

# Research on the Application of Cross-border E-commerce Big Data in the Prediction of U.S. Presidential Election

Haomeng Wang

School of International Trade and Economics, Anhui University of Finance and Economics,  
Bengbu Anhui 233030, China

## Abstract

The U.S. presidential election held every four years is the focus of worldwide attention, and predicting the outcome of the U.S. presidential election is a world-class problem. The emergence and application of Cross-border E-commerce big data has provided new methods and ways to predict the outcome of the U.S. presidential election. The Cross-border E-commerce big data of eBay, AliExpress and other platforms show that during the 2016 and 2020 U.S. presidential elections, Trump's campaign aids sales were far greater than those of rivals Hillary and Biden. Based on the sales volume, it is predicted that Trump will win the two presidential elections, but the actual result is one win and one loss. However, if considering that the 2020 U.S. presidential election was affected by the novel coronavirus pneumonia epidemic, there are significant differences in the election rallies of the two presidential candidates and the demand for campaign aids, the success rate of Cross-border E-commerce big data prediction should be 50% above. The 2024 U.S. presidential election is coming soon. Tracking the big data on the sales of campaign aids on Cross-border E-commerce platforms is a supplementary way that cannot be ignored in predicting the outcome of the campaign.

## Keywords

Cross-border E-commerce Big Data; Campaign Aids; U.S. Presidential Election Prediction.

## 1. Introduction

After the cold war, the United States is the only superpower in the world. In the political structure of the separation of powers, the president of the United States occupies an important position. The US president's political stance and policy agenda will not only affect the domestic political, economic and social trend of the United States, but also have a profound impact on international politics, economy, diplomacy and military affairs. Therefore, the quadrennial U.S. presidential election has become the focus of attention all over the world, and the prediction of the results of the U.S. presidential election has also become a hot topic in global academia, media and economic institutions.

Financial institutions represented by Wall Street claim that some economic indicators can accurately predict the results of the U.S. presidential election. These economic indicators mainly include: stock index, dollar index, manufacturing index, employment rate, economic growth rate, Chicago Option Exchange Volatility Index and so on. However, the prediction results of all the above economic indicators are not ideal. US market research or media organizations represented by Gallup, Fox and the New York Times try to predict the results of the U.S. presidential election through public opinion surveys. However, due to the political position of relevant market research or media organizations and the selection deviation of survey samples, the relevant prediction results are inevitably questioned in all aspects, and even suspected of deliberately misleading the domestic political trend. In short, the quadrennial

U.S. presidential election is not only the focus of the world, but also the prediction of the results of the U.S. presidential election is a world-class problem.

The development of big data, machine learning and artificial intelligence has challenged the statistical prediction methods and technologies of traditional incomplete samples, and greatly improved human prediction ability for the future in many aspects. With the development of information technology and Cross-border E-commerce, more and more consumers are used to online shopping. Therefore, Cross-border E-commerce platforms such as Amazon, eBay and AliExpress have precipitated a wealth of Cross-border E-commerce shopping big data, including the purchase data of U.S. presidential campaign aids. Obviously, buying T-shirts, baseball caps and other campaign aids from different presidential candidates represents the political position of consumers or supports the corresponding presidential candidates. Although the personal information of Cross-border E-commerce consumers cannot be made public (held by Cross-border E-commerce platform companies), the sales volume of aid during the presidential candidate's election campaign can be obtained. Therefore, by comparing the sales of the two presidential candidates' campaign aids on the Cross-border E-commerce platform, we can evaluate their support among voters, and then predict the presidential election results.

## 2. Literature Review

Public opinion survey is an important method and way to predict the U.S. presidential election. There are many polling institutions predicting the U.S. presidential election, such as Gallup, Pew Research Center, Fox and Emerson College. In addition, polling institutions in other countries such as the economist, Reuters and Ipsos also predict the U.S. presidential election. After long-term practice and development, these polling institutions have gradually formed a set of relatively mature polling methods, which have scientific survey norms from determining survey topics, formulating research plans, designing questionnaires, sampling, conducting actual surveys, processing survey data and writing research reports[1].

