Polarity, subjectivity and emotionality of feedback texts: the influence on sales and price premiums in online cryptomarkets

Thesis project Sociology

Faculty of Social and Behavioural Sciences

Student: Etienne Valk

6285953

Supervisor: Ana Macanovic

14-06-2021

Abstract

Online reputation systems serve to overcome the problem of cooperation in online market exchanges, which arises as a consequence of information asymmetries between sellers and buyers. In principle, buyers do not know whether sellers are trustworthy and will ship a product after a buyer has paid. Online reputation systems enable buyers to inquire information on prior behaviour of the seller through reading experiences of other market participants, either expressed in numerical ratings or feedback texts. Buyers can then determine the seller's reputation and whether the seller is trustworthy enough to engage with in economic exchange. These issues of trust are even more applicable for illegal (crypto)markets, as these lack legal, organizational, social and moral guarantees legal markets tend to have for stimulating trust and cooperation. It is yet unknown to what extent the feedback text characteristics polarity, subjectivity and emotionality are of impact on the success of sellers in online illegal cryptomarkets. In our study we show that in cryptomarket AlphaBay, feedback texts with high subjectivity received by a seller leads to more sales of a seller, yet not to more price premiums, while texts with high polarity and emotionality have no effect on either measure of success. While the numerical ratings also show no effect, the reputation measures constructed by the market administrators do prove to be of effect on the sales of a seller. This could spark interest for future research to take into account the institutional measures more, additional to the user-based qualitative and quantitative feedback measures. Our findings contribute to the knowledge of online reputation systems, by uncovering new ground through focusing on the direct and moderating effects of text characteristics of feedback texts on the success of a seller.

Keywords: online reputation systems, feedback texts, cooperation, cryptomarkets, trust, polarity, subjectivity, emotionality, sales, price premiums, ratings

Introduction

Trust can encourage cooperation in many societal interactions; whether it be in exchange between citizens and politicians, civilians and police or consumers and businesses, to name a few (Cook, 2001). Many of these relations are at crisis in the current day when it comes to mutual trust (Bozic, 2017; Murphy, 2017; Van der Meer, 2017). When trust is reciprocal between persons, cooperation is facilitated between them and constructive behaviour, with the potential of being beneficial for both, is fostered instead of mere self-preserving egoism

(Resnick et al., 2000). This raises the question how reciprocal trust can be attained and then, maintained.

A notable interaction where the problem of cooperation arises but is often overcome, is economic exchange. For centuries, reputation systems have been used in social groups to solve these trust problems in economic exchange (Greif, 1989; Hillmann & Aven, 2011; Przepiorka et al., 2017). In its simplest form, reputation systems encompass the circulation of information in a social network about good and/or bad past behaviour of market participants (Macanovic & Przepiorka, 2021). Access to such information enables other market participants to make a grounded decision whether or not to do business with them. Once market participants gain a good reputation, they will be eager not to lose it, as it can cost plenty of time and money to build, and once built, it can yield more profit. Thus, in order to maintain their positive reputation, they are more likely to behave cooperatively and not abuse trust others put in them (Przepiorka et al, 2017).

In our study, we focus on reputation systems of online marketplaces. In these reputation systems, people can leave numerical ratings evaluating a product and/or a seller, which can be accompanied by textual feedback (popularly known as a review). In the current age, people use these online reviews on a daily basis for the most banal (e.g. game apps) to the most essential (e.g. general practitioners) products and services; and increasingly so. A 2020 survey by BrightLocal has shown that in 2019, 81% of consumers read reviews of local businesses, while in 2020 this increased to 87%. With the internet and online reputation systems increasingly being part of our everyday lives and interaction, insights in the workings of these reputation systems is of growing importance (Király et al., 2020; Nowland et al., 2018; Ryan & Lewis, 2017).

An especially interesting, yet underexposed area of research within the literature on online reputation systems, is that of cryptomarkets. Cryptomarkets are online markets that function in the Darknet, where people anonymously exchange illegal goods and services (Chertoff, 2017; Macanovic & Przepiorka, 2021). In cryptomarkets, participants become untraceable through encryption. With this anonymity comes the need for a way to determine whether the other trader is trustworthy. Although problems of trust can also arise in regular (online) markets due to anonymity, it plays an even larger role within illegal cryptomarkets. These markets, after all, lack legal, organizational, social and moral guarantees legal markets tend to have for stimulating trust and cooperation (Diekmann & Przepiorka, 2019; Milgrom et al., 1990). It is therefore of great interest how the problem of trust is still overcome, and cooperation comes about in this specific context of illegal cryptomarkets, with a seemingly

even stronger reliance on reputation systems. We study to what extent textual feedback is of impact on the success of a seller in online illegal cryptomarkets.

In online reputation systems, ratings and feedback texts are both important indicators of the reputation of a seller. Elaborating on the general reputation mechanism for economic exchange, it becomes clear how online reputation systems are responsible for the effect both indicators have on the amount of sales and price premiums of a seller. Price premiums can be defined as the highest – or a higher – demanded price for a product, compared to identical or similar products in a relevant market. When sellers have recently entered an online market, they have yet to build a reputation. For buyers, the risk is higher to trade with them, as these new sellers could potentially deliver a product of lesser quality on purpose, or even just collect the money and leave the market. Posing much less of a threat to buyers, are sellers who have a well-established good reputation as a result of an endless amount of successful transactions. Moreover, these sellers are much less inclined to abuse trust, as losing their good reputation is likely to cost more than it yields. Buyers would thus generally prefer to trade with these lauded sellers and are even prepared to pay a little extra for a product in exchange for the security that comes with the transaction. Sellers who have not been able to build a reputation yet, are therefore forced to lower their prices in order to make themselves more attractive trading partners (Friedman & Resnick, 2001). It has been well established that in legal markets higher numerical ratings enable a seller to sell more and for better prices (Tadelis, 2016; Ye, Law et al., 2009). We will also take into account the effect of numerical ratings on the success of a seller, to ascertain how it relates to the effect of textual feedback.

Studying the influence of feedback texts is important, as information about the product, seller and/or overall experience can be transmitted to other market participants in more detail through text than through numerical ratings. Prior study has indeed shown that positive feedback texts lead to higher price premiums for sellers (Pavlou & Dimoka, 2006). Also, in legal markets, higher subjectivity of feedback texts has also proven to lead to more sales (Ghose & Ipeirotis, 2011). In our context, subjectivity is a possible characteristic of a feedback text and refers to texts or parts of texts which express the opinion, personal experience or feelings of the buyer. Oppositely, objectivity of a text relates to the factual information of e.g. a product, service or delivery which buyers aim to convey without interference of their biases (Liu et al, 2018). Another relevant characteristic of a feedback text is polarity, also known as valence. Polarity relates to how positive or negative the content of a text is, and has proven to be of influence on the success of sellers (Hu et al., 2014; Wang et al., 2017). Furthermore, a study of over 280 million online reviews has shown that the

polarity of texts and numerical ratings don't (necessarily) correspond; the polarity of texts is less positive compared to the polarity of numerical ratings (Schoenmueller et al., 2020). Lastly, emotionality, which in our context, relates to the emotion and feelings of a buyer expressed in text. Generally, emotions are pivotal for causing trust or distrust – happiness and gratitude tend to increase trust, whereas anger decreases trust (Dunn & Schweitzer, 2005). It therefore is essential to investigate how emotionality – and the other two text characteristics – influence the reputation and the success of a seller, expressed in amount of sales and price premiums.

Additionally, these three factors of polarity, subjectivity and emotionality possibly have a moderating effect on the effect of numeric ratings on the amount of price premiums and sales of a seller. The three characteristics can have a supplementary role, meaning they can convey information a numeric rating in itself would not be capable of. In this way, the impact a rating has on the reader's estimation of trustworthiness of the rated person could be strengthened – or diminished – by characteristics of the text.

In short, it remains unclear as of yet to what extent and how the characteristics of feedback texts affect price premiums and sales in illegal online markets. That higher ratings positively affect both has been established for cryptomarkets as well (Przepiorka et al., 2017). Also, research has shown that ratings only affect sales indirectly through sentiment in feedback texts (Hu et al., 2014). Feedback texts may therefore be of way more importance in conveying the trustworthiness and reputation of a seller than has initially been thought, or at least than was put attention to in the literature. Still, the influence (and the possible moderating role) of polarity, subjectivity and emotionality of feedback texts on the price premiums and sales of a seller has yet to be researched. This is especially the case for illegal cryptomarkets and to a considerable extent for legal online markets as well.

This leads to the following research questions:

1a. To what extent does the polarity, subjectivity and emotionality of feedback texts affect price premiums of sellers in cryptomarkets (through trust)?

1b. How does the inclusion of textual features moderate the effect of quantitative reputation measures on prices?

2a. To what extent does the polarity, subjectivity and emotionality of feedback texts affect sales of sellers in cryptomarkets (through trust)?

2b. How does the inclusion of textual features moderate the effect of quantitative reputation measures on sales?

