Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2019 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

6-15-2019

Fake It Till You Make It: An Empirical Investigation of Sales Fraud in E-commerce

Le Wang
Nanyang Technological University, lwang033@e.ntu.edu.sg

Jiahui Mo
Nanyang Technological University, jhmo@ntu.edu.sg

Beibei Li
Carnegie Mellon University, beibeili@andrew.cmu.edu

Fake It Till You Make It: An Empirical Investigation of Sales Fraud in E-commerce

Research-in-Progress

Le Wang
Nanyang Technological University
50 Nanyang Avenue, Singapore
lwang033@e.ntu.edu.sg

Jiahui Mo
Nanyang Technological University
50 Nanyang Avenue, Singapore
jhmo@ntu.edu.sg

Beibei Li
Carnegie Mellon University
4800 Forbes Avenue, Pittsburgh, USA
beibeili@andrew.cmu.edu

Abstract

The competition on e-commerce platforms has become more and more fierce. Among all the different promotion strategies, sales fraud, which is a practice inflating sale volume by using fictitious transactions, is an open secret among e-commerce sellers. Sales fraud will fundamentally undermine the credibility of sales volume, which is one of the major information source for decision making in online purchasing. To shed light on this phenomenon, we empirically investigate circumstance under which sales fraud will take place, using a comprehensive dataset from a mainstream e-commerce website in China. We find that sales cheating is more likely to take place for those products with lower price, from lower-level shops, in their early stages, but with good sales potential. Our empirical findings provide important contributions to the literature on e-commerce, and offer critical managerial implications to online retailers, e-commerce platforms, and consumers.

Keywords: Sales cheating, Online fraud, Sales volume, Electronic commerce

Introduction

The advancement of e-commerce has made it easier for a wide range of retailers to market all over the world (Grandon & Pearson, 2004; Jarvenpaa et al., 2000). According to a report by eMarketer, in 2017, the worldwide e-commerce sales continue to increase at 23.2% to $2.290 trillion, accounting for more than one-tenth of the total retail sales worldwide. On e-commerce platforms, consumers can get easy access to prior consumers’ purchases through the information of previous sales volume, which is a major factor influencing consumer purchasing decision (Cai et al., 2009; Chen & Xie, 2008).

However, sellers can manipulate the sales information by fraud sales. To be specific, the common practice for fraud sales is as follows: first, the vendor finds cheaters and pays them the cost of the products they need to cheat on and an additional amount of fee as an award; second, the cheaters place orders for the products and pay with the money they got from the vendor in advance; third, the vendor delivers parcels that are empty but with a tracking number to outtrick the platform; lastly, the cheaters confirm the receipt of the products.

Although websites have their filtering algorithm to identify suspicious orders and to punish involved sellers, they do not deduct fake orders from the displayed sales volume. That is, when sales cheating takes place, sales volume information will not reflect the real purchasing from previous consumers. Although illegally and explicitly forbidden by most e-commerce platforms, this type of sales cheating is pervasive in e-commerce. For example, the Wall Street Journal reported that 17% of all merchants on Taobao, the largest C2C e-commerce website in China, had faked 500 million transactions in 2013 (Wong et al., 2015).

As sale fraud is getting more and more prevalent in e-commerce, the fraud sales may cause some damage to consumers, sellers, and platforms. For example, sales fraud will mislead consumers because ordinary people do not directly observe whether an order is fake or not. Moreover, sellers who commit the sales cheating have the risk of losing reputation and face the potential of punishment. However, many sellers still insist on doing cheating behaviors.

Little research, however, has been conducted to study the nature of this type of online fraud. To fill in the research gap, we would like to explore when and why sellers cheat on certain products, and specifically what are the characteristics of products that will make sellers choose them to commit sales fraud. Answering these questions will contribute to the literature in several ways. First, to our knowledge, this study will constitute the first effort to study the online sales fraud behavior. Second, our results may help deepen the understanding of fraud motivation in the online purchase environment; thereby we could provide several insights on how to avoid such kind of fraud behaviour in this environment.

