of value (Ali et al., 2014). As overall, Bitcoin, as a crypto-asset, has a bright feature of the continued growth; however, it does not presumably provide the conventional functions of money.

4 Methodology

This section describes advanced machine learning and econometrics methods used to extract information and analyse the effects of news media on Bitcoin pricing. As a first step, an automated processing method extracted information from news articles. Second, the LDA method was used to identify topics from the extracted data. Finally, daily entropies of topics' informational flows in an RD framework were used to determine the reasons behind the growth and bursting of the Bitcoin bubble.

4.1 Text-mining

Text-mining is a general automated process used to extract information from unstructured data. The first and fundamental stage of text mining is that of data preparation. Textual data are created in a format that is understandable to humans; these data must be processed to be understandable to a machine. Text of an unstructured format must be tokenized into smaller, more precise features such as words or groups of words known as tokens. Tokenization involves generating a list of elements of the given text as a character vector. After this stage, the extracted data are still not uniform. Some techniques such as lemmatization and stemming processes must be employed to transform data into base formats. The data may still include some unwanted information, which can create noise in the final results. Therefore, data should be cleaned by extracting common and regular words known as ”stop words”. What results is a text corpus, which is a group of texts with each text including a composition of basic vocabularies of ”words” or ”terms”. In the last stage of data preparation, the text corpus in ”bag-of-words” format must be transferred to the matrix space. The Document-Term Matrix (DTM) is one of the most common presentations of text data in matrix space. The matrix can be presented in different schemes like Term Frequency-Inverse Document Frequency (TF-IDF) schemes (Salton and McGill, 1986) to adjust the frequencies of documents and terms in the matrix. After applying these processes, data are of morphological root form and are ready for subsequent use.

4.1.1 Latent Dirichlet Allocation (LDA)

The main idea expressed in a text document is called a topic. A document covers multiple topics where each topic represents a collection of words and terms. A topic can be subjectively interpreted with given information about its terms. Generally speaking, a document is not labelled with topics. To identify the main idea expressed in a document, probabilistic topic modelling methods can be employed. Such methods use an automatic procedure handled by an algorithm that includes no information on the documents' subjects. Topics are determined by computing the hidden structure of a document. LDA is one of the probabilistic topic models that can be applied to a large number of documents. The intuition and generative processes of LDA can be described as follows. For a given set of documents, a number of topics are distributed across terms and words, and topic modelling methods are designed to determine the characteristics of this distribution automatically. The main computational issue here is to infer the latent topic structure from observed documents and to estimate per-document-topic and per-term-topic links.

LDA is a probabilistic model for text classification and for the topic modelling of series of discrete data known as text corpora. The data can be of two formats: structured or unstructured. LDA is a hierarchical Bayesian model. In such a model, each item of a data series is modelled as a finite mixture over a primary set of topics across multiple levels. In this context, the model provides probabilities of each document and each term with respect to each topic. This section briefly presents an efficient inference technique that can be used to approximate the parameters of the model. The expectation maximization (EM) algorithm empirically estimates Bayes parameters.

In what follows, we describe a simple presentation of the LDA model with K topics, D documents, and N terms in a document. Topics, terms and documents are distributed in a hierarchical format: topics are distributed over documents; terms are distributed over topics, and terms are distributed in documents. The generative process of the LDA model with hidden and observed variables can be illustrated in the following joint distribution function:

$$P(\varphi, \theta, z, w) = P(\theta)P(\varphi)\prod_{d=1}^{D} P(z|\theta)\prod_{n=1}^{N} P(w|\varphi, z),$$

where $\theta_{d(1:D)}$ is the topic distribution of document d, $\varphi_{k(1:K)}$ is the term distribution of topic k, $z_{dn(1:D,1:N)}$ is the topic distribution of the term n in document d, and $w_{dn(1:D,1:N)}$ is the distribution of term n in document d. In a Bayesian statistics framework, the posterior probability of parameter θ is expressed as:

$$P(\theta|\alpha) = \frac{1}{B(\alpha)}\prod_{K} \theta_K^{\alpha_{K-1}}, \tag{1}$$

with given condition $\sum_k \theta_{kD} = 1$, and parameter α is the Dirichlet prior distribution on document-topic links ($\alpha < 1$). Similarly, the posterior probability of parameter φ is expressed as:

$$P(\varphi|\beta) = \frac{1}{B(\beta)}\prod_{K} \theta_K^{\beta_{K-1}}, \tag{2}$$

with given condition $\sum_n \varphi_{kn} = 1$, and parameter β is the Dirichlet prior distribution on the topic-word links.

The hierarchical estimation procedure starts by choosing parameters α and β, then by selecting a topic with distribution $z_{i,j} \sim \text{Multinomial}(\theta_i)$, finally by selecting a term with distribution $w_{i,j} \sim \text{Multinomial}(\varphi_{z_{i,j}})$.

4.2 Measure of Information of Attention

The measuring information of attention to a news' topic can be estimated by how much one can learn, on average, about a topic in a news document (Glasserman and Mamaysky, 2017). Equivalently, it can be seen as the amount of the reducing of uncertainty, on average, about a topic in a news text. The information of attention to a news' topic can be measured based on Shannon's entropy (Shannon, 1948). In doing so, we measure the entropy of the estimator of the link between topic k and document d with parameter θ. It can be expressed as follows:

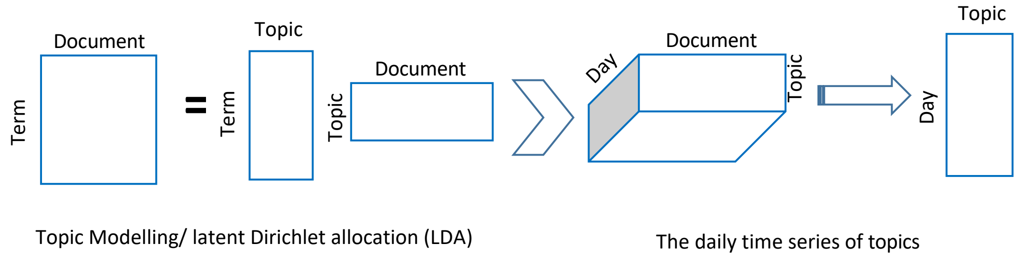
$$H(\theta_{kd}) = -P(\theta_{kd}) \cdot log^{-1}(P(\theta_{kd})). \tag{3}$$

The aggregated daily value of intra-day entropy $H(\theta_{kd})$, in average, is given by:

$$\bar{H}(\theta_{kt}) = \frac{\sum_{i=1}^{N_t} H(\theta_{ki})}{N_t)}, \tag{4}$$

where the number of documents in day t denotes by N_t. In order to have a robust estimation with avoiding the influence of outliers, the aggregated daily value of entropy can be estimated by median of intra-day values. As we focus on the extreme events as well as volatile times of the market, we use average estimations. Figure 1 illustrates the procedure of decomposing the news document into topics and creating daily time series of information entropy of topics.

Figure 1: The Procedure of Creating Time Series of Topics


Document
Term
=
Topic
Term
Document
Topic
Day
Document
Topic
Topic
Day
Topic Modelling/ latent Dirichlet allocation (LDA)
The daily time series of topics

This figure presents the procedure of creating a daily time series of topics. The LDA method uses the Term-Document matrix and extracts the given number of topics. The later step aggregates the entropy of topics on a daily base and creates a time series of topics.

4.3 Discontinuity Regression Design

The Regression Discontinuity (RD) design can be used to estimate a known discontinuity to determine a local treatment effect. The method was first proposed by Thistlethwaite and Campbell (1960) and is used in the econometrics literature following Hahn et al. (2001). This method is a matching approach that pairs objects with the same visible features where one object receives treatment whereas another object does not. The crucial assumption of this method is that the allocation of treatment only depends on observable features. Hence, the untreated object can be a ”missing counterfactual” for the object receiving the treatment.

In what follows, we explain a simple model of RD design with single measured feature, (known as running or forcing variable) x, and threshold score-value $\widetilde{x}$ at the point which the allocation of treatment is disconnected. We consider the following simple model for the conditional dependent variable y:

$$E\left[y|x, D\right] = a_0 + Df\left(x\right), \tag{5}$$

where function $f\left(\cdot\right)$ is a continuous function and variable D is a dummy variable.

In this formulation, the effect of the treatment of dependent variable y is captured by function $f\left(\cdot\right)$ treated by the treatment. This function is a smooth continuous function in the whole area except at around the cut-off point. The regression model can be expressed as follows:

$$y = a_0 + Df(x) + u, \tag{6}$$

with the condition $E(u|x) = 0$. The RD design model focuses the effect running or forcing variable x on the dependent variable y around cut-off points. We can express these effects as follows:

$$y^- \equiv \lim_{x \to \widetilde{x}^-} E\left[y|x\right] \quad \text{and} \quad y^+ \equiv \lim_{x \to \widetilde{x}^+} E\left[y|x\right]. \tag{7}$$

The variable ε measures the difference between the values of dependent variable in two sides of the cut-off point:

$$\Delta\left(\varepsilon\right) = E\left[y|x = \widetilde{x}^-\right] - E\left[y|x = \widetilde{x}^+\right]. \tag{8}$$

Substituting for y from 6, we obtain

$$\Delta\left(\varepsilon\right) = E\left[a_0 + Df\left(x\right) + u|x = \widetilde{x}^-\right] - E\left[a_0 + Df\left(x\right) + u|x = \widetilde{x}^+\right]. \tag{9}$$

Using that the disturbance has zero conditional mean for all x, we then obtain that

$$\Delta\left(\varepsilon\right)=f\left(\widetilde{x}^{-}\right)E\left[D|x=\widetilde{x}^{-}\right]+f\left(\widetilde{x}^{-}\right)E\left[D|x=\widetilde{x}^{+}\right]. \tag{10}$$

Variable D is binary, therefore, we can define $E\left[D|x\right]=p\left(x\right)$. It gives us:

$$\Delta\left(\varepsilon\right)=f\left(\widetilde{x}^{-}\right)p\left(\widetilde{x}^{-}\right)+f\left(\widetilde{x}^{-}\right)p\left(\widetilde{x}^{+}\right). \tag{11}$$

In order to apply the discontinuity in the probability of estimation of treatment effects, we can distinguish between probability of both sides of the cut-off point and define p^{-} and p^{+} as the following:

$$p^{-}\equiv\lim_{x\to\widetilde{x}^{-}}p\left(x\right),$$
$$p^{+}\equiv\lim_{x\to\widetilde{x}^{+}}p\left(x\right).$$

The condition $p^{-}\neq p^{+}$ can be satisfied since we assume a discontinuity at $x=\widetilde{x}$. In order to compare treatment effects the below and above of the cut-off point, we take the limit of $\Delta\left(\varepsilon\right)$, when ε goes to zero. Hence, we have $y^{-}-y^{+}=\lim_{\varepsilon\to0}\Delta\left(\varepsilon\right)$ which can be expressed as:

$$\begin{aligned}\lim_{\varepsilon\to0}\Delta\left(\varepsilon\right) &= \lim_{\varepsilon\to0}f\left(\widetilde{x}^{-}\right)p\left(\widetilde{x}^{-}\right)+f\left(\widetilde{x}^{-}\right)p\left(\widetilde{x}^{+}\right) \quad (12)\\ &= f\left(\widetilde{x}\right)\left(p^{-}-p^{+}\right). \quad (13)\end{aligned}$$

It leads to equation:

$$f\left(\widetilde{x}\right)=\frac{y^{-}-y^{+}}{p^{-}-p^{+}}. \tag{14}$$

The main idea of RD design is that the allocation of treatment depends on the forcing variable x which is cut off at $\widetilde{x}$. However, the assignment of treatment to all objects is not entirely sure. In this case, the difference of two probabilities is not equal one ($p^{-}-p^{+}\neq1$), and we can use fuzzy design. In a particular case, when all objects from one side of cut-off point received the treatment, and ones from others did not, we can apply sharp RD design. In the sharp design, $p\left(x\right)=1$ for all $x\leq\widetilde{x}$ and otherwise 0, it means that $p^{-}=1$ and $p^{+}=0$ or $p^{-}-p^{+}=1$.

5 Data

For this study, we examine news articles of an unstructured format. We construct our text-based variables from news documents about Bitcoin. We collect news items from the Lexis-Nexis database. The database lists the content of news documents and important information about each document. News documents are of different structures and formats and have been published by various publishers. Online news data are text data drawn from online sources such as web publications and news wires. Print news data include news articles published by newspapers, journals and magazines. We apply several strategies to select relevant news data from a large number of documents related to Bitcoin. We select news items first when the title of a news article contains the word ”Bitcoin”, second when this word appears several times in news text, and third when digital currency is cited as a key topic of a document under the Lexis-Nexis database to cover all selected news on Bitcoin. In the next step, we exclude all news articles about Bitcoin prices. This step is import to design unbiased casual inference estimation of relationships of Bitcoin price and news media. Those news items might have a clear message about Bitcoin's price can be influenced by market trends.