Our study contributes to the knowledge of feedback systems, by uncovering new ground through focusing on the direct and moderating effects of text characteristics of feedback texts on the success of a seller. Moreover, contributions made in our study about trust in the context of illegal cryptomarkets can be translated to contributions to the trust literature on online markets in general. This will in turn also contribute to the wider trust literature in the social and behavioural sciences.

We use data from AlphaBay, a former cryptomarket active between 2014 and 2017, which was mostly used for illegal drug trade. In 2017, after it was shut down, the whole marketplace was scraped. We focus on that part of the data that relates to several drug categories. The large dataset includes quantitative feedback (numerical ratings) as well as qualitative feedback (feedback texts) by users about the sellers. The text characteristics 'polarity', 'subjectivity' and 'emotionality' are distilled from the qualitative feedback texts through manual coding followed by machine learning methods. With this data, we can test our hypotheses. First, we will lay out the theoretical underpinnings of these hypotheses. Subsequently, we will introduce the data more extensively and show the methods used. We then proceed to discuss results and finalize with a conclusion and discussion.

Theory

In principle, reciprocally profitable online market exchanges are subject to a problem of cooperation as a consequence of information asymmetries between sellers and buyers (Macanovic & Przepiorka, 2021). This is even more so the case for illegal (crypto)markets, as these lack legal, organizational, social and moral guarantees legal markets tend to have for stimulating trust and cooperation (Diekmann & Przepiorka, 2019; Milgrom et al., 1990).

For both legal and illegal online markets, game theory is helpful in explaining the problem of cooperation. More precisely, the problem of cooperation can be viewed as a trust game with incomplete information, also referred to as a TGI. According to Jiao et al (2021), with the default trust game, the buyer moves first and has to decide whether or not to send money for the product to the seller (see Figure 1, TG of right sub-tree). The seller is second to decide and can either ship or not ship the product after receiving the money. For the seller, it is most advantageous not to ship (T>R). After all, receiving the money but also maintaining the product means no loss, and yields the highest possible profit. As the buyer is aware of this, the buyer declines to buy (P>S). As a consequence, no transaction takes place and the higher possible payoff 'R' for both is lost to a lower equal payoff 'P'.

As can be seen in Figure 1, in the TGI, a buyer has to decide whether the scenario of the aforementioned default trust game is applicable, or that of the assurance game (AG, left sub-tree). In the assurance game, the seller can be rewarded after shipping, or sanctioned after not shipping. These rewards and sanctions can in principle constitute many things. It can be, for example, the social preference of a seller to also be of use to others (the buyers) in economic exchange. Not shipping wouldn't align with this preference and thus constitute a sanction for the seller. Shipping would, on the other hand, give an additional reward, being the knowledge of the seller that the transaction was also of use for the buyer. Still, here, only the seller knows whether the extra award of 'b' makes 'R' more valuable a payoff compared to the payoff of 'T' minus the sanction 'c'. The reputation system has the function of repairing this information asymmetry, as it gives a buyer insight into the possible payoffs for the seller, with the related possible rewards and sanctions. The visible reputation and reviews on the behaviour of the seller in prior exchanges enables the buyer to make out what payoffs are applicable for the seller, and thus in which of the games the buyer will embark (AG or TG). For sellers with a good reputation, it will be costly not to ship and lose reputation as a consequence of a negative feedback that will result from a scam (the sanction). Sellers with a bad reputation on the other hand, have much less to lose and therefore they can collect the money and leave the market, which is more profitable in one 'isolated' game. Good reputation of a seller thus grants an assurance to the buyer that buying poses hardly any risk (for a more extensive explanation of the TGI, see the theory-section of Jiao et al., 2021).
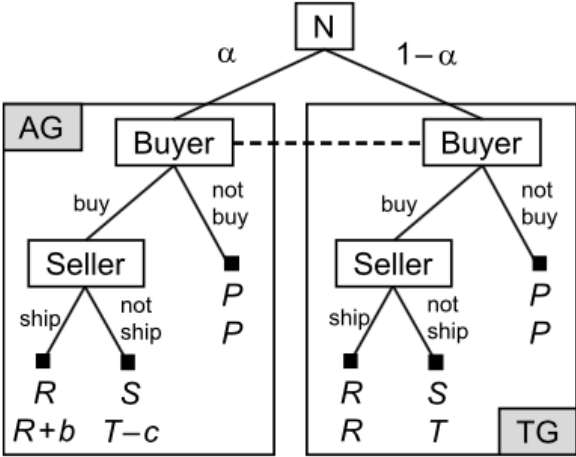

Figure 1. *Trust game with incomplete information (from Jiao et al., 2021)*

In this way, a good reputation leads to more sales of a seller, as the literature confirms (Ye, Li et al., 2009). Furthermore, it also enables sellers to sell their products for higher prices (known as price premiums) (Ba & Pavlou, 2002; Pate, 2006). Buyers are prepared to pay

more in exchange for the assurance that the quality of the service and product will be proper. Buyers thus have a preference to trade with sellers who have a good reputation. As a consequence, sellers who have yet to build a reputation are forced to compete by lowering their prices, making themselves more attractive trading partners (Friedman & Resnick, 2001).

Studies show that indeed, people massively consult these feedback systems on a daily basis for products and services, and increasingly so. A survey by BrightLocal (2020) has shown that in 2019, 81% of consumers read reviews of local business, while in 2020 this increased to 87%. For people aged 35 to 54, this is even higher with 93%. Of much importance as well: 79% of consumers claim they trust online reviews as much as personal recommendations they receive from family or friends. This is even 91% for people between the age of 18 and 34.

In online reputation systems, ratings are one of the indicators of the reputation of the seller. Research has suggested that ratings directly influence the amount of sales (Ye, Law et al., 2009). Moreover, it has also been shown that ratings only affect sales indirectly through sentiment in feedback texts (Hu et al., 2014). Hu et al. only focus on text positivity and negativity in studying the effect of sentiment of a text on sales. However, apart from positivity and negativity (also known also polarity), sentiments in feedback texts are also conveyed in large part through the text characteristics of subjectivity and emotionality. What these text characteristics of polarity, subjectivity and emotionality entail and how they are of influence, will be discussed in the sub parts below.

It is important to note the common distinction made between types of products on the (online) market: search goods and experience goods (Huang et al., 2009). The quality of search goods can be estimated prior to buying them, because they are more characterized by objective information. Examples of search goods are home furnishings, tools and electronics. Experience goods on the other hand, are less easily objectively characterizable and thus the quality of these goods can strictly be established by the buyer after consumption (Ghose & Ipeirotis, 2011). Services such as restaurants and cinemas, as well as products such as recorded media and recreational drugs, are examples of experience goods. This distinction between search goods and experience goods will recur throughout the remainder of this theory section and shall help nuancing the effects of the factors subjectivity and emotionality.

<center>Polarity</center>

Polarity, also known as valence, relates to the positive or negative quality of something. In our context, it characterizes how positive, neutral or negative the content of a buyer's

judgment is on the quality of a product, vendor or experience in general. With ratings, the polarity is simply translated to high ratings for positivity and low ratings for negativity. For feedback texts, this is more fine-grained. Parts of a text can be very positive, while other parts can be very negative, or neutral.

As mentioned before, higher ratings lead to more price premiums and sales. When a buyer sees other buyers rate their experience(s) with the seller positively, the trust of the buyer in the seller increases. Regarding the polarity of texts, Wang et al. (2017) show that 'sentiment polarity' expressed in feedback texts influences sales and prices.

Pertaining to how this works, the trust literature shows that information about extraordinary past behaviour most effectively can create a buyer's trust in a seller (Pavlou & Dimoka, 2006). When a person is surprised, this is most likely to change his or her beliefs (Bikhchandadi et al., 1992). To show how this works in our context, Pavlou & Dimoka (2006) distinguish between outstanding and abysmal comments on the trustworthiness of a seller. Here, outstanding comments don't simply confirm that the seller carried out the proceedings of the transaction as agreed upon, i.e. as expected by the buyer. Outstanding comments exceed a mere positive rating by laying out how the seller did more than obliged, for example in terms of delivery, service or product quality. Abysmal comments on the other hand, surpass a mere negative rating. These comments refer to cases where it is sure that the seller deliberately exploited buyers, for example by delivering products of lesser quality, or by not delivering at all. Buyers would generally look for sellers who are reliable to the extent that they will deliver their products in the manner and under the conditions as agreed upon. When buyers read about outstanding behaviour of a seller, their expectations can be surpassed. As a consequence, they will more easily have trust that the seller will at least adhere to the expectation of following up the agreement – and possibly more. When buyers read abysmal comments describing fraudulent behaviour, this strongly violates their expectations, losing their trust in the seller. In this way, surprise causes a change in the believes of the buyer.

Feedback text is suitable for a buyer to require this information of extraordinary behaviour of a seller. Likely even more so than ratings. A high(est) or low(est) ratings does, after all, not explicitly prove to the buyer that – let alone how – sellers did more than was expected from them, or that they seriously violated another buyer's trust.