In this study, we answer the above research questions via a rich panel dataset obtained from a famous e-commerce website in China on 317,494 products from May 1st 2017 to July 31st 2017. We identify some factors that will influence sales cheating engagement. Specifically, our results show that, on average, sales cheating is more likely to take place for those products with lower price, from lower-level shops, in their early stage, but with good sales potential. Our additional investigations of the contingent factors find that the impact of product-level sales potential on sales cheating engagement would be stronger with longer on shelf time, but the impact of shop-level characteristics on fraud engagement will be weaker if products are on shelf for a longer time.

Related Literature

The Internet economy has made it easier for consumers to get access to various information, e.g. other customers' purchasing and crowd opinion, which cannot be obtained easily from offline channels. However, due to the inherent characteristics of online context, online fraud is ubiquitous in e-commerce. For example, a recent stream of studies have begun to pay attention to the reliability of online reviews, such as objectivity (Goes et al., 2014), and some of them have concerned the problem of fake reviews (Lappas et al., 2016; Luca & Zervas, 2016). Luca and Zervas (2016)) found that nearly 16% of restaurant reviews are fake in Yelp.com, and fake reviews are more likely to take place when a restaurant has a low reputation. Using hotel data across 17 cities, Lappas et al. (2016)) found that even limited injections of fake reviews can have a significant effect on online visibility of hotels on online hotel websites.

While extensive research effort has been devoted to examining the role of sales volume (Chen et al., 2011; Hanson & Putler, 1996; Zhang, 2010), most of them focus on its positive effect in assisting decision making. Another stream of research has concerned the possible negative effects of sales volume. For example, Salganik et al. (2006)) argued that sales volume can be unreliable as it may be caused by randomness from early adopters of products. Other researchers also stated that previous sales volume can be misleading as it simply reflects other consumers' actions but fails to include the reasons behind their decisions (Bikhchandani et al., 1992, 1998).

However, no research has paid attention to the sales fraud behavior. Given that sales cheating has become increasingly common on many e-commerce platforms and served as a new strategy to boost sales by modifying information presented to consumers, examining the actual effect of this online fraud behavior becomes necessary, and we try to fill in this research gap in this study.

Data

To empirically address the above research questions, we obtain data from a mainstream e-commerce website in China. This website is an independent e-commerce platform that facilitates the transactions between individual retailers/stores and consumers. There are numerous online stores on this platform, and each store itself can decide what products to sell. While sales cheating is forbidden, it is ubiquitous for almost every store on the platform. We randomly choose 2973 stores, and found that 2965 (99.73%) of them have sales cheating records. With sophisticated models and algorithm, the platform can follow each order and decide whether it is a cheating order, but only when a certain amount of cheating was caught would the seller be punished. What's more, when an order is confirmed as a cheating order, its review channel will be closed but this order will still be added to the product's sales volume. That is to say, cheating in this platform will only change product sales volume information without changing product ratings.

We obtained 2,593,868 observations for 317,494 products from May 1st, 2017 to July 31st, 2017. The information covers displayed product information (e.g. price, Detail Seller Rating etc.), transaction information. And the platform also provide us the cheating information for each product (i.e. cheating volume in each week). This thus provides us a convenient setting to observe and estimate factors that will influence one product to conduct sales cheating in one week but not the other product in another week.

Model and Analysis

Empirical Model

Based on this panel-level dataset, we conduct our analysis at the product-week. Let subscript *i* denote each individual product in our dataset, and subscript *t* denote each week. To investigate factors that will influence products' likelihood of engaging in sales cheating, our dependent variable, $isCheating_{it}$, is a binary indicator for sales cheating engagement. That is, $isCheating_{it} = 1$ if product *i* has engaged in product cheating on week *t*, zero otherwise. Furthermore, we include several control variables and three sets of independent variables in our empirical models. The meanings of the main independent variables and control variables can be found in Table 1.