Our data sample includes 244,703 online news articles and covers the period running from the start of January 2014 to December 2018. The news dataset includes 11,650 web publications, 22,679 news wires, and 208,217 newsletters, and the rest are web blogs and reports. Price and supply (sent from addresses) information on Bitcoin for the same period are employed as dependent variables. Table 1 presents descriptive statistics on the variables. Figure 2 presents a time series on the supply & price of Bitcoin and number of related news articles. These figures show that after reaching a peak at the end of 2017, the number of news articles published on Bitcoin increased exponentially.

Extracting Information We construct our textual analysis-based variables from news documents about Bitcoin. In doing so, we extract leading sentences and paragraphs containing the word ”bitcoin” in the news text. We then apply natural language processing techniques described in the methodology section to clean and extract relevant terms. The

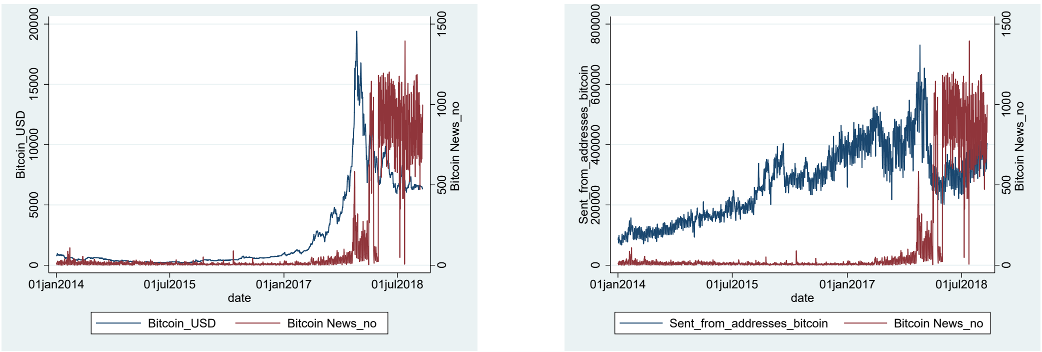
Lexis-Nexis database uses an automatic mechanism to extract important information from news texts and to assign news to different categories. For each news document, information is provided under the following headings: city, company, country, geographic, industry, organization, person, section, state, and subject terms. We use available information on news items including the outlined subject terms and keywords. We apply the same procedure to clean the data and to extract relevant terms. We combine terms taken from key news content with terms taken from leading sentences to generate a rich document-term matrix. We then apply the LDA method to extract topics from this matrix.

Table 1: Discriptive Statistics of Data Sample

Variable	**Mean**	**Sd**	**N**	**Min**	**P25**	**P50**	**P75**	**Max**
Bitcoin Price in USD	2591.806	3605.420	1689	197.577	400.002	632.924	4156	19401
Number of News	140.785	306.709	1689	1	6	13	34	1395
Sent from addresses	2.78e+05	1.22e+05	1689	67236	1.65e+05	2.89e+05	3.67e+05	7.30e+05

This table reports summary statistics information of data used in our study. Our dataset includes price information of Bitcoin in USD, the number of news articles about Bitcoin, number of Bitcoin sent from addresses.

Figure 2: **Time Series of Supply & Price of Bitcoin and Number of News Articles**



(a) Number of News Articles and Price of Bitcoin

(b) Number of News Articles and Supply of Bitcoin

These figures represent the time series of supply & price of Bitcoin and number of related news articles. The period of the study starts from the begging of the 2014 until the end of the 2018. This period includes several volatile phases. To illustrate better relations between news and price and supply of Bitcoin, we use separated figures.

6 Empirical Design

In this section, we discuss how we use the regression discontinuity design to estimate causal effects of the media on Bitcoin prices. We apply RD designs to identify a local average treatment effect (LATE) according to shifts in the mean value of the dependent variable as well as changes in the likelihood of treatment indicator D, once running variables cross the known cut-off points.

We design several experiments to analyse the effects of attention to Bitcoin in the news media. The running variables are taken from textual data and are exogenous. We use two days of lagged information on the entropy of news topics as running variables and the price and number of messages sent on Bitcoin as dependent variables. We use treatment indicator D based on cut-off points in each experiment. Cut-off points are determined independently of the running or forcing variables.

In the first experiment, we compare prices which are before bubble phase with those after that time. We use data in a two-day lag. The topics of arriving news are very similar in all subjects except that the two days before bubble crashing did not receive a particular news topic while the after that time did. Bitcoin price hit an all-time high just below \$20,000 at December 18, 2017, but stopped short of \$20,000. Therefore, we specify the bubble cut-off point at this time. In robustness analysis, we shift the cut-off point to check the robustness of the result related to designing the experiment.

In the second experiment, we analyze news information which attracted the attention of the market and built confidence for investors to enter this market. Bitcoin price reached \$1,000 for the first time on January 3, 2017. At that time, mass media reported the new Bitcoin market participants, and it was anticipated the raising of the Bitcoin price even higher.

We also consider the period after the bubble crashing time in December 2017. In the third experiment, we define the cut-off point when the Bitcoin price dropped to below \$6000. We should emphasis that all cut-off points are defined on the exogenous event which independent of news topics. Almost, all other cryptocurrencies such as Ethereum (ETH), Bitcoin Cash (BCH), and Litecoin (LTC) are roughly tracking Bitcoin's price dynamics. In this research, we focus on the Bitcoin price dynamics; however, all other cryptocurrencies had the same behavior pattern during the study period.

Throughout the paper, we use nonparametric RD design models. A parametric approach is a global estimation strategy which employs all observation in the sample to, even observations far from the cut-off point, to estimate the average outcome around the cut-off point. In contrast, the nonparametric is a local strategy which restricts the analysis to observations that located near to the cut-off point. The parametric approach can be considered as applying the precise model for estimation with given observations, whereas the nonparametric approach is a method which uses accurate observations for estimation with a given model. We focus on the changing of the patterns of news articles arriving about Bitcoin at a specific time. We, therefore, apply local polynomial nonparametric estimators with data-driven bandwidth. The nonparametric approach uses a regression function in an area of the cut-off points, and it requires choosing the bandwidth. As we are using a large data set, it is necessary to improve the power of the nonparametric method to detect effects with limiting the bandwidth.

We apply both fuzzy and sharp RD designs. The central assumption of the sharp design is that all objects receive treatment, however, in the fuzzy RD design, this assumption does not hold. In our quasi-natural experiments, it is possible to assume that some Bitcoin investors did not pay attention to the media or a specific topic did not attract their attention. Based on this fact, the results of the fuzzy model can be valid. Another advantage of fuzzy RD design is that this model does not require including control variables into model and create a more precise estimation (Imbens and Lemieux, 2008).

We implement the fuzzy RD approach in a two-stage regression framework. In the first stage, we estimate the discontinuity in the probability of having structure break in the Bitcoin price caused by attention to news media (entropy or unusualness). In the second stage, we estimate a similar relationship between Bitcoin price and attention to the news media. The fuzzy RD design is carried out in a two-stage least squares (2SlS) framework: First stage equation:

$$D = a_1 + f_1(x) + u_1, \tag{15}$$

Second Stage equation:

$$y = a_2 + \beta D + f - 2(x) + u_2, \tag{16}$$

where variable y is the dependent variable, variable x is running variable, dummy variable D is an indicator of whether receiving the treatment, functions f_1 and f_2 are continuous functions.

We use the conventional RD model as the central point estimator as well as robust and bias-corrected inference procedure proposed by Calonico et al. (2014) for inference and the valid confidence intervals. The robust model is based on different data-driven bandwidth built on the previous model by Imbens and Kalyanaraman (2012) and it is robust to ”large” bandwidth choices.

6.1 Validation

The RD design technique is a quasi-experimental method which can identify causal relationships with high internal validity. The estimation of this method is not biased by endogeneity issues like omitting variables on the condition that forcing variables are continuous around the cut-off point. The key point to our experimental design is to verify the smoothness in decomposed news topics across the three defined cut-off points on the Bitcoin price. We employ the formal validity tests which are instructive to check that entropies of flow of news are continuous across the cut-off points. To verify the requirement of no manipulation, we apply the formal test proposed by McCrary (2008). The test examines a discontinuity in the density of the forcing or running variable around the cut-off point. If we cannot reject the null hypothesis that the density of the running variable is smooth, the RD design with using the forcing variable can produce consistent estimation. These tests can approve validity which there is local randomization to perform a quasi-experiment using sharp and fuzzy regression discontinuity designs around the cut-off points. With satisfying this condition, fuzzy RD design does not require the addition control variables (Imbens and Lemieux, 2008).

Endogeneity (Simultaneity) Problem We address another type of endogeneity (simultaneity) problem related to the joint determination of news articles and Bitcoin prices. Endogeneity bias represents a crucial issue for the investigation of cause and effect relationships with affecting the causal inferences of the hypothesized associations between dependent and independent variables. Since the media can, at the same time, react to changes in Bitcoin prices, the forcing variable might be endogenous. It would cause bias when determining the price relationship econometrically. A standard solution to this problem is replacing a suspected endogenous forcing variable with its lagged information. This solution might not reduce bias in the presence of mutual dependency and autocorrelation of variables. Some approaches, such as instrumental variables and instrument-free techniques, have been employed to deal with the endogeneity problem. In our study, we exclude all news articles about Bitcoin prices. A news item can be assigned to several topics. Those news items might have explicit information about Bitcoin price can be influenced by market trend.

Visualization Another way to examine the internal validity of RD design is plotting the density of the running variables at each data point. If there is no discontinuity detected around the cut-point (no manipulation around the cut-off point), then RD design is valid. In some cases, the discontinuity may not be as certainly determined through visual assessment. Therefore, this approach is supplementary to the formal statistical test. Additionally, a graph provides some insights about the relationship between the dependent and forcing variables and how heterogeneous the data around the cut-off point is likely to be. Therefore, visualization techniques help to choose correct specifications.

6.2 Robust Estimation

There are some issues about RD design which can reduce the validity and acceptability of the results. These issues correspond to misspecification of functional form, choosing a correct kernel function and bandwidth parameter in nonparametric settings, and misspecification of cut-off points.

Alternative Specifications: In order to examine the robustness of our results, we use graphical analyses to check the specification of functional form. We repeat our experiments with alternative specifications like different bandwidths. We shift the cut-off points to the neighbour of the cut-off points to create tie-breaker experiments and check the validation of the results.

Randomization: Generally, RD design can be characterized in two different ways: first, it is conducted as ”discontinuity at a cut-off point,” as our main analysis and, second ”local randomization” (Lee and Card, 2008). The second way is based on differences between objects which miss treatment and those who received treatment are random. It means that any difference, on average, between values of dependent variables of both sides of the cut-off point should, therefore, be caused by treatment.

7 Empirical Results

This section presents the empirical results of our study. The results of the LDA model and our interpretations of the topics explored are described in the first subsection. In the following subsections, the daily entropy values of topics are used as running variables in RD design models to show discontinuity in Bitcoin price dynamics.

7.1 LDA Results

In this section, the results of applying the LDA model to the news data are presented. In an LDA analysis, several parameters should be determined including the number of iterations and the number of Markov chains. These parameters must be carefully chosen, and it is necessary to check the robustness of the results with regard to these parameters. In our experience, results are not highly sensitive to the above parameters. Importantly, the number of topics must also be specified. There are different ways to identify the optimum number of topics. Some automatic procedures have been proposed in the text mining literature (e.g., (Teh et al., 2005; Cao et al., 2009)). In this study, we empirically measure the optimal number of topics to explore. We apply the LDA model to a wide range of topics and test which LDA model serves as a good model for news documents. We find seven to be the optimal number of topics.

We next interpret topics associated with terms and the probability of terms assigned to topics. A single document can include information about multiple topics, but it can be more closely related to only a few topics. A topic is a collection of terms related to a subject. A single term can have a different meaning in a different text; however, we examine a collection of terms that can convey a single message. As described in the methodology section, the LDA-based document model provides information on the probability of a term being assigned to a single topic and of a topic being assigned to a single document. This information helps one interpret topics. Chang et al. (2009) propose two ways to measure the interpretation of topics. The first approach is that of ”word intrusion”, which involves forming a collection of top five terms with five terms less likely to be used to explore a topic and then identifying inserted terms. The second approach is that of “topic intrusion”, which involves applying a set of closely related topics to a single document and combining this set with a set of random topics to identify relevant topics. To interpret the results of the LDA model, we first select the optimal number of interpretable topics due to a trade-off between the goodness of fit of a model and its interpretability. We then use probability information for the LDA model and apply both approaches described in Chang et al. (2009). Table 2 presents the results of the LDA model. The table shows the top 10 topics with the highest values of parameter φ. To improve the readability of terms, we complete the stemmed words. Figure 3 shows the terms' distributions across topics. We measure attention to each topic based on entropy information on the topics. We then aggregate the entropy information to create daily entropy information. Table 3 presents descriptive statistics on the daily entropy information. In what follows, we interpret each topic based on information on each term's distribution of topics.