Pavlou & Dimoka (2006) show that indeed extraordinary behaviour of the seller described in feedback text leads to higher price premiums, even when controlled for the rating of the feedback. The same effect can be expected for sales of a seller, as reputation also

directly influence sales (Ye, Law et al., 2009). Trust (or lack thereof) is indeed likely to be decisive in the consideration of a buyer whether or not to buy.

H1: *More positive feedback texts lead to higher price premiums of sellers in cryptomarkets.*

H2: *More positive feedback texts lead to more sales of sellers in cryptomarkets.*

Moreover, as Pavlou & Dimoka prove, outstanding and abysmal comments are accompanied without exception by positive and negative ratings, respectively. Positive and negative ratings, however, are not accompanied as regularly by outstanding or abysmal comments, but also at times by ordinary comments (Pavlou & Dimoka, 2006). A study of over 280 million online reviews confirms that generally the polarity of texts and numerical ratings don't (necessarily) correspond; the polarity of texts is less positive compared to the polarity of numerical ratings (Schoenmueller et al., 2020). This could mean that people tend to be more harsh in their description, but are still hesitant to give a very low score. What this dissimilarity between feedback texts and numerical ratings means for the effect on reputation and success of a seller, has yet to be uncovered. It would therefore be important to also look at the interaction between feedback texts and ratings.

A plausible mechanism here would be that when a reader consults the numerical rating and the textual feedback of the review, the numerical rating is taken in firstly. In AlphaBay, a numerical rating is shown as a green plus for positive ratings, a blue dot for neutral ratings and a red minus for negative ratings. The appreciation thus shown is more quickly comprehended than that of a feedback text, which requires more scrutiny. If after having seen the rating, a feedback text with a matching polarity is taken in, this could reinforce the impact the numerical rating makes on the reader. Thus, the effect of ratings on the success of a seller is moderated by the (matching) polarity of the feedback.

H3: *Polarity of feedback texts reinforcingly moderates the effect of ratings on the price premiums of a seller.*

H4: *Polarity of feedback texts reinforcingly moderates the effect of ratings on the sales of a seller.*

Lastly, people tend to write longer reviews when writing negative reviews and these reviews are also more detailed (Martens & Johann, 2017). Long reviews are considered to be more helpful by buyers (Mudambi & Schuff, 2010; Peng et al., 2014). Length of the review would thus be a beneficial controlling factor, as it makes it possible to evaluate whether a different influence of positive or negative reviews on trust is (in part) explained by the correlating review length.

<center>Subjectivity</center>

Subjectivity of a text refers to texts or parts of texts which express the opinion, personal experience or feelings of the buyer. Conversely, objectivity of a text relates to the factual information of e.g. a product, service or delivery which buyers aim to convey without interference of their prejudices, feelings or biases (Liu et al, 2018).

Using the distinction between search and experience goods, Ghose and Ipeirotis (2006) find that for search goods, higher subjectivity leads to more sales of products. For experience goods however, no significant effect was found. They theorize that people prefer reading the individual experiences of others when it comes to search goods, but not so much for experience goods. Since the quality of experience goods is more dependent on a person's preference as said before, another person's experience might be considered not sufficiently informative. For both search and experience goods, they find that a combination of objective and very subjective information in reviews has a negative impact on product sales in comparison to strictly objective or subjective reviews.

Still, the findings of their study and the applied explanation might not be equally applicable in our context. Making up the largest percentage of trades in cryptomarkets are illegal drugs (Soska & Christin, 2015). Moreover, in our study we will focus on (recreational) drug categories only. Drugs evidently fall more under the experience goods category, as they are used for a subjective experience. Of course, objective information can be given about for example the specific concentration, proportions and/or ingredients of the drug. Still, in illegal context, buyers are likely to be more suspicious about the given objective information and also in many cases don't have the means to test this after buying it. Furthermore, subjective text is suited for describing sellers in their communication and handling of transactions. In other words, an overall impression of the seller can be given by a reviewing buyer. Given the illegal, anonymous context of cryptomarkets, these subjective features of a seller can be expected to be of more importance to buyers, as trustworthiness of sellers is especially unbeknownst to buyers here. For a buyer seeking confirmation about the product and seller, a

subjective review of another buyer could have more persuasiveness, leading to more sales and higher price premiums of sellers.

H5: *More feedback texts with high subjectivity lead to higher price premiums of sellers in cryptomarkets.*

H6: *More feedback texts with high subjectivity lead to more sales of sellers in cryptomarkets.*

Furthermore, subjectivity in feedback text has the potential of conveying information to other buyers additional to just the quantitative numeric rating. These quantitative ratings are more easily comprehended as being 'objective' descriptions of the item and/or seller (even if perhaps ostensibly so). Feedback texts give a buyer the possibility to also express their subjective experience to other buyers. In this way, subjectivity of feedback texts can strengthen the effect quantitative ratings have on the price premiums and sales of a seller.

H7: *More feedback texts with high subjectivity reinforcingly moderate the effect of ratings on the price premiums of a seller.*

H8: *More feedback texts with high subjectivity reinforcingly moderate the effect of ratings on the sales of a seller.*

Emotionality

In our context, emotionality of a text relates to the emotion and feelings a buyer expresses in text about the experiences of the transaction, product and/or vendor. This emotion can be shown in a variety of ways: either through words, punctuation marks, use of caps lock, as well as cursing and smileys. People know how to detect emotion in a text (Hancock et al., 2007). More specifically, people are able to detect the emotion aimed to convey by the writer of a text with an accuracy of about 73% (Aman & Szpakowicz, 2007). The question is then, how the emotion of a text influences the trust of a person.

In a 2019 study, it has been shown that reviews that contain more factual instead of emotional writing are regarded as more trustworthy (Carbonell et al., 2019). However, this study only focused on one particular type of product, namely laptops; a search good instead of an experience good such as drugs is. A more nuanced take on the impact of emotionality on the trust in online reviews is given by Rocklage and Fazio (2020). They make a distinction

between types of products – search and experience goods – contrary to the study of Carbonell et al. (2019). Through both controlled experiments and computational linguistic analysis of 100,000 reviews from Amazon on 500 products, they studied how emotionality expressed in review texts amplifies or diminishes trust in the review and positivity about the product for the reader. They find that with experience goods, emotion expressed in a review leads to augmented trust, whereas emotions in reviews for search goods causes a reader to be less trustful of the review (Rocklage & Fazio, 2020). They theorize that this is due to violated expectations of the reader. If individuals' expectations are not met, causal reasoning offers a way to explain the unexpected event (Hastie, 1984; Weiner, 1985). When emotion is expressed in a review about a product of which the reader had not expected it to elicit an emotional response in any way - as would be the case with search goods - suspicion is raised with the reader. In such a situation, the explanation is either that the product can indeed bring forth an emotional response contrary to expectations, or the explanation is that the review is not in any way representative of other reviews and the product itself (Rocklage & Fazio, 2020). This is applicable to search goods, which are less likely to evoke emotions as opposed to experience goods of which the purpose is in large part to do so. With search goods, it is more likely to be the review that is not representative instead of the product in fact being able to cause emotional reactions, and thus trust in the review will be lowered. With (reviews of) experience goods however, emotional responses are sought after by readers, causing no violations of expectations and those reviews would actually be considered more representative and trustworthy.

H9: More feedback texts with high emotionality lead to higher price premiums of sellers in cryptomarkets.

H10: More feedback texts with high emotionality lead to more sales of sellers in cryptomarkets.

Here, again, interaction between the numerical ratings and the textual feedback of the reviews is likely. As followed from the literature on emotionality of texts, for experience goods, the emotionality of a text augments the trustworthiness of the review. We can make the assumption that a reader of the text who gains trust in the (writer of the) review, also gains more trust in the numerical part of the review. This would then work reinforcingly for both positive and negative ratings. A positive rating accompanied by an emotional review would
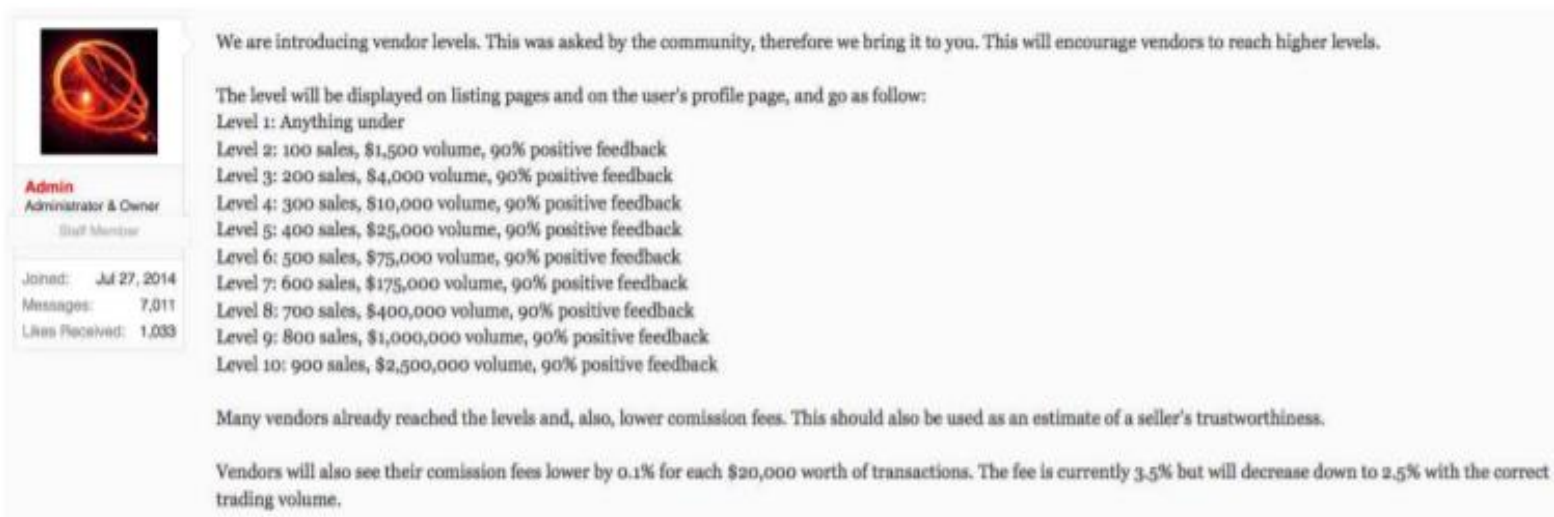
be more relied upon by the seller, and in the same way a negative review backed with an emotional review would also be more relied upon. Following this line of argumentation, emotionality of text has a reinforcing, moderating effect on the effect of quantitative ratings on amount of sales and price premiums.

