ECO2803

Do User Ratings Help Mitigate Risk?

Evaluating the Importance of User Feedback on the Silk Road

Heather Bone and Samantha Cleveland
Student Numbers: 1004213763,1004486683
April 17, 2018

Contents

1 Introduction **1**

2 Silk Road Overview **2**

3 Literature Review **3**
3.1 The Economics of Darknet Markets 3
3.2 The Importance of Reputation 5

4 Data **7**

5 Model **10**

6 Results **11**

7 Conclusions and Recommendations **11**

8 Appendix A: Results with Leading Revenues **12**

1 Introduction

Often referred to as "the eBay of drugs", the Silk Road was an online anonymous marketplace known for being the first modern darknet market. It was launched in 2011 and operated until the website was successfully taken down by the FBI in 2013. Despite its reputation, the Silk Road did not exclusively feature drugs, or even illicit materials. However, due to the anonymous nature of the website, it became overwhelmingly used as a hub for such activity.

The Silk Road differed from conventional drug markets in many ways, but of marked interest to economists is its formalized user feedback feature. When the item arrived to the buyer, the buyer was required to finalize the purchase. Part of doing so required leaving feedback (a rating out of 5 stars with an option to leave additional written comments), which was publicly displayed on the item page for potential buyers to see.[4] The mandatory nature of the feedback provides us with an excellent measure of a seller's reputation for a particular item (free of voluntary response bias), as well as a proxy for the item's total number of sales.

In this paper, we use a panel data model with seller fixed effects to assess the extent to which ratings impact a seller's revenues from a drug depending on the drug's riskiness. Information on the number of overdose deaths associated with the drug's use is used to create a measure of drug riskiness. We find that for new listings, a seller is more sensitive to ratings if they are selling a high risk drug. For established listings, there is no statistically significant relationship. This makes sense, as there is more uncertainty surrounding new products and thus more risk associated with using a product that has the potential to cause harm. These findings suggest that rating systems help mitigate the bodily risk associated with illegal drugs. These results are relevant to policy makers, as many jurisdictions are beginning to consider harm-reduction approaches to drug policy in lieu of criminalization. On dark net markets like the Silk Road, where users can exchange information about drugs anonymously without fear of incarceration, sellers are incentivized to be concerned with the safety of their product.

Our paper is organized as follows: first, we provide a brief overview of the Silk Road and darknet markets, including a description of their economic function. Next, we review some of the literature pertaining to darknet markets as well as the importance of reputation, and explain how it relates to our research which attempts to bridge the gap between these two topics. In section 4, we discuss

our data and how it was collected. After outlining our model in section 5, we present our results in section 6. Finally, we discuss the implications of our findings in section 7, before presenting some concluding thoughts about our results.

2 Silk Road Overview

Darknet markets became popular following the recent development of user-friendly ways to access them, such as the Tor browser bundle, which is an interface that allows users to anonymously browse the web and is required to access the Silk Road. The term "darknet" refers to networks that can only be accessed using non-standard communication protocols, and thus, cannot be accessed with a typical web browser. To ensure anonymity of payment, the Silk Road only accepted the cryptocurrency bitcoin (BTC) as its currency of exchange.[4] Consequently, this darknet market became attractive to sellers and users of illicit products, as its anonymous nature greatly reduced the risk associated with law enforcement. Moreover, the Silk Road did not merely attract the listing of illicit materials, it explicitly promoted them. This is clearly demonstrated by a screenshot of a countdown on the website to "Four Twenty", a holiday dedicated to the use and celebration of cannabis (see figure 1).[4]