Topic 1: Fintech & Blockchain The topic 1 includes terms with regard to financial technology, digital assets, as well as blockchain. The value of parameter φ of the distribution function of terms in topics (Equation 2) is high for some terms such as financial technology, the blockchain. We, therefore, call this topic "Fintech & Blockchain". This topic also talks about the dynamics of Bitcoin price.

Topic 2: Stock Exchange Topic 2 mostly includes information on stock exchange market, transactions about the ownership of companies, and formal agreements. Based on this information, we call this topic "Stock Exchange".

Topic 3: Digital Banking & Finance This topic represents news articles mainly contain information on electronic banking systems such as credit card, electronic commerce, and electronic wallets. However, news articles about venture capital and entrepreneurship are classified in this topic. Since the majority of information is related to the electronic version of the Banking system, we call this topic "Digital Banking & Finance".

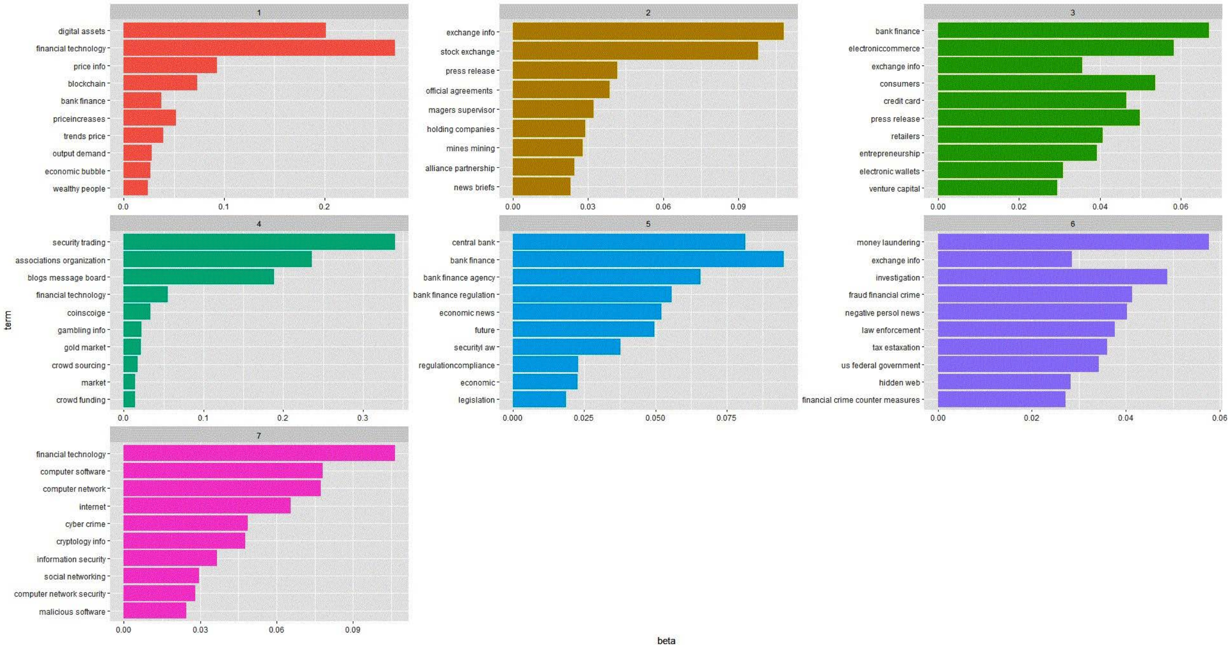
Topic 4: Security Trading This topic includes information on trading financial asset securities in decentralized platforms. Some news articles about different platforms which Bitcoin could be traded. A general platform like crowd-sourcing platform or a specific platform like ComScore can trade in a decentralized system. These platforms implement new technologies from FinTech such as Blockchain. We use a general expression for this topic, and therefore, we call it "Security Trading".

Topic 5: Central Bank & Regulation This topic shows the link between central-banks, regulation, and legislation systems. Recently, central-banks concern about the regulation of digital currency. This topic is mostly about "Central Bank & Regulation".

Topic 6: Law & Tax Investigation This topic assigns to news items which contains negative information about Bitcoin such as money laundering, investigation, fraud, and financial crime. Information about Taxes and taxation are also included in this topic.

Topic 7: Computer & Fintech This topic includes terms related to Fintech and software technology behind Bitcoin. The news items cover technological aspect of Bitcoin like cryptology are assigned to this topic.

Figure 3: The Terms' Distributions in Topics from the LDA Model



This table represents terms' distributions in topics from the LDA Model. Each bar shows the value of parameter φ of the distribution function of terms in topics (Equation 2). The names of topics are shown in Table 2.

In the following subsections, we present the results of the RD design model to determine the impact of news media on Bitcoin prices. We measure this impact from the amount of information contained in news articles to limit uncertainty surrounding different aspects of Bitcoin and to attract the attention of market participants.

7.2 Bubble Phase

At the end of 2017, the price of Bitcoin increased immensely and reached its highest value at close to $20,000 in mid-December 2017, and after a short period, the price crashed. Before this point, the amount of information released by the news media had been increasing exponentially. According to market efficiency theory, price movements are related to changes in investors' beliefs about the fundamental value. With our regression discontinuity design,

Table 2: Terms-Topics in LDA Classification

Topic 1: Fintech & Blockchain		Topic 2: Stock Exchange		Topic 3: Digital Banking-Finance		Topic 4: Security Trading	
Term	φ	Term	φ	Term	φ	Term	φ
financial technology	0.27	exchange info	0.108	bank finance	0.067	security trading	0.34
digital assets	0.201	stock exchange	0.07	electroniccommerce	0.058	associations organization	0.236
decentralize	0.093	press release	0.042	consumers	0.054	blogs message board	0.189
blockchain	0.073	official agreements	0.038	press release	0.05	financial technology	0.055
token-based	0.052	magers supervisor	0.032	credit card	0.047	coinscoige	0.033
market trend	0.039	holding companies	0.029	retailers	0.041	gambling info	0.022
bank finance	0.037	mines mining	0.028	entrepreneurship	0.039	gold market	0.022
output demand	0.028	stock exchange	0.028	exchange info	0.036	crowd sourcing	0.017
economic bubble	0.027	alliance partnership	0.024	electronic wallets	0.031	market	0.015
wealthy people	0.024	news briefs	0.023	venture capital	0.029	crowd funding	0.014
interviews	0.013	share holders	0.022	mobile application	0.028	gambling info	0.012
market changes	0.012	company profits	0.02	mobile application	0.023	market open close	0.011
industrial analysts	0.011	stockprice	0.02	internet ralted	0.022	time information	0.008
hedge fund	0.008	mines mining	0.018	electronic billing	0.021	government info	0.006
market size	0.008	company strategy	0.016	new products	0.019	voter voting	0.005
online trading	0.008	mine operations	0.016	automated teller machines	0.017	gambling info	0.003
intertiol trade	0.007	board directors	0.014	illegal drugs	0.017	productenhancements	0.003
market research	0.007	company earnings	0.014	electronic bank	0.017	entertainments	0.002
market researchalysis	0.007	computer trading systems	0.013	conference convention	0.016	bbelectroniccommerce	0.001
research report	0.007	mine planning magement	0.013	debit card	0.015	business professiol associations	0.001

Topic 5: Central Bank & Regulation		Topic 6: Law & Tax Investigation		Topic 7: Computer Fintech	
Term	φ	Term	φ	Term	φ
bank finance	0.095	money laundering	0.058	financial technology	0.106
central bank	0.081	investigation	0.049	computer software	0.078
bank finance agency	0.066	fraud financial crime	0.041	computer network	0.077
bank finance regulation	0.056	negative persol news	0.04	internet	0.066
economic news	0.052	law enforcement	0.038	cyber crime	0.049
future	0.05	taxes taxation	0.036	cryptology info	0.048
security law	0.038	us federal government	0.034	information security	0.037
economic	0.023	exchange info	0.029	social networking	0.03
regulation compliance	0.023	hidden web	0.028	computer network security	0.028
legislation	0.019	financial crime counter measures	0.027	computer crime	0.024
talk meeting	0.018	litigation	0.027	malicious software	0.024
derivative instruments	0.017	special investigative forces	0.024	online security privacy	0.02
european union	0.017	arrests	0.023	digital signatures	0.019
interest rate	0.017	tax law	0.021	computer equipment	0.018
deflation	0.017	cyber crime	0.019	computer programming	0.016
public policy	0.017	law court tribuls	0.019	web programming	0.016
government info	0.016	lawyers	0.019	artificial intelligence	0.014
risk magement	0.015	terrorism	0.019	computing information technology	0.013
official approvals	0.014	controlled substance crime	0.018	network servers	0.013
us president candidate	0.014	criminal investigation	0.018	cloud computing	0.012

This table shows the results of the topic modeling method. Terms are assigned to each topic. The table shows top 20 terms, with highest probability, for each topic. In the data cleaning process, some techniques such as lemmatization and stemming processes are employed to transform words into base formats. In the term-document matrix, 2-gram words are combined with 1-gram words. In order to make the terms more readable, we complete terms. For instance, the stemmed words ”comput” has been changed to the word ”computer”. The φ presents the parameter value of the distribution function of terms in topics (Equation 2)

Table 3: Descriptive Statistics of Daily Entropy of News Topics

Entropy Variable	mean	sd	N	min	p25	p50	p75	max
Fintech & Blockchain	0.271	0.009	1689	0.227	0.265	0.271	0.277	0.314
Stock Exchange	0.276	0.010	1689	0.235	0.271	0.276	0.281	0.334
Banking & Finance Electronic	0.282	0.012	1689	0.235	0.275	0.279	0.289	0.351
Security Trading	0.267	0.010	1689	0.227	0.261	0.265	0.272	0.293
Central Bank & Regulation	0.276	0.011	1689	0.227	0.270	0.275	0.281	0.345
Law & Tax Investigation	0.278	0.015	1689	0.235	0.270	0.275	0.283	0.366
Computer & Fintech	0.278	0.011	1689	0.235	0.272	0.276	0.282	0.361

This table represents the descriptive statistics of the daily entropy of topics. The table provides simple summaries about the sample and the measures of minimum, maximum, standard division and percentiles. The entropy has monotonicity property, therefore, it is robust to outliers.

we can identify the effects of media attention on the Bitcoin market.

We first run a manipulation test to check the internal validity of the RD design. We employ the test proposed by McCrary (2008), which examines discontinuity in the density of the running variable around cut-off points. The results are presented in Table 4. The table shows P-values for robust and conventional RD design models. From this table, we can conclude that there is no statistical evidence of systematic manipulation among the running variables “Central bank & regulation”, “Law & tax investigation”, “Security trading” and “Stock exchange” from the robust model estimation. However, in the conventional model, suitable running variables for the RD design model include “Bank & finance electronic”, “Fintech & blockchain”, “Central bank & regulation” and “Security trading”. Figure 8 plots the density of the above running variables at each data point. The figure shows an absence of discontinuity around the cut-point (no manipulation around the cut-off point) for running variables “Bank finance electronic”, “Fintech & blockchain”, “Central bank & regulation”, and “Stock exchange”. The manipulation results are mixed, therefore, for inference, the confidence intervals are based on valid RD design models.

Table 5 presents the results of estimations of the RD design model for the bubble phase. The table presents the results of sharp and two-stage fuzzy RD designs. Running variables “Central bank & regulation”, “Fintech & blockchain” and “Law & tax investigation” are statistically significant. Following the results of manipulation tests (Table 4), we use the robust version of RD models to make point estimations of the running variables “Central bank & regulation” and “Law & tax investigation”. The economic magnitude of “Law and tax investigation” on the order of 5% is roughly 131.1 (235) basis points on the logarithm transformation of the Bitcoin price in the sharp (the second stage fuzzy) RD design models. The running variable “Fintech & blockchain” topic is statistically significant in both the conventional fuzzy and sharp RD designs. The result of the fuzzy RD design model shows that “Fintech & blockchain” has an economic magnitude of roughly 30 basis points on the discontinuity of the probability and of 698 basis points on the logarithm of the Bitcoin price. Figure 9 shows that entropies of news topics about “Fintech & blockchain” and “Law & tax investigation” gradually change across the cut-off points. Table 12 of the appendix presents the results of the RD design model estimating the impacts of news media on Bitcoin supply.