H11: More feedback texts with high emotionality reinforcingly moderate the effect of ratings on the price premiums of a seller.

H12: More feedback texts with high emotionality reinforcingly moderate the effect of ratings on the sales of a seller.

Data and Methods

We use data from a now inactive cryptomarket named AlphaBay, which was mostly used for illegal drug trade. It was active between 2014 and 2017. Shortly before being shut down, all available data was scraped. The part of the data we use, is of items that were posted between 1/5/2017 and 16/6/2017 by sellers on the market. These items fall under several drug categories. These categories are: Weed, Hash, MDMA, Cocaine, Ketamine, Meth & Heroin. The dataset consists of 1655 posted items by a total of 555 individual sellers. It also contains information about quantitative and qualitative feedback for each seller on a given transaction up to the moment they posted the item.

We are introducing vendor levels. This was asked by the community, therefore we bring it to you. This will encourage vendors to reach higher levels.

The level will be displayed on listing pages and on the user's profile page, and go as follow:
Level 1: Anything under
Level 2: 100 sales, $1,500 volume, 90% positive feedback
Level 3: 200 sales, $4,000 volume, 90% positive feedback
Level 4: 300 sales, $10,000 volume, 90% positive feedback
Level 5: 400 sales, $25,000 volume, 90% positive feedback
Level 6: 500 sales, $75,000 volume, 90% positive feedback
Level 7: 600 sales, $175,000 volume, 90% positive feedback
Level 8: 700 sales, $400,000 volume, 90% positive feedback
Level 9: 800 sales, $1,000,000 volume, 90% positive feedback
Level 10: 900 sales, $2,500,000 volume, 90% positive feedback

Many vendors already reached the levels and, also, lower comission fees. This should also be used as an estimate of a seller's trustworthiness.

Vendors will also see their comission fees lower by 0.1% for each $20,000 worth of transactions. The fee is currently 3.5% but will decrease down to 2.5% with the correct trading volume.

Figure 2. *The assignment of Vendor levels in AlphaBay*

This quantitative information is firstly expressed by a rating of a buyer's experience with a seller as either 'Positive' (+), 'Neutral' (.) or 'Negative' (-). In the reputation system of the

market, this was then accumulated for each seller into a '% of positive' number. If, for example, a seller was rated only with Positive ratings and no Neutral or Negative ratings, the number was 100%. Apart from that rating, in the data there is also information on a seller's 'Trust level' and 'Vendor level', as these occurred in the market. Vendor levels were assigned to sellers based on a combination of the amount of sales, the amount of money made from these sales and a 90% positive feedback score. See Figure 2 above for the level assignment per combination.

On how Trust levels were assigned, less clarity was given by the administrators of AlphaBay. The reason behind this was so that market users could not purposely try to boost their own reputation via other users or by themselves through fake accounts - or decrease the reputation of others for their own sake (Kalberg, 2017). Trust levels have in common with Vendor levels that they also range from 1 to 10. Trust levels, however, were also applicable to buyers instead of being limited just to sellers, as was the case with Vendor levels.

For our study, we will use the three variables relating to quantitative reputation scores separately from one another. Trust levels and Vendor levels are operationalized simply by leaving the scores from 1 to 10 each seller scored on these levels and will be treated as continuous variables. The ratings will also be handled in the same way they occurred in the market. By dividing the amount of positive ratings by the total amount of ratings gathered by the seller (received up until the moment of posting the item), a percentage between 0 to 100 will be applicable to each seller. The total amount of quantitative and qualitative ratings will also be included as control variables; additional control variables will be discussed later in this section.

The qualitative feedback consists of textual feedback by buyers about the sellers. Writing a feedback text was voluntary for buyers in the market, so not every rating is accompanied by a feedback text. For our study, a subset of the feedback texts has been manually coded on polarity, subjectivity and emotionality. This coding has been carried out by thesis-students and supervisors, as well as by people who were briefly introduced to the general topic (ELSE lab participants)[1]. The manual coding was followed by machine learning methods to code the remaining feedback texts. In short, this entails that through text mining the remainder of the texts were assigned the degrees of polarity, subjectivity and emotionality.

During the coding, a text could be coded on polarity as either 'Positive', 'Neutral', or 'Negative'. In the operationalization, we will use the percentage of positive feedback texts a

---

[1] ELSE lab is an experimental laboratory affiliated with Utrecht University. Here, people (mostly students) participate in social science and economic experiments and earn around 8 to 10 euros per hour for participating.

seller received up until the moment of posting the item. For subjectivity, a text could be coded as 'Objective', 'Somewhat subjective', 'Rather subjective' and 'Very subjective'. Similarly, for emotionality a text could be coded as 'Unemotional', 'Somewhat emotional', 'Rather emotional' and 'Very emotional'. In both the cases of subjectivity and emotionality, the 'Rather' and 'Somewhat' variants were merged later on, thus resulting in three possible degrees for each description. Subjectivity and emotionality are operationalized in the same manner as polarity. Thus, by using the percentage of 'Very subjective' and 'Very emotional', respectively, of the total amount of feedback texts acquired by a seller received up until the moment of posting the item. This operationalization does introduce missing values for sellers who had not received any feedback up until the moment the item was posted.

Table 1

Descriptive statistics of relevant variables

Variable	*n*	*M*	*SD*	Median	Mode	Min	Max	Skewness
Price per gram (log)	1655	2.87	1.07	2.55	2.55	.21	7.31	.83
Amount of sales	1655	5.75	11.60	2	2	1	198	7.08
Perc. positivity	1176	.89	.12	.90	.90	0	1	-4.33
Perc. subjectivity	1176	.46	.16	.47	.47	0	1	-.32
Perc. emotionality	1176	.01	.03	.009	.009	0	.5	9.09
Per. pos. ratings	1182	.98	.04	.99	.99	0	1	-13.81
Item days online	1655	26.58	11.41	27	27	0	47	-.07
Feedback length	1655	8.17	6.22	9.96	9.96	0	40	0.38
Trust level	1655	4.45	1.39	4	4	3	10	1.04
Vendor level	1655	3.34	2.58	2	2	1	10	.83
Weight in grams	1655	56.48	235.76	7	7	.05	4535.92	11.67
Price category	1655	1.50	.69	1	1	1	3	1.01
Amnt. of fb. texts	1655	420.70	1058.25	52	52	0	11884	5.01
Amnt. of ratings	1655	556.36	1379.47	78	78	0	16927	5.17

The variable for sales of a seller is operationalized simply by using the total amount of sales as of the moment the item was posted. Descriptions of this variable show, however, that it is highly skewed (7.08) (see Table 1). The significant p-value of the Shapiro-Wilk test shows that indeed normality cannot be assumed (W = 0.404, p-value < .001). This would pose a problem for multiple linear regression, as multivariate normality is an important assumption there. However, negative binomial regression does not assume (multivariate) normality. We will therefore use this type of regression for the hypotheses concerning the sales of a seller. Negative binomial regression does assume independence of observations, overdispersion and that the dependent variable is a count variable, among other things (UCLA, 2021). A count variable is a variable which only has non-negative, whole-numbered values. The variable for the amount of sales is a count variable and thus this assumption for negative binomial regression is met. The assumption of overdispersion has to be met as well. There is overdispersion when the observed variance is higher than the theoretical model assumes. Testing a Poisson model with the same variables shows that the distribution of the variables is indeed over-dispersed (dispersion = 18.09, z = 4.22, p < .001). A problem for the independence of observations, is that in our dataset observations are handled per item, instead of per seller. A seller with several items is common in the dataset, which violates the assumption of independence. To mitigate this problem, we will work with clustered standard errors and seller fixed effects.