Table 1 Variable Description

Variables	Description
Control variables (CV)	
LC	Whether the item has fraud sales in last week
PB	The price of the item in the beginning of this week
TB	The number of days since the item was first put on shelf
SL	The level of the shop in the beginning of this week
PV	The page view of the item in last week (log)
Lag1-factors (LF)	
LSI	The sales of the item in last week (log)
LSS	The sales of the shop in last week (log)
LRI	The average displayed ratings of the item in last week
LRS	The average displayed ratings of the shop in last week
Historical factors (HF)	
HSI	The historical sales of the item
HSS	The historical sales of the shop
HRI	The historical average ratings the item obtains
HRS	The historical average ratings the shop obtains

Incremental factors (IF)	
RII	(The item displayed ratings in the beginning of the week-displayed ratings in the beginning of last week)/ displayed ratings in the beginning of last week
RIS	(The shop displayed ratings in the beginning of the week-displayed ratings in the beginning of last week)/ displayed ratings in the beginning of last week
SII	(The item sales in last week t-average item sales by last week)/ average item sales by last week
SIS	(The shop sales in last week t- average shop sales by last week)/ average shop sales by last week

To address our research questions, we model the three sets of factors on sales cheating engagement separately. The panel-level linear model is specified in Equation (1)-(4):

$$isCheating_{it}= \beta_0+\beta_1 CV_{it}+m_t+\alpha_i+\varepsilon_{it} \quad (1)$$

$$isCheating_{it}= \beta_0+\beta_1 CV_{it}+ \beta_2 LF_{it}+m_t + \alpha_i+\varepsilon_{it} \quad (2)$$

$$isCheating_{it}= \beta_0+\beta_1 CV_{it}+ \beta_2 HF_{it}+m_t + \alpha_i+\varepsilon_{it} \quad (3)$$

$$isCheating_{it}= \beta_0+ \beta_1 CV_{it}+ \beta_2 IF_{it}+m_t+ \alpha_i+\varepsilon_{it} \quad (4)$$

where α_i captures unobserved product-specific effect and m_t denotes the month dummy variables.

Results

We first estimate a logit model of whether the product engages in sales cheating on all control variables. As reported in table 2, Column (1), various control variables have significant relationships with sales cheating engagement. Specifically, sales cheating is a strategy with continuity, thus sales cheating in the past week is positively related to the likelihood of sales cheating engagement in this week. Next, product price has a negative relationship with cheating engagement. This might be because as the price goes up, sellers need to pay more for hiring cheaters as they need to pay cheaters the cost of the products in advance. Product on shelf time is negatively related to cheating engagement. This is consistent with sellers' original intention to conduct sales cheating: attracting buyers in an early stage of products. Higher-level shops are less likely to engage in cheating as they have more other means in boosting sales and cheating may do more harm than good to them (e.g. ruining their reputation). Moreover, comparing with those products with little page view, products with high page view are more likely involved in sales cheating. Overall, these significant relationships imply that our control variables have good explanatory power.

Beyond the control variables, we then estimate other factors by including three more sets of other independent variables. We summarize the results in Table 2, Column (2)-(4). As indicated in Column (2), the coefficient of *LSI*, 0.763 ($p < 0.001$), is positive and statistically significant, suggesting a positive relationship with cheating engagement. On the contrary, the coefficient of *LSS* (β=-0.0870, $p < 0.001$) suggests a negative relationship with cheating engagement. Furthermore, both *LRI* (β=0.0708, $p < 0.001$) and *LRS* (β=0.125, $p < 0.001$) have a positive relationship with cheating engagement. The results are consistent if we use historical indicators instead of one-week-lag indicators, thus we may interpret the results altogether as follows: sales cheating is more likely to take place for products that have potential, i.e., a good past sales or a good rating accumulated from past selling, however, sales cheating may be just an expedient strategy and when the overall sales of a shop is large enough it will avoid conducting sales cheating. This is consistent with the effect of shop level, which is highly correlated with overall sales. These effects are further consolidated by including incremental variables. As indicated in Table 2, Column (4), the coefficients of *RII* (β=0.165, $p < 0.001$) and *SIS* (β=3.311, $p < 0.01$) are both significantly positive, suggesting that an increasing in product rating or sales, even a transitory one, will incur more sales cheating, and the larger the increasing the stronger the effects on sales cheating engagement. However, the coefficient of *RIS* (β=-

0.194, $p < 0.001$) is significantly negative, suggesting that when the overall ratings of a shop increase, the products in this shop are less likely to be involved in sales fraud.