The results indicate that during a bubble period, investors are more concerned with Bitcoin security and blockchain technologies. Investors were notified that Bitcoin’s blockchain could be lost to its competitor in a market like Ethereum. However, the market believed that Bitcoin’s blockchain could be the next big thing. Perhaps the most important issue with the strongest impacts on market expectations was Bitcoin’s security. Several major cyber-attacks to Bitcoin platforms like Bitfinex had occurred. The market believed that Bitcoin had the potential to enter deep trouble. Every day, investors received news including keywords such as ”theft”, ”lost keys”, ”lost memory”, ”exchange collapse”, ”value collapse”, and ”hacking”. Our empirical results confirm that attention paid to a news article based on its unusualness can create discontinuity in the dynamics of Bitcoin prices.

Table 4: Falsification Test of RD Design: Bubble Phase

	Robust Effect	Robust P-value	Conventional Effect	Conventional P-value
Bank Finance Electronic	-14.779	0	37.735	.70342
Central Bank Reguation	21.973	.64856	20.933	.26165
Computer Fintech	111.113	.01016	92.544	.00024
Fintech blockchain	29.409	.03982	33.51	.86314
Law tax investigation	35.948	.15534	34.194	0
Security trading	63.872	.53353	65.39	.14893
Stock exchange	36.234	.12139	36.898	0

This table represents the results of manipulation tests by using local polynomial density functions. The null hypothesis of this test is that there is a discontinuity in running values. The P-value above 0.05 indicates the rejection of null of hypothesis.

Table 5: RD Design: Attention to News Article about Bitcoin (Bubble Phase)

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin
Running Variable	Central Bank Reguation	Fintech blockchain	Law tax investigation	Security trading	Stock exchange
Sharp:					
Conventional	0.486^{**}	0.485^{***}	1.201^{***}	0.177	0.0656
	(0.159)	(0.0740)	(0.147)	(0.135)	(0.139)
Bias-corrected	0.617^{***}	1.011	1.311^{***}	-0.0690	0.322^{*}
	(0.159)	(0.154)	(0.147)	(0.135)	(0.139)
Robust	0.617^{**}	1.011	1.311^{***}	-0.0690	0.322
	(0.218)	(0.198)	(0.196)	(0.179)	(0.185)
Fuzzy:					
First Stage					
Conventional	$-.1211^{***}$	$.3065^{***}$	$.3703^{*}$	$-.0375^{***}$	$.1543^{*}$
	(.0547)	(.0768)	(.063)	(.035)	(.0632)
Robust	-.0269	$.2276\ ^{***}$	.0776	-.051	-.0681
	(.0863)	(.1137)	(.0979)	(.0483)	(.0931)
Second Stage					
Conventional	2.979	6.982^{**}	2.622^{***}	7.585	0.718
	(2.472)	(4.616)	(0.574)	(7.233)	(1.409)
Bias-corrected	1.946	3.642	2.354^{***}	0.252	-1.245
	(2.472)	(2.408)	(0.574)	(7.233)	(1.409)
Robust	1.946	3.642	2.354^{**}	0.252	-1.245
	(3.694)	(3.796)	(0.887)	(10.26)	(2.049)
Observations	1687	1687	1687	1687	1687

This table represents the results of RD design with fuzzy and sharp approaches. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin price. Each column presents a regression model with each running variable. The kernel function is the Epanechnikov function, and the optimal bandwidth is obtained based on the Calonico et al. (2015) procedure. The cut-off point is defined as the highest value of Bitcoin price during December 2017. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level. Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.3 Before Bubble Phase

Table 7 reports the results of the second experiment related to changes in the price of Bitcoin, which reached $1,000 for the first time on January 3, 2017. This growth in the price of Bitcoin was mostly related to its advanced Blockchain technology. The technology attracted different sectors of the economy, rendering it as revolutionary as the Internet. The media created positive news about blockchain, and Bitcoin was one of its successful products. In our second experiment, we first checked the internal validity of the RD design models. The results of the manipulation test of this experiment are given in Table 6. The running variable “Fintech and blockchain” satisfies the requirement for an absence of systematic manipulation of the running variable in the robust RD design model. We then visually examine the density plot for this running variable to identify discontinuities around the cut-off point. The result of the sharp design shows a jump in the Bitcoin price of roughly exp(-1.36)= 0.256 as a result of attention to this news topic. Similarly, from the fuzzy RD design model, we find a significant economic magnitude of roughly exp(5.035)= 153.69 affecting the Bitcoin price. Due to information constraints, we cannot expect all investors to have received information on this news topic or to have similarly interpreted the news. However, there is enough evidence to believe that investors viewed Bitcoin as a successful application of blockchain technology.

Table 6: Falsification Test of RD Design (Before Bubble Phase)

	Robust Effect	Robust P-value	Conventional Effect	Conventional P-value
Bank Finance electronic	-5.198	0	42.604	.03873
Central Bank Reguation	37.951	.69388	30.077	.92877
Computer Fintech	41.515	00.24464	39.487	.90912
Fintech blockchain	75.599	.29723	86.193	0
Law tax investigation	12.946	.61264	11.348	.36145
Security trading	9.832	.22941	9.868	.88982
Stock exchange	40.166	.01991	48.036	0

This table represents the results of manipulation tests by using local polynomial density functions. The null hypothesis of this test is that there is a discontinuity in running values. The P-value above 0.05 indicates the rejection of null of hypothesis.

7.4 After Bubble Phase

Bitcoin is a decentralized currency, and it was designed to operate without governmental or central bank control or regulation. Bitcoin is a form of payment offering a high degree of privacy without regulation. However, after its market crashed in December 2017, many central banks started to work on the regulation of cryptocurrencies. Financial regulation is crucial to ensuring the stability of the market by controlling the demand and supply of money. Thus, after the Bitcoin bubble phase, there was increased attention to Bitcoin and discussion in the media on regulating this market.

In our third experiment, we run RD design models to understand the main contributions of media reports to changes in the dynamics of Bitcoin prices. Table 9 reports the results of RD design models with flow entropy in news topics used as running variables. From the internal validation test results given in Table 8, we find that the flow of information on

Table 7: RD Design: Attention to News Article about Bitcoin (Before Bubble Phase)

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin
Running Variable	Central Bank Reguation	Computer Fintech	Fintech blockchain	Law tax investigation	Security trading
Sharp:					
Conventional	-0.0947	0.649	-1.430***	0.0243	0.526
	(0.373)	(0.259)	(0.224)	(0.326)	(0.626)
Bias-corrected	-0.214	0.553	-1.364***	-0.0810	0.714
	(0.373)	(0.220)	(0.224)	(0.326)	(0.626)
Robust	-0.214	0.553	-1.364***	-0.0810	0.714
	(0.442)	(0.244)	(0.256)	(0.385)	(0.765)
Fuzzy:					
First Stage:					
Conventional	.2614**	.1458 *	-.2077	.0912	.26*
	.1362	.0982	.1054	.2186	.0941
Robust	.18	.0752	-.1859 *	.0993	.2502***
	.1396	.1017	.1079	.2207	.1013
Second Stage:					
Conventional	2.644	0.358	4.743**	-7.156	9.105
	(2.613)	(1.072)	(1.650)	(20.92)	(9.012)
Bias-corrected	2.321	0.0276	5.035**	-6.399	9.341
	(2.613)	(0.0826)	(1.650)	(20.92)	(9.012)
Robust	2.321	0.0276	5.035**	-6.399	9.341
	(2.640)	(0.0870)	(1.710)	(21.20)	(9.084)
Observations	1687	1687	1687	1687	1687

This table represents the results of RD design with fuzzy and sharp approaches. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin price. Each column presents a regression model with one running variable. The data are in a two-day lag. The kernel function is the Epanechnikov function, and the optimal bandwidth is obtained based on the Calonico et al. (2015) procedure. The cut-off point is defined as a time when the Botcoin price, for the first time, reached to $1,000. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level. Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

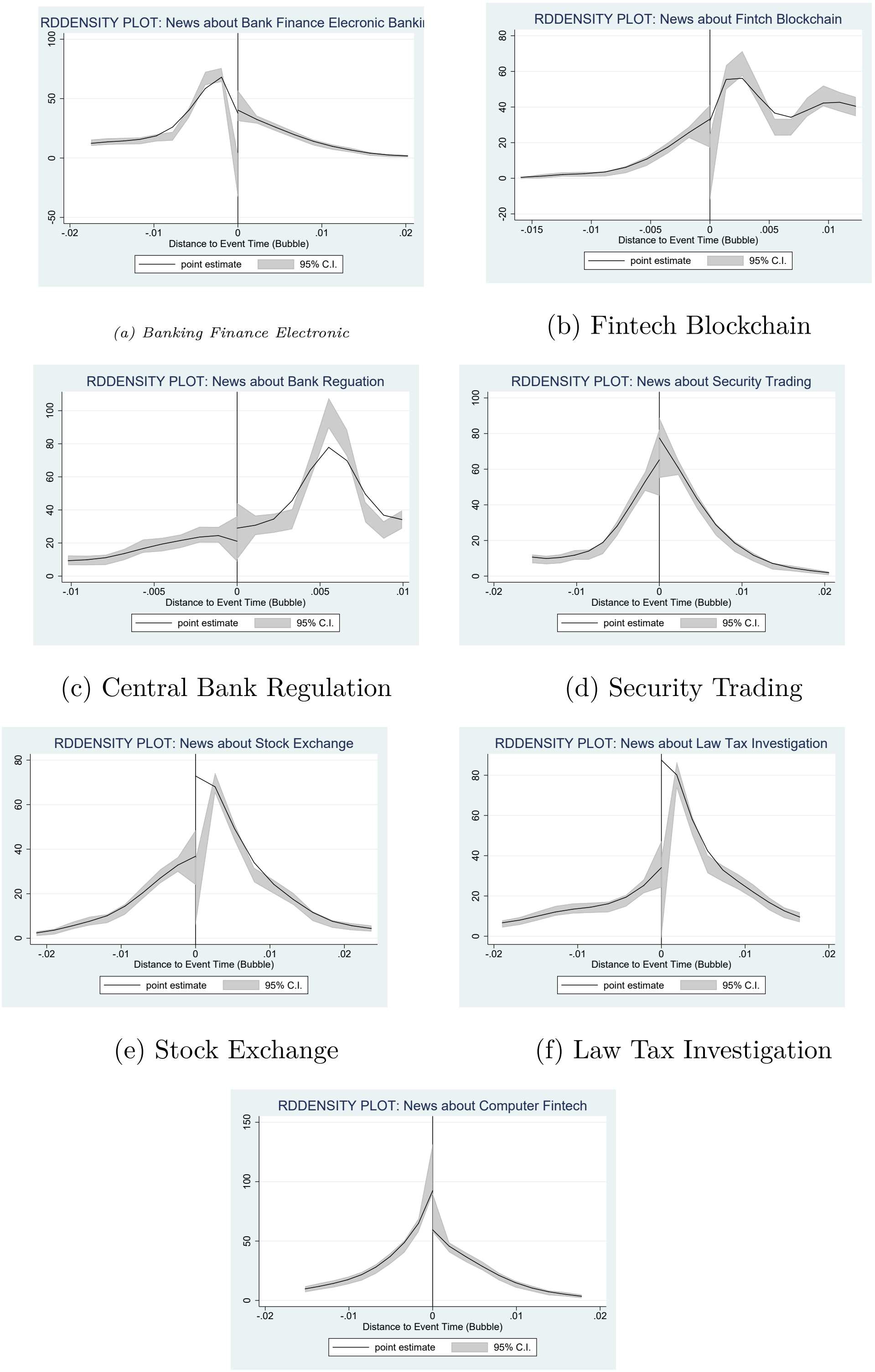
Table 8: Falsification Test of RD Design (After Bubble Phase)

	Robust Design Effect	Robust P-value	Conventional Effect	Conventional P-value
Bank Finance Electronic	104.394	.8066	69.97	.28848
Central Bank Reguation	110.806	.63608	79.488	.19063
Computer Fintech	45.853	.96341	42.498	.06734
Fintech blockchain	67.11	.09642	54.025	.26215
Law tax investigation	113.549	.26358	77.537	.58741
security trading	12.876	.06602	12.876	.14323
stock exchange	118.535	.07958	96.477	.00505

This table reports the results of falsification Test of RD Design (After Bubble Phase). The null hypothesis of this test is that there is a discontinuity in running values. The P-value above 0.05 indicates the rejection of null of hypothesis.

Bitcoin's regulation altered Bitcoin price dynamics significantly. The economic magnitude of this impact according to the sharp RD design model is extremely high, as the media dominated news reports on ways of regulating the cryptocurrency market. We can conclude that discontinuity in the Bitcoin price was mostly governed by news articles focused on regulation, which attracted the attention of market participants.