Price premiums will be operationalized by using the price per gram of the item. However, as the histogram shows (Figure 3), the variable isn't normally distributed either. The Shapiro-Wilk test confirms this (W = 0.415, p-value < .001). Since price per gram is not a count variable, negative binomial regression isn't an option for this model. Instead, we transform the dependent variable into a log variable, which results in a distribution that does approach normality, as can be seen below in Figure 3. We will use this log variable as the dependent variable in a multiple linear regression.
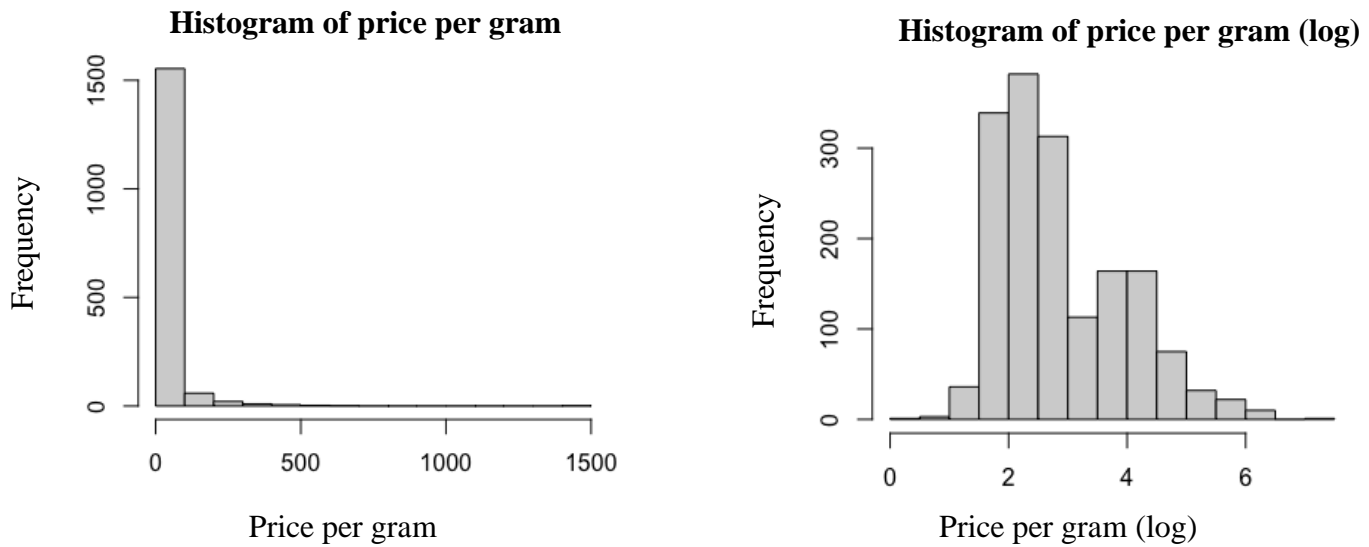
Figure 3. *Histograms showing the distributions of price per gram variable*

Multiple regression has a larger set of assumptions; we will only discuss those that are possibly violated in our dataset. Again, to mitigate the problem of dependence of observations, we will work with clustered standard errors seller fixed effects. Checking the VIF values of all variables, the assumption of no multicollinearity seems to be violated for multiple variables. However, the variables for the moderation effects are made out of the textual feature variables, thus correlation between all of these variables is unavoidable. Accordingly, multicollinearity is irrelevant for included moderation effects (McClelland et al., 2017). The total amount of quantitative ratings and qualitative ratings are also highly correlated (VIF's of 97.99 and 96.14, respectively). To counter this possible problem of multicollinearity, we will exclude the variable relating to the total amount of quantitative feedback. Ultimately, this variable might be slightly less important, since the main focus of this study is on qualitative feedback. We will also leave out this variable in the negative binomial model.

Figure 4 below shows the model which will test the hypotheses regarding the effects of the qualitative measures on price premiums of a seller (H1, H3, H5, H7, H9 & H11). The possible moderating role of the textual features in the effect of the quantitative ratings on price premiums is also shown in this model.
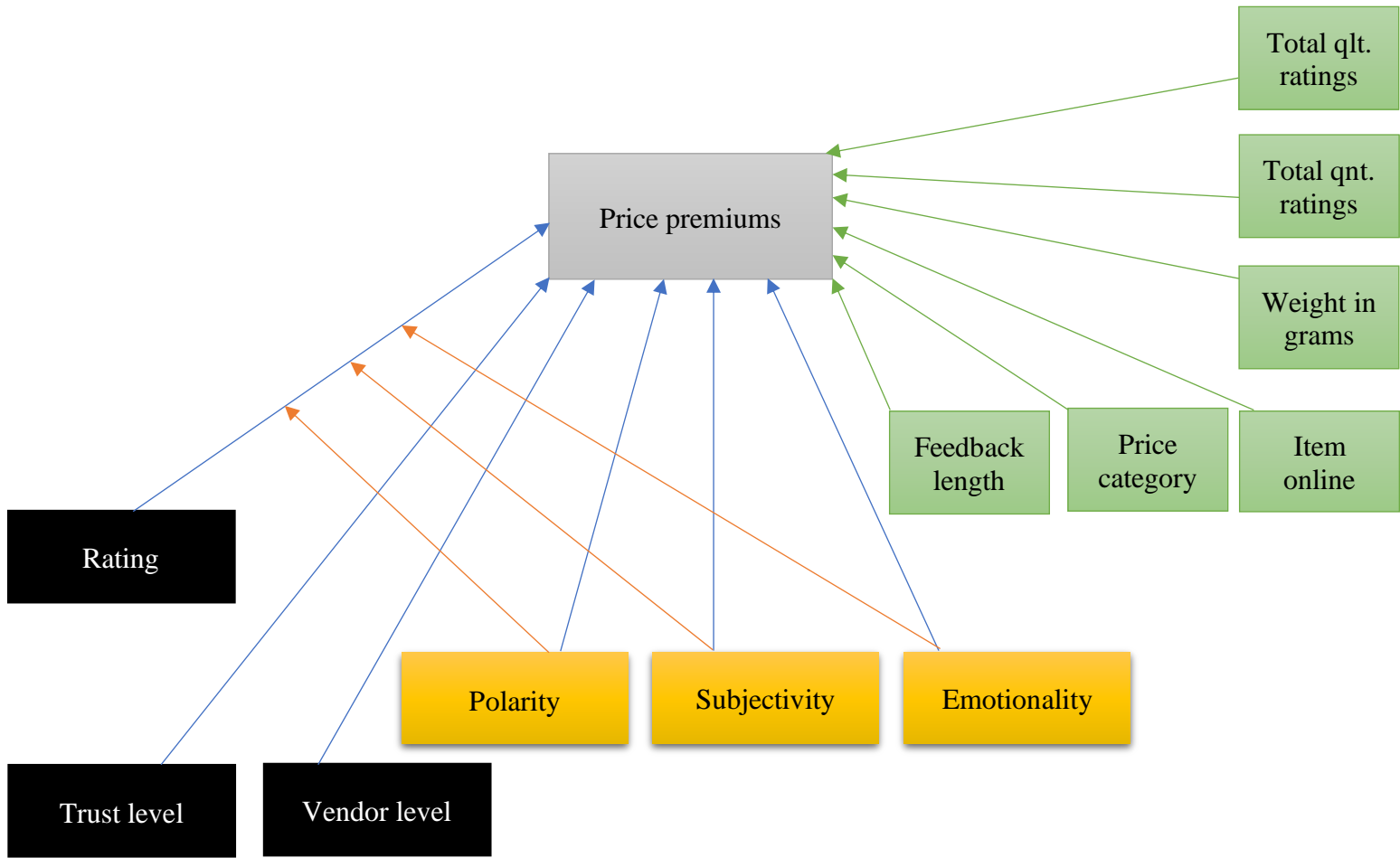
Figure 4. *Full Model 1 - the effect of qualitative and quantitative measures on price premiums*
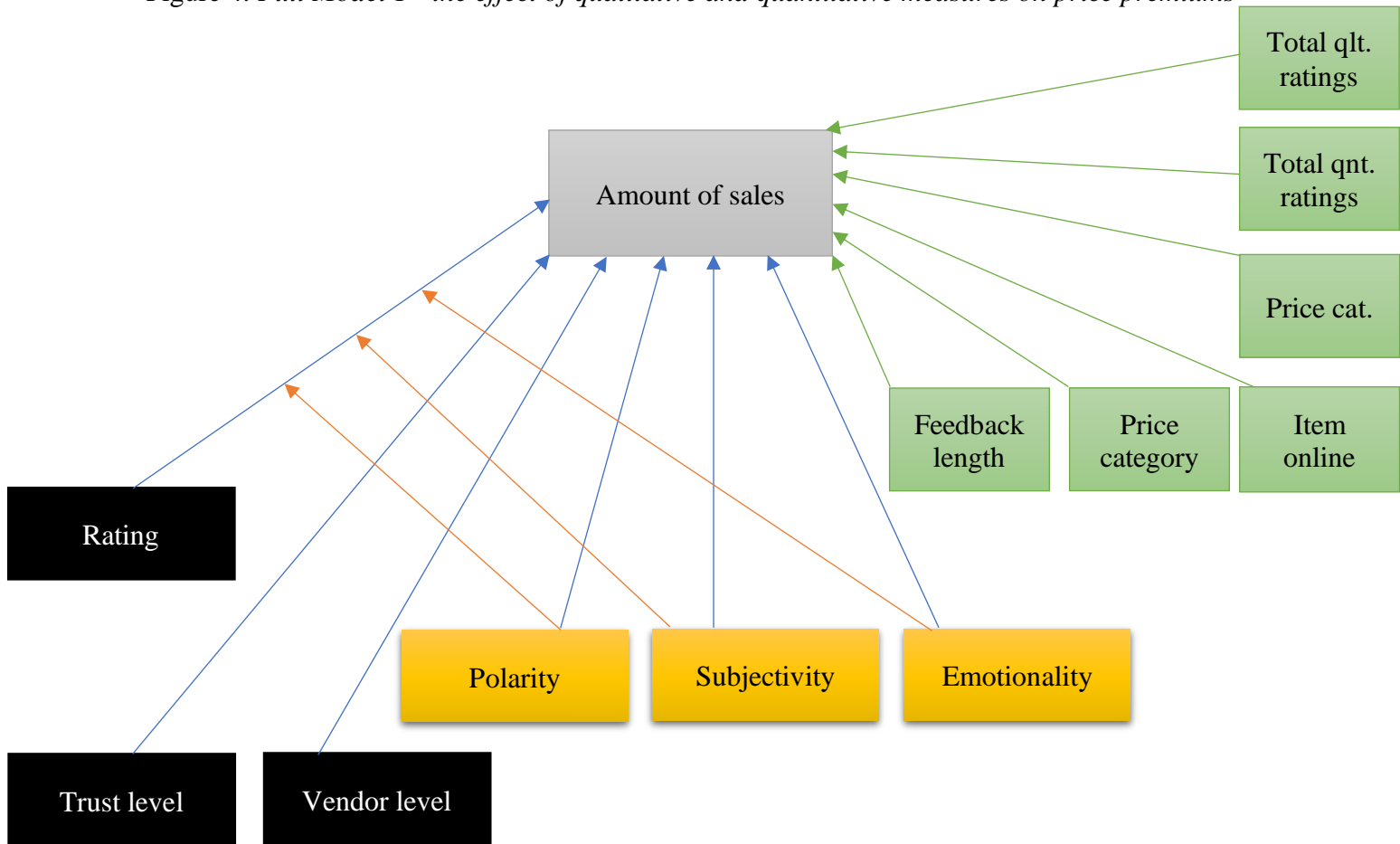