Table 2 Results

Variables	(1) Logit Control	(2) Logit Lag1-effect	(3) Logit Historical-effect	(4) Logit Incremental-effect
LC	1.178***	0.518***	1.119***	1.239***
	(0.00802)	(0.0106)	(0.00943)	(0.0117)
PB	-0.000539***	-0.000757***	-0.00396***	-0.00456***
	(0.0000663)	(0.000113)	(0.000113)	(0.000137)
TB	-0.00860***	-0.00129***	-0.00749***	-0.00614***
	(0.000215)	(0.000273)	(0.000286)	(0.000326)
SL	-0.103***	-0.0176***	-0.108***	-0.0972***
	(0.00319)	(0.00510)	(0.00427)	(0.00497)
PV	0.663***	0.185***	0.677***	0.722***
	(0.00220)	(0.00479)	(0.00286)	(0.00322)
LSI		0.763***		
		(0.00625)		
LSS		-0.0870***		
		(0.00450)		
LRI		0.0708***		
		(0.00934)		
LRS		0.125***		
		(0.0294)		
HSI			0.000589***	
			(0.0000216)	
HSS			-0.00000138*	
			(0.000000543)	
HRI			0.154***	
			(0.00755)	
HRS			0.201***	
			(0.00781)	
RII				0.165**
				(0.0578)
RIS				-0.194**
				(0.0709)
SII				0.0627
				(0.0525)
SIS				3.311***
				(0.491)
constant	-4.047***	-2.475***	-5.378***	-4.022***
	(0.0139)	(0.134)	(0.0485)	(0.0278)
time dummies	-included-	-included-	-included-	-included-
N	1911186	540035	1146523	882212
pseudo R^2	0.298	0.287	0.350	0.379

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

Contingent Factors: Product On-shelf Time

After identifying some factors that will influence sales engagement, we further explore potential contingent factors which may moderate the identified relationships. We expect that the factors may exert different influence basing on the stages of the product. To empirically analyse these moderating effects, we construct and include the interaction terms in our model estimation. As show in Table 3, the estimates of *TB*LSI*, *TB*LSS*, *TB*LRI* are all positive and significant. These results thus indicate that as the product is at its later stage, the effect of product sales and displayed ratings in the past week on sales cheating engagement will be stronger, while the effect of shop sales in the past week will be weaker. That is, when an product is on shelf for longer time, shop level characteristics will exert weaker influence, but if the product still has good sales potential, it is very likely to be involved in sales cheating. We get similar results if we use historical indicators, product on shelf time also negatively moderates the relationship between sales cheating engagement and historical shop sales and ratings. However, we get opposite results for *HSI* and *HRI*: product on shelf time negatively moderates the relationship between sales cheating engagement and historical product sales but positively moderates the relationship between sales cheating engagement and historical product ratings. We believe the reason behind is as follows: comparing with historical ratings, historical sales is a weaker indicator of product potential, thus, as more days a product is on shelf, the positive effect of past sales will be weaker, but if the product still has high ratings, indicating a long-lasting potential, the incurring effect will be even stronger.

Table 3 Interaction Effects

Variables	(1) Logit control	(1) Logit Lag1-effect	(2) Logit Historical-effect	(3) Logit Incremental-effect
LC	1.179***	0.517***	1.115***	1.239***
	(0.00798)	(0.0106)	(0.00948)	(0.0117)
PB	-0.000482***	-0.000711***	-0.00397***	-0.00456***
	(0.0000625)	(0.000113)	(0.000113)	(0.000137)
TB	-0.0281***	-0.00372	-0.00119	-0.00608***
	(0.000504)	(0.00707)	(0.00272)	(0.000328)
SL	-0.101***	-0.0159**	-0.103***	-0.0973***
	(0.00318)	(0.00514)	(0.00431)	(0.00497)
PV	0.522***	0.183***	0.678***	0.722***
	(0.00377)	(0.00480)	(0.00291)	(0.00322)
*TB*PV*	0.00495***			
	(0.000112)			
LSI		0.695***		
		(0.00834)		
LSS		-0.124***		
		(0.00774)		
LRI		0.0290		
		(0.0162)		
LRS		0.157***		
		(0.0435)		
TB LSI*		0.00230***		
		(0.000193)		
TB LSS*		0.00129***		
		(0.000214)		
TB LRI*		0.00180**		
		(0.000590)		
TB LRS*		-0.00278		
		(0.00157)		
HSI			0.000893***	
			(0.0000605)	
HSS			-0.0000126***	
			(0.00000140)	
HRI			0.0795***	
			(0.0138)	
HRS			0.284***	
			(0.0120)	
TB HSI*			-0.00000594***	
			(0.00000105)	
TB HSS*			0.000000219***	
			(2.47e-08)	