However, the disadvantage of this method is that it is easy to produce sample deviation. One of the reasons leading to sample deviation is objective technical problems and cost problems. How to ensure the consistency between sample distribution and overall distribution is always the core problem in statistical analysis and prediction. At present, the samples of institutions mainly come from three ways: random dialing (RDD), registration-based sampling (RBS) and Internet opt in [2]. Among them, RDD and RBS belong to probability sampling, in other words, random sampling[3]. Random sampling ensures that the selection probability of each sample is known, and the reciprocal of the selection probability is the weight of the sample. When estimating, the sample can be restored to the population [4], so the sample distribution can be consistent with the population distribution. However, in practice, even random sampling cannot be completely random, such as no response [5]. With fewer and fewer people responding to the public opinion survey and the higher and higher cost of the public opinion survey, researchers began to use the nonprobability sampling method. Opt in belongs to the nonprobability sampling. The respondents are usually not randomly selected into the sample, and their characteristics may be different from those who did not participate in the survey [3], so the sample distribution cannot be guaranteed to be consistent with the overall distribution. Another reason is subjective. The prediction subject itself has a political position. In order to guide the votes, it deliberately selects a biased sample. During the general election, some public opinion survey institutions carried out "biased public opinion survey" by using the mentality of "conformity" or "voting for the winner" of ordinary people, and adjusting the sample structure of respondents is one of the means [6]. M.V. Schönberger and others believe that once there is

any bias in the sampling process, the results will be very different [7], so the sample deviation will affect the prediction accuracy.

Big data prediction can avoid the sample deviation of poll prediction to a great extent. On the one hand, big data is based on mastering all data, at least as much data as possible [7], rather than sampling data. In the era of limited information processing capacity, sampling methods have been produced for data analysis [7]. However, with the popularity of the Internet, the Internet of things and social media, the operations on any computer, mobile device or sensor can be automatically recorded, stored or transmitted by the machine, and then big data is produced. It gives the impression that it can record anything anyone does[4], and the improvement of data processing and storage capacity, the progress of analytical technology and cheap data collection methods make it easier to collect these comprehensive and complete data[7]. In this case, there is no need for sampling or investigation [4]. Only certain technologies need to be used to collect these comprehensive data, so that big data prediction can avoid the sample deviation caused by sampling. On the other hand, the prediction subject of big data prediction is big data, which usually comes from online E-commerce transaction data, supermarket scanning and settlement data, online search engine data, social media data, etc. [8]. They record people's normal state, so there is no need to worry about the subjective bias of investigators in the investigation [7], Therefore, there will be no sample deviation caused by the intentional selection of biased samples by the prediction subject with political position. In short, big data prediction can improve the sample deviation and improve the prediction accuracy.

Compared with the poll forecast, big data forecast also has other advantages. First, big data can better reflect people's real psychology. The data source of big data describes people's digital footprint [9], which makes the prediction based on facts to a greater extent. People will hide their real ideas, but these data generated by people's actual behavior will not. While investigating the sample, the respondents choose not to respond or to lie for various reasons, which reduces the accuracy of prediction. Secondly, big data prediction is more immediate. Through the timely monitoring, collection and sorting of big data, the time difference between the information used for prediction and the content to be predicted can be reduced to a great extent[10], while the survey information is easy to lag and can not keep up with the changes of the situation.

In short, using the traditional public opinion survey method to predict the U.S. presidential election is easy to fail due to the sample deviation, while the big data prediction can make up for this defect to a great extent, and is better than the poll prediction in reflecting the authenticity and timeliness of the content. At present, the most commonly used data source in the literature on big data prediction of U.S. presidential election is social media big data on twitter[11,12], but few use Cross-border E-commerce big data. The innovation of this paper is to establish a link between Cross-border E-commerce sales and the results of the U.S. presidential election, and apply the new data source - Cross-border E-commerce big data - to the prediction of the U.S. presidential election. Therefore, it adopts the real sample prediction rather than the traditional sampling prediction, so as to explore a new way of the prediction of the U.S. presidential election.

### **3. Theoretical Mechanism of Cross-border E-commerce Big Data Predicting the Outcome of the U.S. Presidential Election**

With the continuous popularization of the Internet and the continuous progress of information technology, Cross-border E-commerce platforms such as Amazon, AliExpress and eBay have developed rapidly. All kinds of goods from all over the world are sold on these Cross-border E-commerce platforms, and more and more consumers are used to buying goods on Cross-border E-commerce platforms, The massive records of huge consumer groups trading goods on Cross-

border E-commerce platforms have formed Cross-border E-commerce big data, including the shopping big data of U.S. presidential campaign aid.