Figure 5. *Full Model 3 - the effect of qualitative and quantitative measures on amount of sales*

Above in Figure 5, the model is shown which will test the hypotheses regarding the sales of a seller (H2, H4, H6, H8, H10 & H12). For the main independent variables of interest - polarity, subjectivity and emotionality - the direct effect on amount of sales will be tested. This model also contains the moderating effect of these textual features on the effect of the quantitative reputation measures on amount of sales.

For testing the hypotheses regarding the moderation effects, we will compare the model fit of the full models shown in Figure 4 and 5 with their respective restricted models, which leave out the moderation effects. Then, for testing the hypotheses regarding the main effects of the qualitative measures on sales and price premiums, we will use the models that have the best model fit.

As followed from the theory, a useful control variable is that of feedback length. What we will use is the average feedback length in words of all the feedback texts the seller received. The theory suggested this control variable only for polarity. However, we will also include it for subjectivity and emotionality to exclude a possible suppressor effect there as well. We also control for the amount of days the item was online before the data collection, to ascertain that an item of a seller wasn't sold more merely by the fact that it was for sale for a longer period of time than another item. Additionally, we will control for price categories, as this factor assumedly can affect both the amount of sales and price premiums. In the model testing the effect on price premiums, we also control for the weight in grams of the product, since higher weights tend to lead to lower prices. In order to keep both models as parsimonious as possible, some factors which are of lesser interest, are left out (e.g. item category, shipping category),

All methods will be carried out in statistical software R (R Core Team, 2017)[2].

Results

We will first compare the model fits for the models regarding the price premiums (H1, H3, H5, H7, H9 & H11) (see Table 2). The outcome of the F-test for Model 1 including the moderation effects is significant ($F_{(15, 1160)} = 116.7$, $p < .001$, $R^2 = .596$). The residual standard error for the model is .679 on 1161 Df. The restricted Model 2 without the moderation effects also has a significant F-test outcome ($F_{(12, 1163)} = 141.5$, $p < .001$, $R^2 = .589$). The F-statistic of Model 2 is significantly higher than that of Model 1, as shown by the

---

ANOVA test comparing the models. However, the BIC of Model 2 (BIC = 2531.893) is slightly higher than the BIC of Model 1 (BIC = 2530.253). Also, the Adjusted R squared is higher for the full Model 1, albeit again, minutely. Thus, the full model explains circa 0.7% more of the variance than the restricted Model 2. Also, the residual standard error is slightly smaller for the full Model 1 (RMSE = .679) than for the restricted Model 2 (RMSE = .685). A smaller RMSE usually indicates a better model fit, but in this case, we take the minute difference to be negligible. Taking into account that for Model 2 the F-statistic is significantly higher while all other differences are rather diminutive, we assume that overall the restricted Model 2 without the moderation effects is a better fitted model.

Table 2. *Model fit comparison (Model 1 & 2)*

Model	F-statistic	BIC	RMSEA	Adjusted R2
Full (Model 1)	116.7***	2530.253	.679	.596
Restricted (Model 2)	141.5***	2531.893	.685	.589

*N=1176, *p<.05, **p<.01, ***p<.001*

The following hypotheses therefore cannot be accepted: *H3 'Polarity of feedback texts reinforcingly moderates the effect of ratings on the price premiums of a seller.', H7 'More feedback texts with high subjectivity reinforcingly moderate the effect of ratings on the price premiums of a seller.', H11: 'More feedback texts with high emotionality reinforcingly moderate the effect of ratings on the price premiums of a seller.'*

We'll therefore use Model 2 for interpreting the main effects of the qualitative measures on price premiums (see Table 3). Firstly, we will look at the main effect of the quantitative measure. The percentage of positive quantitative ratings has no significant effect on the price per gram ($\beta = 2.066$, $t = 1.653$, $p = .073$). Then, the qualitative measures. The percentage of positive qualitative feedback texts shows no significant effect on price per gram ($\beta = .157$, $t = .864$, $p = .538$). As a consequence, the following hypothesis cannot be accepted: H1 '*More positive feedback texts lead to higher price premiums of sellers in cryptomarkets.'*

Concerning the effect of percentage of highly subjective feedback texts on the price per gram, there is a significant positive effect ($\beta = .551$, $t = 2.359$, $p = .003$). Because the dependent variable is log-transformed here, we need to exponentiate the coefficient, subtract one from this outcome and then divide that by 100. Carrying this out, gives an outcome of 73%. This means that for every unit increase in percentage of received highly subjective feedback texts by sellers, their items are on average priced 73% higher (in USD), when all

other variables are hold constant. Thus, the following hypothesis can be accepted: H5 '*More feedback texts with high subjectivity lead to more price premiums of sellers in cryptomarkets.*'

The percentage of highly emotional qualitative feedback texts shows no significant effect on price per gram (β = -.530, t = -.649, p = .517). It is worth noting here that the effect is negative, while we expected a positive effect. We clearly cannot accept the following hypothesis: *H9 'More feedback texts with high emotionality lead to higher price premiums of sellers in cryptomarkets.'*

Feedback length has no significant effect contrary to what the literature suggested, and also has a negative effect instead of the expected positive effect on price premiums (β = -.014, z = -1.630, p = 1.04). The control variable weight in grams has a very significant, yet weakly negative effect on price premiums (β = -.0007, t = -3.484, p < .001). This is in line with expectations, as higher weights tend to lead to lower prices. The dummy variables of price category have a very significant, positive effect on price premiums (d2: β = 1.435, t = 15.337, p < .001) (d3: β = 2.062, t = 14.900, p < .001). Thus, the higher the price category, the higher the price per gram. This is only logical, as price category indicates the level of the price.