Figure 4: **Density Plot of Entropy of News Topics (Bubble Phase)**



(a) Banking Finance Electronic

(b) Fintech Blockchain

(c) Central Bank Regulation

(d) Security Trading

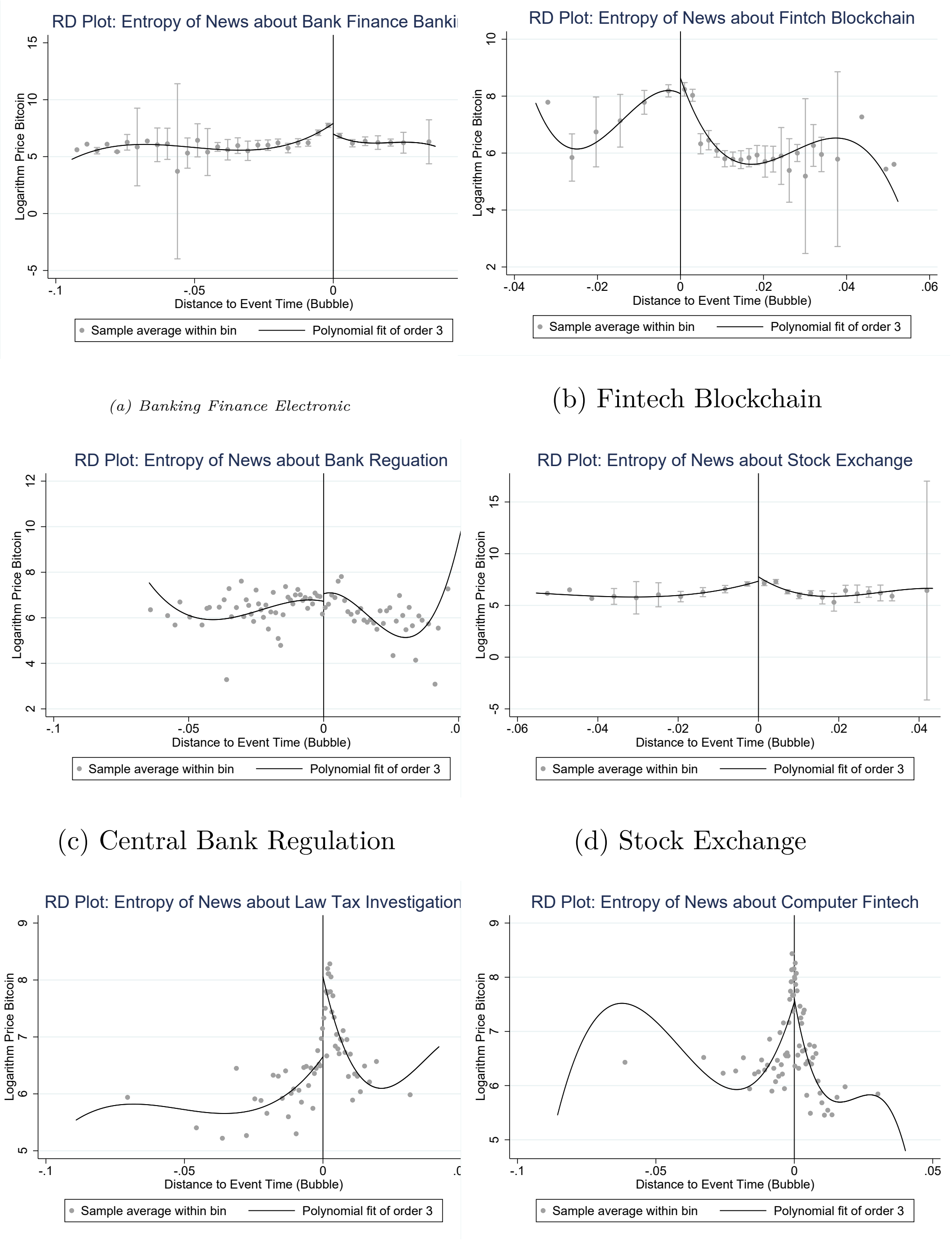
(e) Stock Exchange

(f) Law Tax Investigation

(g) Computer Fintech

These figures show the histogram, estimated density of the forcing variables. The cut-off point is specified as the highest value of Bitcoin price during December 2017. The figures confirm there is no manipulation. The findings from these figures consist with the formal manipulation McCrary (2008) test for each running variable which is not statistically significant.

Figure 5: **RD Design Plots of Entropy of News Topics (Bubble Phase)**



(a) Banking Finance Electronic

(b) Fintech Blockchain

(c) Central Bank Regulation

(d) Stock Exchange

(e) Law Tax Investigation

(f) Computer Fintech

These figures show the RD design plot with a three-degree polynomial function. The plots show the value of the function at the running variable. The horizontal line indicated the cut-off point. Solid lines correspond to nonlinear fit with 95% confidence intervals for two sides of the cut-off point. The cut-off point is specified as the highest value of Bitcoin price during December 2017.

Table 9: RD Design: Effects of Attention to News on Bitcoin Price (After Bubble Phase)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent	Ln.Price	Ln.Price	Ln.Price	Ln.Price	Ln.Price	Ln.Price
Variable	Bitcoin	Bitcoin	Bitcoin	Bitcoin	Bitcoin	Bitcoin
Running	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy
Variable	Bank Finance electronic	Centr- Bank Regulation	Computer Fintech	Fintech Blockchain	Law tax investigation	security trading
Sharp: :						
Conventional	0.341	2.498***	-0.751	-0.219	0.228	0.0916
	(0.184)	(0.520)	(0.483)	(0.168)	(0.184)	(0.120)
Bias-corrected	0.330	2.732***	-0.968*	-0.164	0.206	0.329**
	(0.184)	(0.569)	(0.483)	(0.168)	(0.184)	(0.120)
Robust	0.330	2.732***	-0.968	-0.164	0.206	0.329
	(0.205)	(0.629)	(0.529)	(0.191)	(0.206)	(0.192)
Fuzzy:						
First Stage						
Conventional	.0957***	.306***	-.1712***	-.031	.0613	.0381
	.0415	.0484	.0857	.063	.0486	.0317
Robust	.1085***	.3588***	-.2027***	-.0071	.0762	.059
	.0468	.0545	.089	.0701	.0556	.056
Second Stage						
Conventional	-64.94	193.8**	27.43	3.492	-168.0	-0.261
	(495.0)	(317.9)	(26.09)	(2.886)	(4781.6)	(0.492)
Bias-corrected	-200.7	159.8**	3.769	3.487	-1990.1	-0.935
	(495.0)	(262.1)	(26.09)	(2.886)	(4781.6)	(0.492)
Robust	-200.7	159.8**	3.769	3.487	-1990.1	-0.935
	(563.4)	(309.7)	(29.63)	(3.318)	(5495.6)	(0.681)
Observations	1687	1687	1687	1687	1687	1687

This table represents the results of RD design with fuzzy and sharp approaches. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin price. Each column presents a regression model with one running variable. The data are in a two-day lag. The kernel function is the Epanechnikov function, and the optimal bandwidth is obtained based on the Calonico et al. (2015) procedure. The cut-off point is when Bitcoin price reached $1,000 for the first time on January 3, 2017. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level. Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

8 Robustness Analysis

This section provides examinations of the robustness of our results to alternative specifications. Generally, There are some issues about RD designs with regards to the misspecification of functional form, the selection of kernel functions and obtaining the bandwidth parameters. We report the reuslts of bias correction on confidence interval (Calonico et al., 2018) with regards to a selection of kernel densities and the degrees of local polynomial regression. We repeat the RD designs with different kernels functions and degrees of local polynomial. Table 16, shown in appendix, presents the results. As mentioned earlier, we use the procedure explained in Calonico et al. (2015) to obtain the optimal bandwidths. As a robustness check, we run Rd designs with a set of bandwidths. Table 17, shown in appendix, reports the results. We can observe that the conclusion of no significant difference in the models' outcomes, i.e., the baseline model is not sensitive to the selection of bandwidth in the local linear regression estimator.

As discussed earlier, RD design can be characterized as ”discontinuity at a cut-off point,” or ”local randomization” approaches. The latter approach is based on differences between objects which miss treatment and those who received treatment are random. Table 15 presents results of the implementation of the local randomization method (Cattaneo et al., 2015) as a robustness comparison to conventional RD inference methodologies. The results show that news articles contain information on ”Law & tax investigation”, ”Fintech & blockchain” and ”Security trading” topics had impacts on the Bitcoin price. The results of the ”local randomization” approach confirmed the validation of the output of the conventional approach.

Another uncertainty in our RD design framework might be the misspecification of cut-off points. We shift two days the cut-off points to create tie-breaker experiments and examine the validation of the results of the baseline model. We repeat the same procedure of RD design during the bubble phase with a new cut-off point. Table 11 reports the results of sharp and fuzzy RD design approaches. We can observe slightly differences in the outcomes of the RD design models. We speculate this difference is due to earlier considerations of investors to the security of the digital currency market. However, after while investors

Table 10: RD Design: Randomized (Bubble Phase)

	limit window (left)	limit window (right)	sample size (left)	sample size (right)	statistics statistics	randomization P-value	asymptotic P-value
Central Bank & Reguation	-.0646	.0538	424	1187	.0546	.437	.4545
Fintech & blockchain	-.0349	.0524	251	1362	-1.2684	0	0
Law & tax investigation	-.0891	.0419	617	995	.848	0	0
Security trading	-.0285	.0376	852	761	-.7038	0	0
Stock exchange	-.0553	.0434	1612	1049	-.0367	.613	.5997

This paper presents the results of RD design with local randomization approach. We repeat the baseline model. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin price. The table presents information on the left and right limits of the window used, sample size in the used window, the sample size in the used window to the right and left of the cut-off point moreover, an observed statistic with randomization and asymptotic p-values.

decided to exit the market with receiving more information about some illegal activities.

Table 11: RD Design: Attention to News Article about Bitcoin Shift two days (Bubble Phase)

	(1)	(2)	(3)	(4)	(5)
Dependent					
Variable	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin
Running					
Variable	Central Bank Reguation	Fintech blockchain	Law tax investigation	Security trading	Stock exchange
Sharp:					
Conventional	0.301^{*}	0.237^{***}	0.668^{***}	0.131	0.205
	(0.151)	(0.0198)	(0.0951)	(0.0879)	(0.109)
Bias-corrected	-0.0245	0.362^{***}	0.546^{***}	-0.136	0.522^{***}
	(0.151)	(0.0303)	(0.0951)	(0.0879)	(0.109)
Robust	-0.0245	0.362^{***}	0.546^{***}	-0.136	0.522^{***}
	(0.204)	(0.0376)	(0.122)	(0.117)	(0.146)
Observations	1613	1613	1613	1613	1613
Fuzzy:					
Conventional	8.325	7.800^{***}	3.301	4.340	3.715^{**}
	(16.20)	(2.629)	(2.046)	(2.758)	(1.353)
Bias-corrected	17.80	6.700^{***}	7.230^{***}	4.030	8.032^{***}
	(16.20)	(2.258)	(2.046)	(2.758)	(1.353)
Robust	17.80	6.700^{***}	7.230^{**}	4.030	8.032^{***}
	(23.27)	(2.906)	(2.517)	(3.904)	(2.005)
Observations	1613	1613	1613	1613	1613

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table represents the results of replication of base line model for the bubble pahse. The cut-off point is shfted to two before of bubble crashing. RD design with fuzzy and sharp approaches. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin price. Each column presents a regression model with one running variable. The data are in a two-day lag. The kernel function is the Epanechnikov function, and the optimal bandwidth is obtained based on the Calonico et al. (2015) procedure. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level. Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The findings given in this section underscore how the robustness of baseline results corresponds to alternative specifications.

9 Conclusion

The price of digital currencies like Bitcoin varies considerably over time and is governed by attention to news media. In this paper, we study the causes of the burst of the Bitcoin bubble. From a large sample of news articles on Bitcoin, we construct daily textual variables. Each textual variable represents the entropy value of topics extracted from news documents. We then develop a set of regression discontinuity designs to make inferences about which news topic spurred changed in Bitcoin prices in different periods.

Based on this quasi-natural experiment, we find that early on, informational flows in technology focus on a digital asset known as blockchain, and its application in Fintech attracted the attention of investors to this unregulated market, ultimately resulting in the growth the bubble. During the bubble phase, uncertainty surrounding the security of this currency led investors to exit the market and resulted in the bursting of the bubble. The open-ended debate on the regulation and governing of this digital currency is the principal factor that has shaped fluctuations in this market. Importantly, our results illustrate the importance of the timing of news items in addition to their content.

Our findings contribute to the ongoing debate on the role and economic impacts of the media. We present important findings for understanding the causes of the growth and bursting of bubbles in the financial market. Apart from identifying the effects of news media on the growth and bursting of the bubble in the digital currency market, future research should examine the impacts of news media on fluctuations and volatility in this market. It is worth investigating how attention to certain topics leads to high- or low-volatility regimes in this market. This paper offers a framework to construct a soft informational index from high-frequency textual data and to cover a lack of insufficient hard information about economic phenomena.

References

Abadi, J., Brunnermeier, M., 2018. Blockchain economics. Technical Report. mimeo Princeton University.

Agarwal, R., 2015. Breaking through the zero lower bound. International Monetary Fund.