TB HRI*			0.00272^{***}	
			(0.000436)	
TB HRS*			-0.00452^{***}	
			(0.000459)	
RII				-0.0875
				(0.125)
RIS				-0.0900
				(0.165)
SII				0.168
				(0.115)
SIS				4.470^{***}
				(1.085)
TB RII*				0.00917^{*}
				(0.00398)
TB RIS*				-0.00289
				(0.00419)
TB SII*				-0.00398
				(0.00391)
TB SIS*				-0.0328
				(0.0270)
constant	-3.533^{***}	-2.234^{***}	-5.373^{***}	-4.023^{***}
	(0.0176)	(0.204)	(0.0807)	(0.0279)
time dummies	-included-	-included-	-included-	-included-
N	1911186	540035	1146523	882212
pseudo R^2	0.300	0.287	0.350	0.379

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

We further corroborate our findings by checking the robustness and consistency in multiple ways. Most of findings remain consistent.

Discussion and Conclusion

Our research object was to shed light on sales fraud, a phenomenon ignored by extant research, by identifying when and why sellers commit sales cheating in e-commerce, specifically the characteristics of products that will make sellers choose them to commit sales fraud. Using a rich dataset from a large e-commerce website, we identify several factors that will influence product sales cheating engagement. Our results show that, on average, sellers are more likely to commit sales cheating on those products with lower price, from lower-level shops, in their early stages, but with good sales potential. Our additional investigations of the contingent factors find that in a later stage of a product, the impact of product-level sales potential on sales cheating engagement would be stronger, but the impact of shop-level characteristics on fraud engagement will be weaker.

Our findings make several contributions to the literature on e-commerce. This is one of the first studies to investigate sales cheating behavior. It also helps understand the nature of this kind of cheating behavior. From a practical perspective, our study reminds consumers not to rely too much on sales volume when making purchasing decisions, but to take other information (e.g. product review) into consideration. Our results provide some insights why shops want to commit sales fraud behavior. For example, they may want to boost sales and promote some products with high potential. Platforms should consider various ways for shops to promote their high-potential products; thereby shops may not rely on sales fraud to boost sales and promote products.

References

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political economy*, *100*(5), 992-1026.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of economic perspectives*, *12*(3), 151-170.

Cai, H., Chen, Y., & Fang, H. (2009). Observational learning: Evidence from a randomized natural field experiment. *American Economic Review*, *99*(3), 864-882.

Chen, Y., Wang, Q., & Xie, J. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of marketing research, 48*(2), 238-254.

Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science, 54*(3), 477-491.

Goes, P. B., Lin, M., & Au Yeung, C.-m. (2014). “Popularity effect” in user-generated content: Evidence from online product reviews. *Information Systems Research, 25*(2), 222-238.

Grandon, E. E., & Pearson, J. M. (2004). Electronic commerce adoption: an empirical study of small and medium US businesses. *Information & management, 42*(1), 197-216.

Hanson, W. A., & Putler, D. S. (1996). Hits and misses: Herd behavior and online product popularity. *Marketing letters, 7*(4), 297-305.

Jarvenpaa, S. L., Tractinsky, N., & Vitale, M. (2000). Consumer trust in an Internet store. *Information technology and management, 1*(1-2), 45-71.

Lappas, T., Sabnis, G., & Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research, 27*(4), 940-961.

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science, 62*(12), 3412-3427.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science, 311*(5762), 854-856.

Wong, G., Chu, K., & Osawa, J. (2015). Inside Alibaba, the Sharp-Elbowed World of Chinese E-Commerce. Retrieved from https://www.wsj.com/articles/inside-alibaba-the-sharp-elbowed-world-of-chinese-e-commerce-1425332447

Zhang, J. (2010). The sound of silence: Observational learning in the US kidney market. *Marketing Science, 29*(2), 315-335.