Table 3

Model 2 Multiple regression analyses: Feedback text characteristics on price premiums (SE's clustered by seller)

Model 4 Negative binomial analyses: Feedback text characteristics on sales of item (SE's clustered by seller)

Variable	Estimate M2	Estimate M4
(Intercept)	.113 (1.201)	1.375 (1.910)
Perc. positive ratings	2.066 (1.473)	-.751 (1.867)
Perc. positivity	.157 (.256)	.546 (.407)
Perc. subjectivity	.551 (.186)*	.180 (.446)
Perc. emotionality	-.530 (.817)	2.300 (1.453)
Item days online	-.005 (.003)	.029 (.005) ***
Amount of feedback texts	-.0000006 (.000002)	.0001 (.00009)
Feedback length	-.014 (.008)	.020 (.017)
Weight in grams	-.0007 (.0002)***	
Trust level	.005 (.045)	-.252 (.077) **
Vendor level	.007 (.023)	.124 (.038) **
Price category (d2)	1.435 (.094)***	.056 (.156)
Price category (d3)	2.062 (.138)***	.234 (.191)

$N = 1176$, *** = p < .001, ** = p < .001, * = p < .05, d2 = dummy 2, d3 = dummy 3

Now, we will compare the model fits for the models regarding the sales of a seller (H2, H4, H6, H8, H10 & H12). We will look at the BIC for both models, where a lower BIC indicates better fit. As can be seen in Table 4, Model 4 without the interaction effects has a BIC of 6533.7, which is lower than the BIC of Model 3 (6553.3). Moreover, we'll have to look at the Log likelihood. The Log likelihood is marginally higher for Model 3 (-3223.6) than for Model 4 (-3224.5). For negative binomial regression, Pseudo R2 is commonly used, as an alternative for R2. The adjusted Pseudo R2 represents the improvement in model likelihood compared to a null model instead of the proportion of explained variance, as is the case with an OLS Adjusted R2 (Hemmert et al., 2016). The adjusted Pseudo R2 is slightly higher for Model 4 (.0234) compared to Model 3 (.0226). Taking into account both the BIC as well as the adjusted Pseudo R2, Model 4 has better model fit than Model 3. As a consequence, we will work with Model 4 (see Table 3) to ascertain the main effect of the qualitative measures on sales of a seller.

Table 4. *Model fit comparison (Model 3 & 4)*

Model	BIC	Log Likelihood	Adjusted Pseudo R2
Full (Model 3)	6553.3	-3223.6	.0226
Restricted (Model 4)	6533.7	-3224.5	.0234

N=1176

It also gives us reason to assume the following hypotheses regarding the moderation effects cannot be accepted: *H4 'Polarity of feedback texts reinforcingly moderates the effect of ratings on the sales of a seller.', H8 'More feedback texts with high subjectivity reinforcingly moderate the effect of ratings on the sales of a seller.', H12 'More feedback texts with high emotionality reinforcingly moderate the effect of ratings on the sales of a seller.'*

We'll now interpret the results of Model 4 (see Table 3 above). Again, we'll firstly look at the main effect of the quantitative measure. The percentage of positive quantitative ratings has no significant effect on the price per gram (β = -.751, z = -.402, p = .688). Contrary to what was expected, it even has a negative effect. Then, the effect of the qualitative measures on the sales of sellers. The percentage of positive qualitative feedback texts shows no significant effect on sales (β = .546, z = 1.314, p = .189). We therefore cannot accept the following hypothesis: H2 '*More positive feedback texts lead to more sales of sellers in cryptomarkets.'*

With regards to the effect of the percentage of highly subjective feedback texts, there is no significant effect either (β = .180, z = .408, p = .683). The following hypothesis cannot be accepted: H6 '*More feedback texts with high subjectivity lead to more sales of sellers in cryptomarkets.'*

The percentage of highly emotional qualitative feedback texts shows no significant effect on sales of sellers (β = 2.300, z = 1.453, p = .119). The following hypothesis thus cannot be accepted: *H10 'More feedback texts with high emotionality lead to more sales of sellers in cryptomarkets.'*

The control variable for the days the item was online does have a significant, positive effect on the amount of sales (β = .029, z = 5.544, p < .001). This means that, all other variables held constant, for each 1 increase in the days the item is online, the number of items sold can be expected to increase by 2.9%. Feedback length, although positive in this model, again has no significant effect on the reputation measure of sales (β = .002, z = 1.260, p = .208). Trust- and Vendor level both have a significant effect on amount of sales as well.

Curiously, though, Trust level has a negative effect (β = -.252, z = -3.250, p = .001), whereas Vendor level has a positive effect (β = .124, z = 3.269, p = .001).

Conclusion and discussion

We study to what extent feedback text characteristics are of impact on the success of a seller in online illegal cryptomarkets. These feedback texts are part of online reputation systems, which are of increasing popularity (BrightLocal, 2020). Apart from feedback texts, numerical ratings are given to market participants by other participants, which are often aggregated into one or several reputation scores. Reputation systems function to overcome the problem of cooperation, which arises as consequence of lack of trust, since there is an information asymmetry between buyers and sellers. Buyers in principle do not know whether the seller has good intentions and will ship a product after the buyer has paid. An online reputation system is a centralized digital place which gives buyers an opportunity to inquire wanted information on any product or service as well as (prior behaviour of) the vendor, by reading experiences of other buyers. This way, buyers can determine the seller's reputation and whether the seller is trustworthy enough to engage with in economic exchange. Just as importantly, it is a means for sellers to display their behaviour for every buyer to see (Macanovic & Przepiorka, 2021). Our study focuses on illegal cryptomarkets, where participants acquire anonymity through encryption, in order to exchange illicit goods – most commonly drugs (Soska & Christin, 2015).

Mainly pertaining to legal markets, scholars have initially found that the quantitative, numerical ratings are effective predictors of a seller's success, often expressed in the amount of sales and price premiums (Ba & Pavlou, 2002; Ye, Li et al., 2009). Over the recent years, more research is focused on the qualitative feedback texts. More information about the product, seller and overall experience can be transmitted through text than through numerical ratings. Polarity of feedback texts, relating to the positive and negative nature of the text, has often proven to be of influence on the success of sellers (Wang et al., 2017). Subjectivity is another possible characteristic of a text, which can showcase the buyer's subjective experience with of a transaction, seller and/or product. Here, a distinction is commonly made between search goods (utilitarian of nature) and experience goods (experience based). Previous literature has found that subjectivity of texts tends to increase success only for search goods (Ghose & Ipeirotis, 2006). Considering our illegal context, we theorized the contrary. Most buyers would not be able to test the objective information on the drugs, plus subjective information can mitigate the supposed untrustworthiness of an anonymous person

in an illegal market. Emotionality of text is another characteristic of feedback text. It has been shown in previous literature, that emotionality positively influences success for experience goods (Rocklage & Fazio, 2020). We further expand on the literature by taking into account all three of these text characteristics simultaneously. Moreover, we study their moderating effect on the effect of numerical ratings on success of a seller, as we theorized that these qualitative characteristics can strengthen the trust put in the quantitative numerical ratings.

The data we base our findings on, is from cryptomarket AlphaBay. The offered items in our data only pertain to several drug categories. We operationalize polarity, subjectivity and emotionality by taking the percentage of highly positive, subjective and emotional texts received by each seller upon the moment of posting an item. For the multiple regression model on price premiums, we log-transform the dependent variable, as the variable price per gram is not normally distributed. For the model on sales, we use negative binomial regression, as the operationalization of the dependent variable, amount of sales, is a count variable which is not normally distributed.

We find that more feedback texts with high subjectivity lead to more price premiums of sellers in cryptomarkets. More feedback texts with high subjectivity does not prove to be of influence on the amount of sales of sellers in cryptomarkets. We also did not find any evidence for an influence of the text characteristics of polarity and emotionality on the amount of price premiums or sales of a seller. We also take into account the effect of quantitative ratings on the amount of price premiums and sales. This has, however, not shown to be of any effect on the success of a seller, contrary to what the current literature suggests. More positive, subjective or emotional feedback texts also don't seem to strengthen the effect of the positive numerical ratings on the success of a seller, while we did expect this based on our theory.

That only subjectivity of texts leads to higher prices, while polarity and emotionality are of no effect, could implicate that experience goods are most easily promoted through subjectivity. Human experience and subjectivity are, after all, closely tied. Still, the same could be said for experience and emotionality, although perhaps to a lesser extent for the specific case of drugs in our context. That polarity was of no effect on either success measures, could indicate that surprise was either not elicited by the highly positive text, or that surprise simply was not of sufficient effect on buyers in their decision the buy and for how much money. Possibly, another mechanism is at play, which is more determinative. For instance, the opposites of the characteristics could perhaps be of most importance. As a result of the operationalization of polarity, subjectivity and emotionality of feedback text

characteristics, taking the percentage of their 'highly' equivalents, the effect of their neutral equivalents and of negative, objective and unemotional feedback texts has not been taken into account. This does not, however, take away from the findings on polarity, subjectivity and emotionality per se.