Ali, R., Barrdear, J., Clews, R., Southgate, J., 2014. The economics of digital currencies. Bank of England Quarterly Bulletin , Q3.

Assaf, A., 2018. Testing for bubbles in the art markets: An empirical investigation. Economic Modelling 68, 340–355.

Athey, S., 2017. Beyond prediction: Using big data for policy problems. Science 355, 483–485.

Athey, S., 2018. The impact of machine learning on economics, in: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press.

Auer, R., Claessens, S., 2018. Regulating cryptocurrencies: assessing market reactions. BIS Quarterly Review September .

Bandiera, O., Hansen, S., Prat, A., Sadun, R., 2017. Ceo behavior and firm performance. Technical Report. National Bureau of Economic Research.

Ben-Rephael, A., Da, Z., Israelsen, R.D., 2017. It depends on where you search: Institutional investor attention and underreaction to news. The Review of Financial Studies 30, 3009–3047.

BIS, 2018. Annual economic report 2018 .

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research 3, 993–1022.

Böhme, R., Christin, N., Edelman, B., Moore, T., 2015. Bitcoin: Economics, technology, and governance. Journal of Economic Perspectives 29, 213–38.

Bordo, M.D., Levin, A.T., 2017. Central bank digital currency and the future of monetary policy. Technical Report. National Bureau of Economic Research.

Boyd, J.H., Hu, J., Jagannathan, R., 2005. The stock market's reaction to unemployment news: Why bad news is usually good for stocks. Journal of Finance 60, 649–672.

Brooks, C., Katsaris, A., 2003. Rational speculative bubbles: an empirical investigation of the london stock exchange. Bulletin of Economic Research 55, 319–346.

Budak, C., Goel, S., Rao, J.M., Zervas, G., 2014. Do-not-track and the economics of third-party advertising. Boston University, School of Management Research Paper 2505643.

Calonico, S., Cattaneo, M.D., Farrell, M.H., 2018. On the effect of bias estimation on coverage accuracy in nonparametric inference. Journal of the American Statistical Association 113, 767–779.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82, 2295–2326.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2015. Optimal data-driven regression discontinuity plots. Journal of the American Statistical Association 110, 1753–1769.

Cao, J., Xia, T., Li, J., Zhang, Y., Tang, S., 2009. A density-based method for adaptive lda model selection. Neurocomputing 72, 1775–1781.

Cattaneo, M.D., Frandsen, B.R., Titiunik, R., 2015. Randomization inference in the regression discontinuity design: An application to party advantages in the us senate. Journal of Causal Inference 3, 1–24.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J.L., Blei, D.M., 2009. Reading tea leaves: How humans interpret topic models, in: Advances in neural information processing systems, pp. 288–296.

Chang, T., Gil-Alana, L., Aye, G.C., Gupta, R., Ranjbar, O., 2016. Testing for bubbles in the brics stock markets. Journal of Economic Studies 43, 646–660.

Charemza, W.W., Deadman, D.F., 1995. Rational and intrinsic bubbles: a reinterpretation of empirical results. Applied Financial Economics 5, 199–202.

Chen, M.A., Wu, Q., Yang, B., 2019. How Valuable Is FinTech Innovation? The Review of Financial Studies 32, 2062–2106.

Chiu, J., Koeppl, T.V., 2019. Blockchain-Based Settlement for Asset Trading. The Review of Financial Studies 32, 1716–1753.

Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. The Journal of Finance 63, 1977–2011.

Cong, L.W., He, Z., 2019. Blockchain Disruption and Smart Contracts. The Review of Financial Studies 32, 1754–1797.

Corbet, S., Lucey, B., Urquhart, A., Yarovaya, L., 2019. Cryptocurrencies as a financial asset: A systematic analysis. International Review of Financial Analysis 62, 182–199.

Diba, B.T., Grossman, H.I., 1988. The theory of rational bubbles in stock prices. The Economic Journal 98, 746–754.

Drescher, C., Herz, B., 2016. What determines simultaneous asset bubbles? an empirical analysis. Applied Economics 48, 35–51.

Easley, D., O'Hara, M., Basu, S., 2019. From mining to markets: The evolution of bitcoin transaction fees. Journal of Financial Economics .

Engelberg, J., Gao, P., 2011. In search of attention. The Journal of Finance 66, 1461–1499.

Engsted, T., 2016. Fama on bubbles. Journal of Economic Surveys 30, 370–376.

Foley, S., Karlsen, J.R., Putni??, T.J., 2019. Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies? The Review of Financial Studies 32, 1798–1853.

Fung, B., Hendry, S., Weber, W.E., et al., 2018. Swedish Riksbank notes and enskilda bank notes: lessons for digital currencies. Technical Report. Bank of Canada.

Gandal, N., Hamrick, J., Moore, T., Oberman, T., 2018. Price manipulation in the bitcoin ecosystem. Journal of Monetary Economics 95, 86–96.

Gentzkow, M., Kelly, B.T., Taddy, M., 2017. Text as data. Technical Report. National Bureau of Economic Research.

Giglio, S., Maggiori, M., Stroebel, J., 2016. No-bubble condition: Model-free tests in housing markets. Econometrica 84, 1047–1091.

Glasserman, P., Mamaysky, H., 2017. Does unusual news forecast market stress? Technical Report. working paper.

Hahn, J., Todd, P., Van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69, 201–209.

Hansen, S., McMahon, M., Prat, A., 2017. Transparency and deliberation within the fomc: a computational linguistics approach. The Quarterly Journal of Economics 133, 801–870.

Hirshleifer, D., Lim, S.S., Teoh, S.H., 2009. Driven to distraction: Extraneous events and underreaction to earnings news. The Journal of Finance 64, 2289–2325.

Huberman, G., Leshno, J., Moallemi, C.C., 2017. Monopoly without a monopolist: An economic analysis of the bitcoin payment system .

Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. The Review of Economic Studies 79, 933–959.

Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. Journal of Econometrics 142, 615–635.

Lee, D.S., Card, D., 2008. Regression discontinuity inference with specification error. Journal of Econometrics 142, 655–674.

Lei, V., Noussair, C.N., Plott, C.R., 2001. Nonspeculative bubbles in experimental asset markets: Lack of common knowledge of rationality vs. actual irrationality. Econometrica 69, 831–859.

Makarov, I., Schoar, A., 2019. Trading and arbitrage in cryptocurrency markets. Journal of Financial Economics .

Manela, A., Moreira, A., 2017. News implied volatility and disaster concerns. Journal of Financial Economics 123, 137–162.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics 142, 698–714.

Menzly, L., Ozbas, O., 2010. Market segmentation and cross-predictability of returns. The Journal of Finance 65, 1555–1580.

Moinas, S., Pouget, S., 2013. The bubble game: An experimental study of speculation. Econometrica 81, 1507–1539.

Mueller, H., Rauh, C., 2018. Reading between the lines: Prediction of political violence using newspaper text. American Political Science Review 112, 358–375.

Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system .

Nimark, K.P., Pitschner, S., et al., 2016. Delegated information choice. volume 11323. Centre for Economic Policy Research.

Odean, T., 1998. Are investors reluctant to realize their losses? The Journal of finance 53, 1775–1798.

Odean, T., 1999. Do investors trade too much? American economic review 89, 1279–1298.

Salton, G., McGill, M.J., 1986. Introduction to modern information retrieval .

Schilling, L., Uhlig, H., 2018. Some simple bitcoin economics. Technical Report. National Bureau of Economic Research.

Shannon, C.E., 1948. A mathematical theory of communication. Bell system technical journal 27, 379–423.

Sims, C.A., 2003. Implications of rational inattention. Journal of Monetary Economics 50, 665–690.

Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M., 2005. Sharing clusters among related groups: Hierarchical dirichlet processes, in: Advances in neural information processing systems, pp. 1385–1392.

Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. Journal of Finance 62, 1139–1168.

Thistlethwaite, D.L., Campbell, D.T., 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. Journal of Educational Psychology 51, 309.

Thorsrud, L.A., 2018. Words are the new numbers: A newsy coincident index of the business cycle. Journal of Business & Economic Statistics , 1–17.

Weber, W.E., 2016. A Bitcoin standard: Lessons from the gold standard. Technical Report. Bank of Canada Staff Working Paper.

Yermack, D., 2017. Corporate governance and blockchains. Review of Finance 21, 7–31.

A Supply of Bitcoin

The number of Bitcoin sent from addresses can be a good proxy for the supply of the Bitcoin. All Payments are made by Bitcoin addresses. Bitcoin uses unique addresses for all transactions and allows anyone to see the amount of volume has been transferred from one address to another. This digital currency has a finite supply due to its cryptography system. Once it reaches the limit of its supply, it might be converted a deflationary currency. Before that time, almost the Bitcoin miners can control this market and indirectly decide about the volume of Bitcoin must enter the market. On the other hand, the decision of miners can be governed by media news about this market. Therefore, we replicate our experiments with the number of Bitcoin sent from addresses as the dependent variable. The results of experiments during, after and before the bubble phase are shown in Tables 12, 13 and 14, respectively. The results of the effects of attention to media on the supply of Bitcoin is quite similar to the results the price Bitcoin. The main difference is the magnitude of coefficients. All together, we can conclude attention of suppliers is also governed by media.

Table 12: RD Design: Attention to News about Bitcoin (Bubble Phase) (Supply)

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Ln.Sent.From Address	Ln.Sent.From Address	Ln.Sent.From Address	Ln.Sent.From Address	Ln.Sent.From Address
Running Variable	Central Bank Reguation	Fintech blockchain	Law tax investigation	security trading	stock exchange
Sharp					
Conventional	-0.0438	0.864***	0.191***	-0.133***	-0.0534
	(0.0459)	(0.0332)	(0.0388)	(0.0398)	(0.0398)
Bias-corrected	0.0496	1.044	0.121**	-0.191***	0.0248
	(0.0459)	(0.0401)	(0.0388)	(0.0398)	(0.0398)
Robust	0.0496	1.044	0.121*	-0.191***	0.0248
	(0.0622)	(0.0521)	(0.0496)	(0.0517)	(0.0520)
Fuzzy :					
First Stage					
Conventional	.2328***	-.0778***	.3709***	.1***	.1886***
	.0297	.0463	.031	.0202	.0322
Robust	.2607 ***	.0791	.444***	.0329	.2591***
	.0389	.0604	.043	.028	.0448
Second Stage					
Conventional	-0.188	6.584	0.516***	-1.332*	-0.283
	(0.205)	(7.900)	(0.103)	(0.525)	(0.232)
Bias-corrected	0.236	25.59**	0.226*	-2.805***	0.237
	(0.205)	(30.70)	(0.103)	(0.525)	(0.232)
Robust	0.236	25.59*	0.226	-2.805***	0.237
	(0.277)	(40.00)	(0.132)	(0.703)	(0.305)
Observations	1613	1613	1613	1613	1613

This table shows the results of RD design with the logarithm transformation of the volume of Bitcoin sent from addresses as the dependent variable. The cut-off point is the time when the Bitcoin reached its highest value in December 2017. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin price. Each column presents a regression model with one running variable. The data are in a two-day lag. The kernel function is the Epanechnikov function, and the optimal bandwidth is obtained based on the Calonico et al. (2015) procedure. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level. Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: RD Design: Attention to News about Bitcoin (Before Bubble Phase) (Supply)

	(1)	(2)	(3)	(4)	(5)
Dependent	Ln.Sent.From	Ln.Sent.From	Ln.Sent.From	Ln.Sent.From	Ln.Sent.From
Variable	Address	Address	Address	Address	Address
Running					
Variable	Central Bank Reguation	Computer Fintech	Fintech blockchain	Law tax investigation	Security trading
Sharp					
Conventional	0.595*	1.426**	-0.331***	0.223	0.241
	(0.231)	(0.166)	(0.0751)	(0.289)	(0.293)
Bias-corrected	0.654**	1.362**	-0.0943	0.0308	0.208
	(0.231)	(0.158)	(0.0751)	(0.289)	(0.293)
Robust	0.654**	1.362**	-0.0943	0.0308	0.208
	(0.234)	(0.163)	(0.0801)	(0.293)	(0.298)
Fuzzy					
First Stage					
Conventional	.2214 **	.3243***	-.4996***	.3366**	.2916**
	.0572	.0498	.0375	.0684	.0732
Robust	.2864**	.3542 **	-.3811***	.1172 **	.2501**
	.0677	.0582	.0472	.0818	.0887
Second Stage					
Conventional	0.510*	1.934***	0.551***	0.716***	0.453
	(0.244)	(0.253)	(0.0558)	(0.217)	(0.249)
Bias-corrected	0.719**	1.828***	0.352***	0.823***	0.651**
	(0.244)	(0.240)	(0.0558)	(0.217)	(0.249)
Robust	0.719*	1.828***	0.352***	0.823**	0.651*
	(0.285)	(0.277)	(0.0680)	(0.259)	(0.299)
Observations	1613	1613	1613	1613	1613