The big data of campaign aids shopping contains the sales information of campaign aids, such as hats, T-shirts and flags printed with the name, avatar and campaign slogan of the presidential candidate, which are applied by the presidential candidate's campaign team and supporters to the campaign rally and other activities during the U.S. presidential campaign. On the one hand, the large purchase of these aids is a symbol of the financial resources of the presidential candidate. The greater the amount of aids purchased, the stronger the candidate's financial resources, the more funds he raises, and the stronger the strength of his supporters. Moreover, these funds usually come from some people who have important influence on all aspects of American society, who have a strong help for the candidate to win the election. Data show that most of the recent U.S. presidential elections were won by the candidates with the largest amount of fundraising[13]. Although the financial advantage of candidates is sometimes not necessarily equal to the election advantage, in most cases, money will have a decisive impact on the election results, especially in those close elections, the amount of money often has a direct impact on the election results [14]. On the other hand, the sales volume of aid also represents people's support for candidates. In public opinion polls, respondents may choose to hide their true attitude or even lie for a variety of reasons, but when it comes to personal money and interests, people will be more faithful to their true thoughts. Therefore, people will buy the aid of candidates they support to help them. This is just like the truth contained in the explicit preference theory in microeconomics. The behavior of people buying campaign aids itself has shown their real preference, that is, which candidate they support. Therefore, if a candidate sells more aids, it means that he has more support from the people, and more people are willing to vote for him and appeal to others to support the candidate through various channels. In the era of highly developed Internet, the speed and scope of information dissemination are large, and huge and complex relationship network has been established between people. This influence force can not be underestimated.

To sum up, the sales volume of the campaign aids can be used as a symbol of the candidate's financial resources and a reflection of the people's real attitude to predict the results of the presidential election, that is, the more the sales volume of aid, the more likely the candidate is to win the election. The rapid development of data mining technology makes it easier to obtain Cross-border E-commerce big data, so the Cross-border E-commerce big data reflecting the sales of aid can predict the U.S. presidential election. Therefore, through data mining, we can get the shopping big data of American presidential candidate's aid on the Cross-border E-commerce platform, and then compare the sales of these aids, we can predict the results of the U.S. presidential election.

## **4. Empirical Analysis of Cross-border E-commerce Big Data Predicting the Results of the U.S. Presidential Election**

### **4.1. Data Source and Description**

The original data of this paper is obtained by using Octopus collector version 8.2.2.111911 to crawl the data of Cross-border E-commerce platforms. Amazon, eBay and AliExpress, as the three major Cross-border E-commerce platforms in the world, have huge traffic. In particular, both Amazon and eBay are Cross-border E-commerce platforms in the United States, with more users in the United States. However, because there is no public commodity sales information on Amazon's web page, they are not available legal information, Therefore, the commodity sales data on this platform cannot be crawled. Therefore, this paper only uses the data crawled from eBay and express, which are the big Cross-border E-commerce shopping data of the aids of the U.S. presidential election candidates in 2016 and 2020 on these two platforms. Among them,

the presidential candidates in 2016 are Trump and Hillary Clinton, and the presidential candidates in 2020 are Trump and Biden. The aid is divided into three categories: hat, T-shirt and flag. Searching AliExpress (<https://www.aliexpress.com/>) and eBay (<https://www.eBay.ca/>) respectively in the browser and opening the website of the Cross-border E-commerce platforms, then entering keywords in the search box of the website, you can search the aids by category, and get the website corresponding to each keyword(the keywords are shown in Table 1). In this way, each type of campaign aid can be screened out and their sales data can be crawled.

**Table 1.** Keywords

2016 trump hat	2016 trump shirt	2016 trump flag	2016 hillary hat	2016 hillary shirt	2016 hillary flag
2020 trump hat	2020 trump shirt	2020 trump flag	2020 biden hat	2020 biden shirt	2020 biden flag

### 4.2. Results

As shown in Table 2 and table 3, the sales of hats, T-shirts and flags of the two presidential candidates during the 2016 and 2020 U.S. presidential elections are crawled. It can be seen that Trump's sales of hats, T-shirts and flags during the two elections are more than that of the other candidate.

As the more aid sales, the more funds the candidates raise and the more supporters they have, the more likely they are to win the election. Based on this, the prediction of Cross-border E-commerce big data on the results of the U.S. presidential election is made: trump will win the election in 2016 and 2020. Compared with the actual election results, the 2016 election forecast is successful, while the 2020 election forecast is failed.