It is interesting to note that for the model estimating the amount of sales, both Trust level and Vendor level do have a significant effect, while none of the qualitative and quantitative feedback have any effect. Trust and Vendor level were both measures created by the AlphaBay market administrators. This could indicate that 'institutional' reputation systems turn out to have more impact on users than the pure user-based reputation systems. Perhaps especially so in an illegal market, as the anonymous and immoral nature of the market causes little trust to arise between participants. The reputation measures that are based on the 'institutional' reputation system, then, are more relied upon by lack of a better mainstay. Possible other mechanisms could (also) be at play here. The Vendor levels were an aggregation of the number of sales, amount of proceeds and a total of 90% received positive feedback of a seller. As the user-based feedback does play a role in this reputation measure and the measure also immediately shows the prior success of a seller in the amount of sales and proceeds, perhaps users were accustomed to simply base their decisions on the aggregated Vendor levels? Future study should therefore look into the effect of prior success, expressed by sales and proceeds as an indication of reputation, and the impact it has on future success of sellers in (illegal) markets. Quite odd still, and difficult to explain, is the negative effect a higher Trust level has on the amount of sales of a seller. As it was widely hidden how these Trust levels were made by the AlphaBay administrators, perhaps suspicions circulated with users that these levels could be bribed, which would lead to a decreased trust in the sellers who possibly undertook such immoral actions. Hopefully, future research could give more clarity.

Furthermore, it would be of interest to study how our findings on the interplay of text characteristics, numerical ratings, and institutional measures hold in a legal context. Also, our study only focused on experience goods. There could be further expanded on the knowledge of reputation systems in illegal cryptomarkets by also focusing on search goods, which presumedly works through different mechanisms for the text characteristics.

In short, our findings develop the current knowledge on illegal cryptomarkets; their workings and deficiencies. This study helps us understand how multiple characteristics of the feedback texts of buyers in these markets do impact the sales and prices of products, and how they do not. We also introduce the theoretical distinction between search goods and

experience goods into the literature on reputation systems of cryptomarkets. Moreover, we assess the interplay between a multitude of reputation measures, considering both qualitative and quantitative reputation measures, as well as 'institutional' measures. In order to attain these insights, we integrated previous literature on feedback texts in online reputation systems. The findings of this study emphasize how important it is to take into account as many reputational factors as can reasonably be expected to be of relevance in studying cooperation in online (illegal) markets.

References

Aman, S., & Szpakowicz, S. (2007). Identifying expressions of emotion in text. In International *Conference on Text, Speech and Dialogue* (pp. 196-205). Springer, Berlin, Heidelberg.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, 243-268.

Bergé, L (2018). "*Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm*." CREA Discussion Papers.

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.

Bozic, B. (2017). Consumer trust repair: A critical literature review. *European Management Journal, 35*(4), 538-547.

Carbonell, G., Barbu, C. M., Vorgerd, L., & Brand, M. (2019). The impact of emotionality and trust cues on the perceived trustworthiness of online reviews. *Cogent Business & Management, 6*(1), 1586062.

Chertoff, M. (2017). A public policy perspective of the Dark Web. *Journal of Cyber Policy*, 26-38.

Cook, K. (Ed.). (2001). Trust in society. Russell Sage Foundation.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: the influence of emotion on trust. *Journal of personality and social psychology, 88*(5), 736.

Fox, J., Weisberg, S., (2019). An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Friedman, E. J., & Resnick, P. (2001). The social cost of cheap pseudonyms. *Journal of Economics & Management Strategy, 10*(2), 173-199.

Ghose, A., & Ipeirotis, P. G. (2006). Designing ranking systems for consumer reviews: The impact of review subjectivity on product sales and review quality. In *Proceedings of the 16th annual workshop on information technology and systems* (Vol. 303, No. 10).

Ghose, A., and Ipeirotis, P. G. (2011). "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics." *IEEE Transactions on Knowledge and Data Engineering* 23(10):1498–1512.

Hancock, J. T., Landrigan, C., & Silver, C. (2007). Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 929-932).

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of personality and social psychology, 46*(1), 44.

Hemmert, G. A., Schons, L. M., Wieseke, J., & Schimmelpfennig, H. (2018). Log-likelihood-based pseudo-R 2 in logistic regression: Deriving sample-sensitive benchmarks. *Sociological Methods & Research, 47*(3), 507-531.

Hillmann, H., & Aven, B. L. (2011). Fragmented networks and entrepreneurship in late imperial Russia. American Journal of Sociology, 117(2), 484-538.

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision support systems, 57*, 42-53.

Huang, P., Lurie, N. H., & Mitra, S. (2009). Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods. *Journal of marketing, 73*(2), 55-69.

Kalberg, Å. H. (2017). An Endeavour in the Domain of Cybercrime: Exploring the Structural and Cultural Features of the Darknet Market AlphaBay (Master's thesis).

Király, O., Potenza, M. N., Stein, D. J., King, D. L., Hodgins, D. C., Saunders, J. B., ... & Demetrovics, Z. (2020). Preventing problematic internet use during the COVID-19 pandemic: Consensus guidance. *Comprehensive Psychiatry, 100*, 152180.

Kleiber, C., Zeileis, A. (2008). Applied Econometrics with R. Springer-Verlag, New York.

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management, 56*(2), 172-184.

Liu, S. Q., Ozanne, M., & Mattila, A. S. (2018). Does expressing subjectivity in online reviews enhance persuasion?. *Journal of Consumer Marketing.*

Macanovic, A., Przepiorka, W. (2021). The Moral Foundations of Immoral Markets: Text Mining Feedbacks on Economic Exchanges in the Darknet. (To be published)

McClelland, G. H., Irwin, J. R., Disatnik, D., & Sivan, L. (2017). Multicollinearity is a red herring in the search for moderator variables: A guide to interpreting moderated multiple regression models and a critique of Iacobucci, Schneider, Popovich, and Bakamitsos (2016). *Behavior research methods, 49*(1), 394-402.

Milgrom, P. R., North, D. C., & Weingast*, B. R. (1990). The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics, 2(*1), 1-23.

Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly*, 185-200.

Murphy, K. (2017). Challenging the 'invariance'thesis: procedural justice policing and the moderating influence of trust on citizens' obligation to obey police. *Journal of Experimental Criminology, 13*(3), 429-437.

Nowland, R., Necka, E. A., & Cacioppo, J. T. (2018). Loneliness and social internet use: pathways to reconnection in a digital world?. *Perspectives on Psychological Science, 13*(1), 70-87.

Pate, J. (2006). Seller reputation as a determinant of price in online auction: Theory and evidence from gift card sales. *Retrieved March, 23*, 2010.

Pavlou, Paul, and Dimoka Dimoka. 2006. "The Nature and Role of Feedback Text Comments in Online Marketplaces- Implications for Trust Building, Price Premiums, and Seller Differentiation." *Information Systems Research* 17(4):392–414.

Peng, C. H., Yin, D., Wei, C. P., & Zhang, H. (2014). How and when review length and emotional intensity influence review helpfulness: Empirical evidence from Epinions. com.

Przepiorka, W., Norbutas, L., & Corten, R. (2017). Order without law: Reputation promotes cooperation in a cryptomarket for illegal drugs. *European Sociological Review, 33*(6), 752-764.

R Core Team. (2017). "R: A Language and Environment for Statistical Computing."

Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. *Communications of the ACM, 43*(12), 45-48.

Rocklage, M. D., & Fazio, R. H. (2020). The enhancing versus backfiring effects of positive emotion in consumer reviews. *Journal of Marketing Research, 57*(2), 332-352.

Ryan, C. L., & Lewis, J. M. (2017). Computer and internet use in the United States: 2015. *Washington, DC: US Department of Commerce, Economics and Statistics Administration, US Census Bureau.*

Schoenmueller, V., Netzer, O., & Stahl, F. (2020). The polarity of online reviews: Prevalence, drivers and implications. *Journal of Marketing Research, 57*(5), 853-877.

Soska K., Christin N. (2015). Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *24th USENIX Security Symposium (USENIX Security 15)*. Washington, DC: USENIX Association, pp. 33–48.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8, 321-340.

UCLA: Statistical Consulting Group (2021). Negative Binomial Regression - SPSS data analyses examples. From https://stats.idre.ucla.edu/sas/modules/sas-learning-moduleintroduction-to-the-features-of-sas/ (accessed April 22, 2021).

Van der Meer, T. W. (2017). Political trust and the "crisis of democracy". In *Oxford research encyclopedia of politics*.

Van Lissa, C. J. (2019). *tidySEM: A tidy workflow for running, reporting, and plotting structural equation models in lavaan or Mplus*. R package version 0.1.6. https://github.com/cjvanlissa/tidySEM/

Venables, W. N., Ripley, B. D. (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, https://www.stats.ox.ac.uk/pub/MASS4/.

Wang, Q., Wang, L., Zhang, X., Mao, Y., & Wang, P. (2017). The impact research of online reviews' sentiment polarity presentation on consumer purchase decision. *Information Technology & People*.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4), 548.

Wickham, H., & Miller, E. (2018). *haven: Import and export SPSS, Stata, and SAS files*. Retrieved from https://CRAN.R-project.org/package=haven

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A Grammar of Data Manipulation*. R package version 0.7.6. https://CRAN.R-project.org/package=dplyr

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management, 28*(1), 180-182.

Ye, Q., Li, Y., Kiang, M., & Wu, W. (2009). The Impact of Seller Reputation on the Performance of online sales: evidence from TaoBao buy-it-now (BIN) data. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 40*(1), 12-19.