This table reports the results of RD design from the replication of the second experiment for supply of the Bitcoin. The cut-off point is defined as the time the price of Bitcoin reached first time to $1000. This table represents the results of replication of base line model for the bubble pahse. RD design with fuzzy and sharp approaches. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin sent from addresses. Each column presents a regression model with one running variable. The data are in a two-day lag. The kernel function is the Epanechnikov function, and the optimal bandwidth is obtained based on the Calonico et al. (2015) procedure. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level. Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: RD Design: Attention to News on Bitcoin Supply (After Bubble Phase) (Supply)

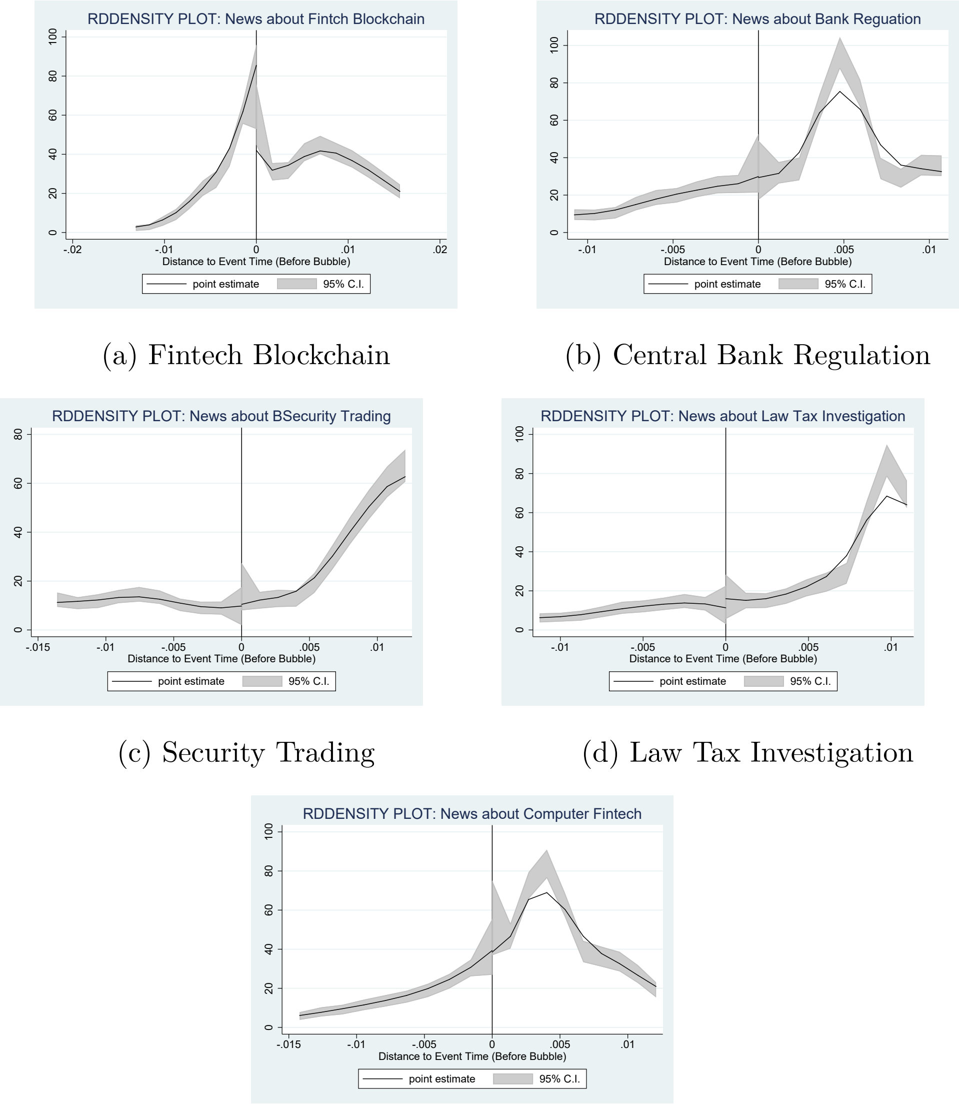
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent	Ln.Bitcoin__sent	Ln.Bitcoin__sent	Ln.Bitcoin__sent	Ln.Bitcoin__sent	Ln.Bitcoin__sent	Ln.Bitcoin__sent
variable	from__addresses	from__addresses	from__addresses	from__addresses	from__addresses	from__addresses
Running						
Variable	Bank Finance electronic	Central Bank Regulation	Computer Fintech	Fintech Blockchain	Law tax investigation	security trading
Sharp						
Conventional	0.0413	1.151**	-0.113	-0.00122	0.0174	-0.116
	(0.0481)	(0.0553)	(0.0928)	(0.0393)	(0.0477)	(0.0816)
Bias-corrected	0.0390	1.174***	-0.150	0.00657	0.0125	-0.0653
	(0.0481)	(0.0564)	(0.0928)	(0.0393)	(0.0477)	(0.0816)
Robust	0.0390	1.174**	-0.150	0.00657	0.0125	-0.0653
	(0.0546)	(0.0618)	(0.104)	(0.0440)	(0.0539)	(0.132)
Fuzzy :						
First stage						
Conventional	.0009 ***	.1559***	-.0626***	-.0583**	.0166 ***	-.3453
	.0478	.0475	.0206	.0683	.0495	.3583
Robust	.0158	.1911***	-.1163***	-.0365	.0371	-.2149
	.0549	.0559	.0252	.0762	.0566	.4467
Second stage						
Conventional	0.492	1.425**	0.616	-0.329	0.459	0.797
	(0.436)	(0.177)	(0.644)	(2.052)	(0.676)	(0.688)
Bias-corrected	0.365	1.493**	0.582	-0.869	0.331	1.786**
	(0.436)	(0.186)	(0.644)	(2.052)	(0.676)	(0.688)
Robust	0.365	1.493**	0.582	-0.869	0.331	1.786
	(0.508)	(0.216)	(0.669)	(2.289)	(0.777)	(1.003)
Observations	1613	1613	1613	1613	1613	1613

This table shows the results of RD design from the replication of the third experiment for supply of the Bitcoin. This table shows the results of RD design with the logarithm transformation of the volume of Bitcoin sent from addresses as the dependent variable. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin price. Each column presents a regression model with one running variable. The data are in a two-day lag. The kernel function is the Epanechnikov function, and the optimal bandwidth is obtained based on the Calonico et al. (2015) procedure. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level.

Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B RD Plots

Figure 6: **Density Plot of Entropy of News Topics (Before Bubble Phase)**



(a) Fintech Blockchain

(b) Central Bank Regulation

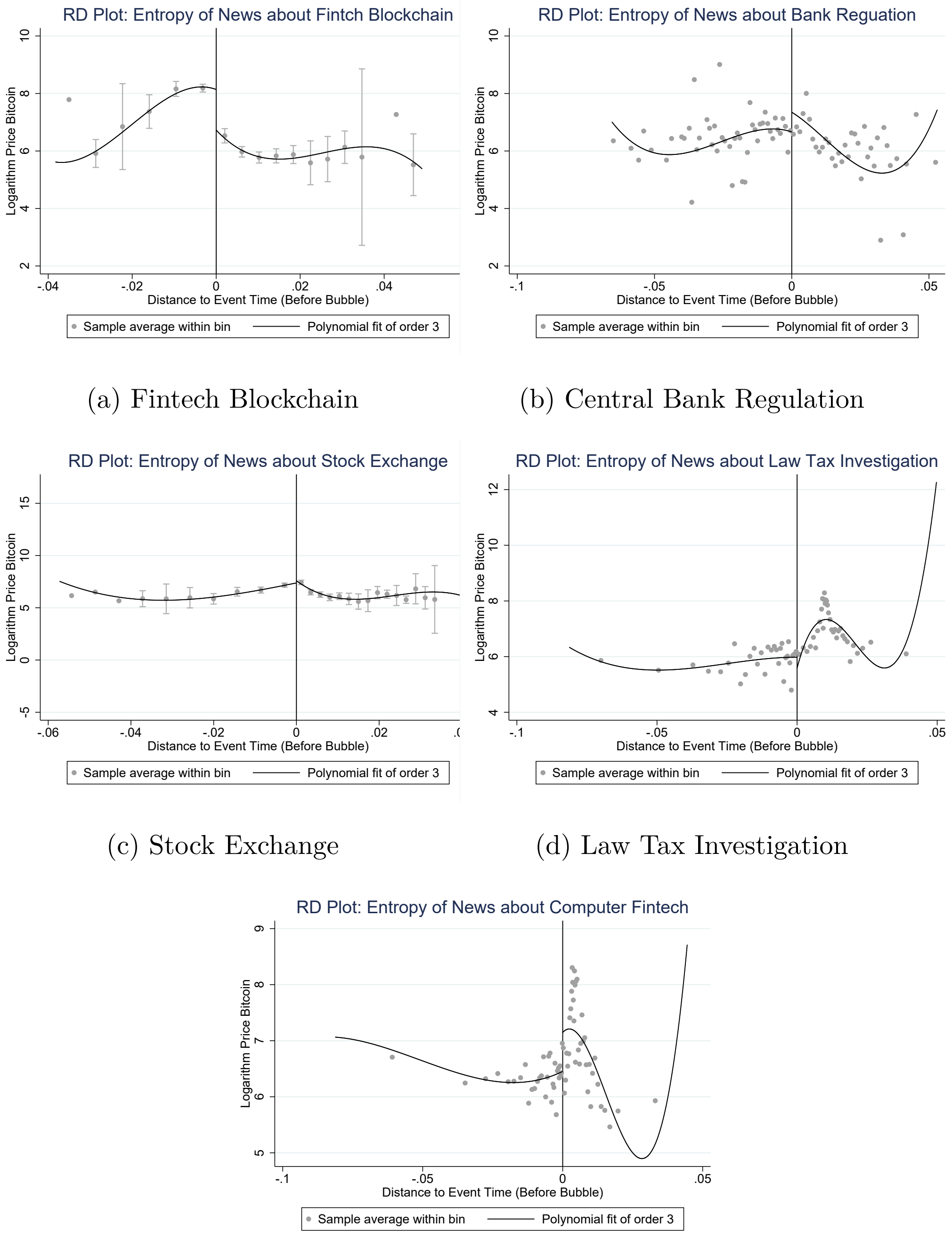
(c) Security Trading

(d) Law Tax Investigation

(e) Computer Fintech

These figures show the histogram, estimated density of the forcing variables. The cut-off point is specified as the time of jumping off the price of Bitcoin and reached the $1,000 for the first time on January 3, 2017. The figures confirm there is no manipulation. The findings from these figures consist with the formal manipulation McCrary (2008) test for each running variable which is not statistically significant.