**Table 2.** Sales of U.S. presidential campaign aids in 2016

		Hat	T-shirt	Flag
AliExpress	Trump	3	7	4
	Hillary	0	0	0
eBay	Trump	11377	2021	2928
	Hillary	6	187	472

**Table 3.** Sales of U.S. presidential campaign aids in 2020

		Hat	T-shirt	Flag
AliExpress	Trump	2819	16	1103
	Biden	626	10	2
eBay	Trump	27611	9974	29106
	Biden	358	3198	354

### 5. Conclusion

In this paper, octopus collector version 8.2.2.111911 is used to crawl the Cross-border E-commerce big data of the candidates' campaign aids during the 2016 and 2020 U.S. presidential elections on the Cross-border E-commerce platform AliExpress and eBay. The raw data crawled are processed and summarized to obtain the sales of various kinds of campaign aids, compare the relative sales of all kinds of aids of the candidates, and combine the theoretical mechanism of the Cross-border E-commerce big data in predicting the U.S. presidential election, Trump was found to have won both presidential elections. However, by comparing with the actual election results, it is found that Cross-border E-commerce big data can successfully predict the results of the U.S. presidential election in 2016, but can not successfully predict the election results in

2020. Therefore, using Cross-border E-commerce big data to predict the U.S. presidential election is accidental and can not be used as a completely reliable prediction method.

During the two presidential campaigns, there were more campaign aids of Trump sold on eBay and AliExpress. However, Trump only won the election in 2016. This may be because the Amazon platform also sells the aids of the candidates. However, we cannot obtain the sales information of the Amazon platform. Otherwise, we can sum up the sales of the three Cross-border E-commerce platforms for comparative analysis. In addition, the 2020 U.S. presidential election is an election held after the outbreak of the COVID-19 in 2019. Against the background of the epidemic, trump is more keen than Biden to hold campaign rallies, and the epidemic prevention requirements, especially the restrictions on the number of participants in the rallies, are also more relaxed, which may also cause the difference in the demand for campaign aids between the both sides. In fact, the U.S. presidential campaign is affected by many factors, and the traditional poll prediction and emerging big data prediction can not take all factors into account. Therefore, other factors affecting the election results, especially unexpected factors, will interfere with the prediction results. Moreover, the election process is long and complex, and any situation during the period may make the election situation change sharply, which makes the prediction of the election results more difficult.

Although Cross-border E-commerce big data cannot be used as a reliable prediction method for the U.S. presidential election, Cross-border E-commerce big data itself has some advantages that public opinion polls lack: it can avoid sample deviation to a certain extent, its cost is lower, it is easier to reflect people's real ideas, it can timely reflect changes in the situation, etc. Moreover, the U.S. presidential election is affected by many factors, so it is not easy to predict the election results, Therefore, the prediction function of Cross-border E-commerce big data cannot be completely denied. It can be used as a supplementary way to predict the future U.S. presidential election and expand the prediction method of U.S. presidential election.

## Acknowledgments

Postgraduate research and innovation fund project of Anhui University of Finance and Economics (ACYC2020086).

## References

- [1] H. J. Dong, J.J. Tang: Analysis on the history and current situation of foreign public opinion polls, Learning and Practice, (2012) No.2, P.103-108.
- [2] C. Kennedy, M. Blumenthal, S. Clement, et al: An Evaluation of the 2016 Election Polls in the United States, Public Opinion Quarterly, Vol.82 (2018) No.1, p.1-37.
- [3] Information on: [https:// www. aapor. org/ Education-Resources/ Election-Polling-Resources/ Sampling-Methods-for-Political-Polling.aspx](https://www.aapor.org/Education-Resources/Election-Polling-Resources/Sampling-Methods-for-Political-Polling.aspx).
- [4] L.Y. Ren: Survey Research Methods in American Political Studies, American Studies, Vol.34 (2020) No.3, p.84-106+7.
- [5] Information on: <http://www.ncpp.org/?q=node/4>.
- [6] L.L. Yang: From the general failure of US election opinion polls, China Statistics, (2017) No.2, p.23-26.
- [7] M. V. Schönberger, K. Cukier: Big Data (Harcourt Publishing Company, U.S. 2013).
- [8] Q. He: Big data prediction, China Statistics, (2016) No.3, p.18-20.
- [9] T. Breur: US elections: How could predictions be so wrong?, Journal of Marketing Analytics, Vol.4 (2016) No.4, p.125-134.
- [10] Y. L. Wang: Big data prediction and economic prediction in the era of big data, Academics, (2016) No. 11, p.77-90+325.

- [11] B. Heredia, J.D. Prusa, T.M. khoshgoftaar: Social media for polling and predicting United States election outcome, *Social Network Analysis and Mining*, Vol.8(2018), No.1.
- [12] Z. Y. Gong, T.T. Cai, T.J. Claude, et al: Measuring relative opinion from location-based social media: A case study of the 2016 U.S. presidential election, *Plos One*, Vol.15(2020), No.5.
- [13] B.X.Gao, W.R.Li: Money erosion: a big problem to test American democracy, *Jianghai Tribune*, (2016) No. 09, P.56-57.
- [14] C.N.Ni: Political contributions and American election politics, *Contemporary World and Socialism*, (2018) No.5, p.134-142.