Figure 7: **RD Design Plots of Entropy of News Topics (Before Bubble Phase)**



(a) Fintech Blockchain

(b) Central Bank Regulation

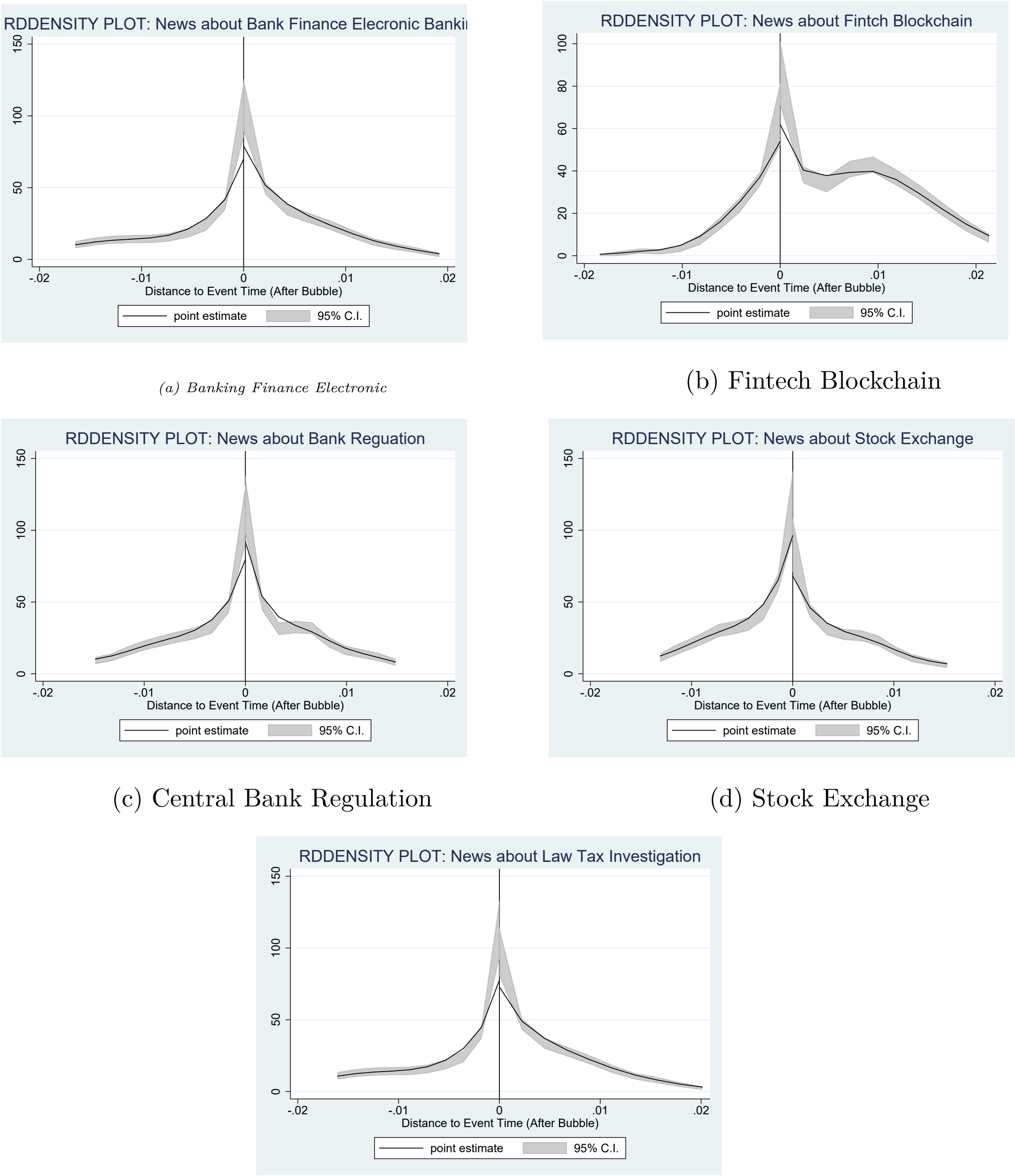
(c) Stock Exchange

(d) Law Tax Investigation

(e) Computer Fintech

These figures show the RD design plot with a three-degree polynomial function. The plots show the value of the function at the running variable. The horizontal line indicated the cut-off point. Solid lines correspond to nonlinear fit with 95% confidence intervals for two sides of the cut-off point. The cut-off point is specified as the time of jumping off the price of Bitcoin and reached the $1,000 for the first time on January 3, 2017.

Figure 8: **Density Plot of Entropy of News Topics (After Bubble Phase)**



(a) Banking Finance Electronic

(b) Fintech Blockchain

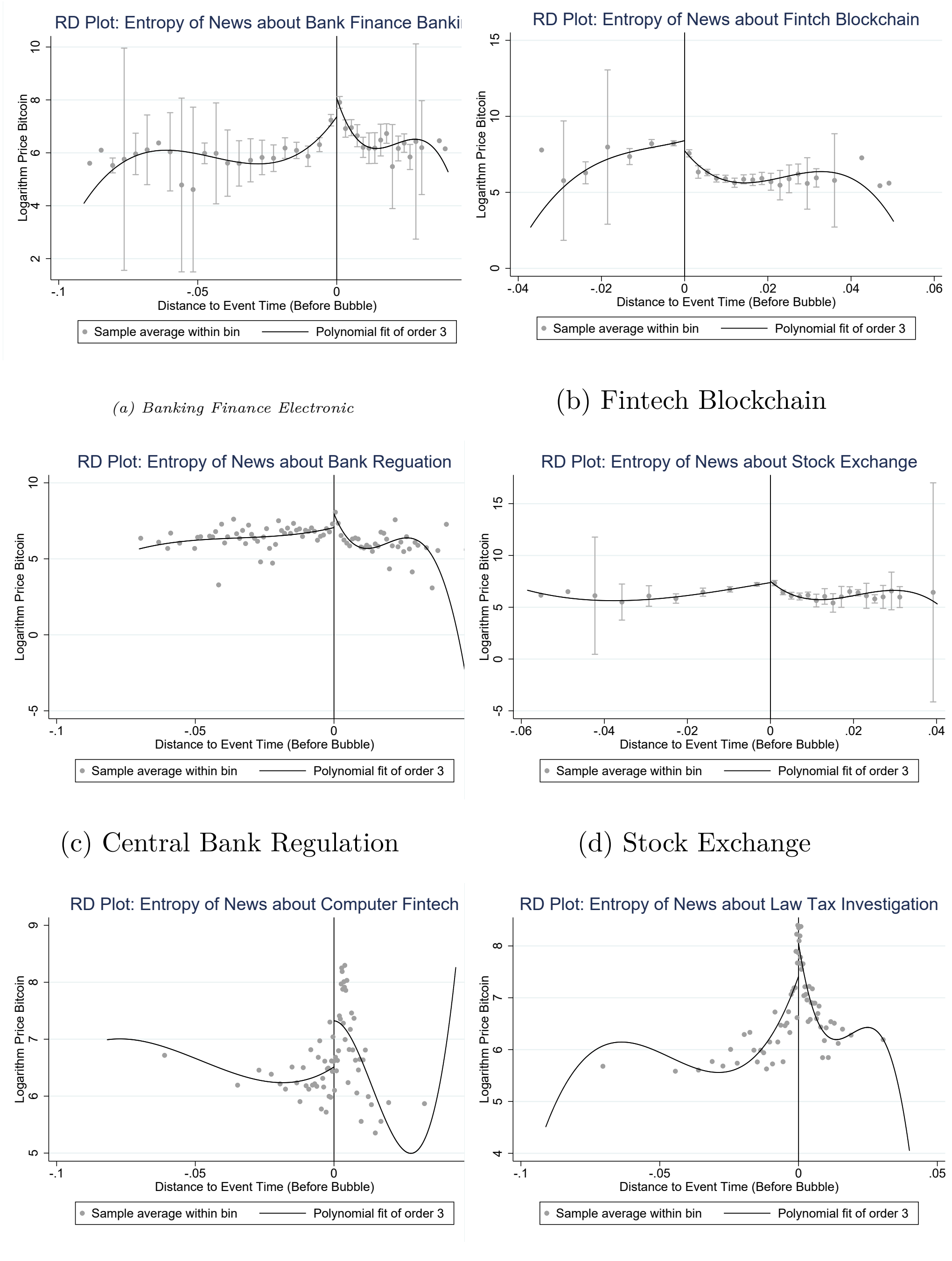
(c) Central Bank Regulation

(d) Stock Exchange

(e) Law Tax Investigation

These figures show the histogram, estimated density of the forcing variables. The cut-off point is specified as the time when the Bitcoin price dropped to below $6000 (after bubble phase). The figures confirm there is no manipulation. The findings from these figures consist with the formal manipulation McCrary (2008) test for each running variable which is not statistically significant.

Figure 9: **RD Design Plots of Entropy of News Topics (After Bubble Phase)**



(a) Banking Finance Electronic

(b) Fintech Blockchain

(c) Central Bank Regulation

(d) Stock Exchange

(e) Computer Fintech

(f) Law Tax Investigation

These figures show the RD design plot with a three-degree polynomial function. The plots show the value of the function at the running variable. The horizontal line indicated the cut-off point. Solid lines correspond to nonlinear fit with 95% confidence intervals for two sides of the cut-off point. The cut-off point is specified as the time when the Bitcoin price dropped to below $6000 (after bubble phase).

C Robustness

Table 15: RD Design: Randomized Robustness on Supply (Bubble Phase)

	limit window	limit window	sample size	sample size	statistics	randomization	asymptotic
	(left)	(right)	(left)	(right)	statistics	P-value	P-value
Central Bank Reguation	-.0646	.0538	424	1187	-.0587	.056	.0527
Fintech & blockchain	-.0349	.0524	251	1362	-.3684	0	0
Law & Tax Investigation	-.0891	.0419	617	995	.226	0	0
Security trading	-.0285	.0376	852	761	-.1937	0	0
Stock exchange	-.0553	.0434	563	1049	-.0497	.059	.0643

This paper presents the results of RD design with local randomization approach. We repeat the baseline model. The running (forcing) variables are daily entropy of news topics, and the dependent variable is logarithm transformation of the Bitcoin sent from addresses. The table presents information on the left and right limits of the window used, sample size in the used window, the sample size in the used window to the right and left of the cut-off point moreover, an observed statistic with randomization and asymptotic p-values.

Table 16: Robustness for Specifications of Degree Polynomial and Kernel Functions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Running	Law & Tax	Law & Tax	Law & Tax	Law & Tax	Law & Tax	Law & Tax	Law & Tax
Variable	Investigation	Investigation	Investigation	Investigation	Investigation	Investigation	Investigation
Dependent							
Variable	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin
Kernel	epanechnikov	triangular	uniform	epanechnikov	epanechnikov	epanechnikov	epanechnikov
Bandwidth	h(0.05 10)	h(0.05 10)	h(0.05 10)	h(0.05 10)	h(0.05 10)	h(0.05 10)	h(0.05 10)
Degree	P(1)	P(1)	P(1)	P(2)	P(3)	P(4)	P(5)
Sharp							
Conventional	1.111***	2.949***	1.146***	1.152***	1.067***	0.869***	0.506*
	(0.0916)	(0.277)	(0.0883)	(0.123)	(0.155)	(0.194)	(0.239)
Bias-corrected	1.152***	3.084***	1.198***	1.067***	0.869***	0.506**	0.126
	(0.0916)	(0.290)	(0.0883)	(0.123)	(0.155)	(0.194)	(0.239)
Robust	1.152***	3.084***	1.198***	1.067***	0.869***	0.506*	0.126
	(0.123)	(0.387)	(0.119)	(0.155)	(0.194)	(0.239)	(0.285)
Fuzzy							
Second stage							
Conventional	3.063***	21.51***	3.042***	2.617***	2.244***	1.971***	1.655**
	(0.176)	(3.982)	(0.164)	(0.178)	(0.204)	(0.280)	(0.513)
Bias-corrected	2.522***	12.16***	2.549***	2.215***	1.991***	1.752***	1.312*
	(0.176)	(2.250)	(0.164)	(0.178)	(0.204)	(0.280)	(0.513)
Robust	2.522***	12.16***	2.549***	2.215***	1.991***	1.752***	1.312*
	(0.229)	(2.920)	(0.214)	(0.223)	(0.256)	(0.340)	(0.613)
Observations	1613	1613	1613	1613	1613	1613	1613

This table shows the results of RD design with the logarithm transformation of the price of Bitcoin s as the dependent variable. The cut-off point is the time when the Bitcoin reached its highest value in December 2017. The running (forcing) variable is daily entropy of news with topic Tax Investigation & Law. Each column presents a regression model with different kernel functions and degrees of polynomial function. The optimal bandwidth is obtained based on the Calonico et al. (2015) procedure. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level. Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 17: Robustness for Specifications of Bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
Running	Law & Tax	Law & Tax	Law & Tax	Law & Tax	Law & Tax	Law & Tax
Variable	Investigation	Investigation	Investigation	Investigation	Investigation	Investigation
Dependent						
Variable	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin	Ln.Price Bitcoin
Kernel	epanechnikov	epanechnikov	epanechnikov	epanechnikov	epanechnikov	epanechnikov
Bandwidth	h(0.01)	h(0.05)	h(0.1)	h(0.5)	h(5)	h(10)
Degree	P(1)	P(1)	P(1)	P(1)	P(1)	P(1)
Sharp						
Conventional	0.987***	3.127***	1.175***	1.186***	1.186***	1.186***
	(0.156)	(0.289)	(0.0844)	(0.0821)	(0.0820)	(0.0820)
Bias-corrected	0.458**	3.211***	1.275***	1.291***	1.292***	1.292***
	(0.156)	(0.297)	(0.0844)	(0.0821)	(0.0820)	(0.0820)
Robust	0.458	3.211***	1.275***	1.291***	1.292***	1.292***
	(0.242)	(0.401)	(0.111)	(0.109)	(0.109)	(0.109)
Fuzzy						
Second stage						
Conventional	2.288***	20.10***	3.023***	3.028***	3.028***	3.028***
	(0.219)	(3.356)	(0.151)	(0.146)	(0.146)	(0.146)
Bias-corrected	1.979***	11.83***	2.522***	2.536***	2.537***	2.537***
	(0.219)	(1.976)	(0.151)	(0.146)	(0.146)	(0.146)
Robust	1.979***	11.83***	2.522***	2.536***	2.537***	2.537***
	(0.337)	(2.599)	(0.192)	(0.188)	(0.188)	(0.188)
Observations	1613	1613	1613	1613	1613	1613

This table shows the results of RD design with the logarithm transformation of the price of Bitcoin s as the dependent variable. The cut-off point is the time when the Bitcoin reached its highest value in December 2017. The running (forcing) variable is daily entropy of news with topic Tax Investigation & Law. Each column presents a regression model with different bandwidths. Panel A presents the results of sharp RD design approach and Panel B presents the results of the two-stage fuzzy RD design approach. The time horizon is from January 2014 to December 2018. Robust standard errors clustered at time level. Note: standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$