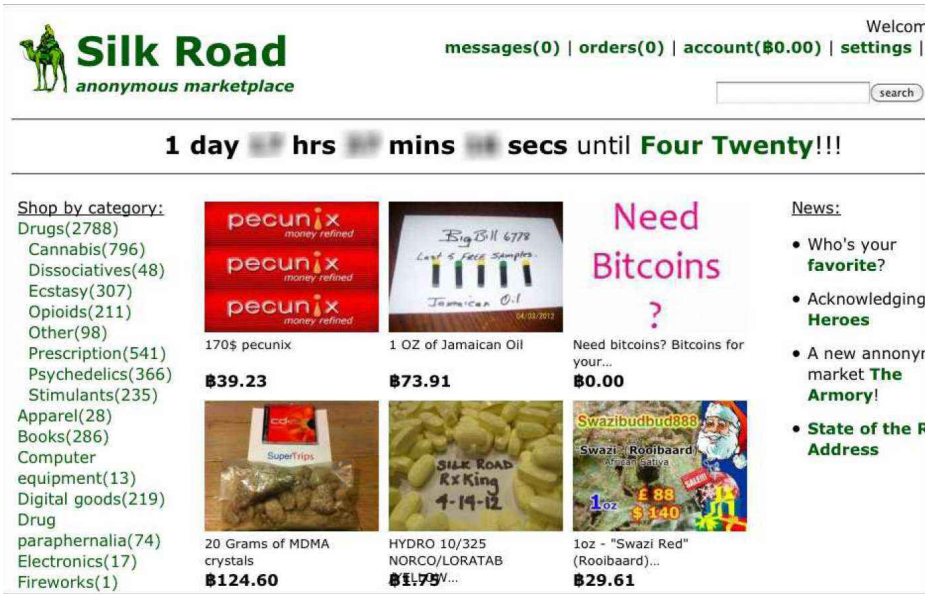

Figure 1: The front page of the Silk Road

Aside from the benefits the Silk Road provided to its users in terms of anonymity, it also reduced the risk associated with transactions in other ways. Notably, the Silk Road mandated the use of an escrow system, which not only allowed the Silk Road to accurately calculate commission fees, but also provided a means through which to solve disputes between buyers and sellers. Once the buyer received the item, they finalized the purchase, after which the money was released from the escrow to the seller. If buyers forgot to do so, the orders was automatically finalized by the Silk Road after a set period of time. Failure to use the escrow system was punishable by expulsion from the platform. Users with more than 35 successful transactions were given the ability to ask buyers to finalize early, that is, before they received the item. During this finalization process, the user was forced to leave feedback about the purchase before having even received the item. The Silk Road generally discouraged this practice, as it delegitimized the accuracy and inference of ratings.[4]

3 Literature Review

3.1 The Economics of Darknet Markets

The recent advancements in digital currencies and encryption have led to the emergence of anonymous online markets, also called cryptomarkets or darknet markets. The anonymous nature of these markets has greatly reduced the ability of law enforcement to pinpoint buyers and sellers. Consequently, cryptomarkets often specialize in the exchange of illegal products and services due to the reduced risk imposed by the anonymity of the parties involved in trade. The mysterious and illicit nature of these relatively new markets has become a topic of intrigue among researchers and policy makers. Hence, various papers have been published on this topic over the past two decades. Many of these papers have focused on characterizing and measuring the economic importance of cryptomarkets. In particular, Christin (2012) performed a measurement analysis of the Silk Road, which he describes as an "anonymous, international online marketplace that operates as a Tor hidden service and uses Bitcoin as its exchange currency" (p.1). According to his findings, the Silk Road was predominantly used for trading narcotics and controlled substances. The economic importance of this darknet market, like many other cryptomarkets, is reflected by its substantial size and large revenues: the Silk Road was very large, with approximately 30,000 to 150,000 customers, and a transaction volume of approximately USD 15 million per year. Total seller revenues amounted to over USD

1.2 million per month, while the Silk Road operators collected about USD 92 thousand in monthly commissions. Hence, these numbers suggest that darknet markets play a significant role in the shadow economy.[4]

To continue, anonymous online markets often encounter problems derived from asymmetric information, notably adverse selection and moral hazard. The latter occurs when the buyer makes a purchase and the seller receives the money but refuses to deliver the product. In the case of adverse selection, buyers are unable to distinguish between high and low quality products before making a purchase if they rely only on product descriptions, thus creating ineffective price signals. Hence, seller reputation plays an important role in these scenarios. Subsequently, many online markets, such as eBay and Amazon, allow buyers to rate a product. In fact, buyers in the Silk Road were required to leave a rating whenever they purchased a product.

A recent paper by Janetos and Tilly (2017) explores the importance of reputation as a factor for mitigating adverse selection issues. With the use of data retrieved from Agora, a popular cryptomarket for illicit products, the authors study the interaction of ratings, reviews (one to five starts, along with optional commentary) and prices for transactions related to four different drug categories. Using an advanced web-scraping mechanism, the authors were able to extract data on the quantity sold for each sale and calculated the normalized price in dollars per gram. In the Agora marketplace, buyers have the ability to impact the reputation of sellers by leaving reviews, which affect a seller's rating, as the latter is the average of all reviews left for a seller. After analyzing their data, the authors report the presence of a significantly positive, but small, relationship between price and a seller's rating. This relationship is amplified as the number of reviews left for a seller increases. Ratings have no significant impact on price for new sellers, while a mature seller with a high rating can charge a price 20 percent higher than a mature low-rated seller. They also report a positive relationship between the number of reviews left and the price charged by sellers: mature sellers (more than five thousand reviews) have prices that are on average ten percent higher than young sellers. These findings imply that adverse selection is an issue when sellers are new: buyers are aware that a seller's rating is based on very few reviews and does not reflect the true quality of the seller. Hence, ratings for new sellers have no significant impact on price. However, when the number of reviews increases, buyers attribute more importance to a seller's rating, and this is reflected in the price charged by the seller. Therefore, this paper demonstrates that a feedback

system is more effective at mitigating the risks of adverse selection as the number of reviews increases.[6]

In addition, the authors also note that sellers with lower ratings are more likely to exit the market rather than face lower prices. Furthermore, in the absence of a rating system, sellers would quickly exit and the market would collapse due to adverse selection. Hence, in addition to reporting a dynamic, positive relationship between price and a seller's rating, this paper highlights that rating systems are a crucial feature to the well-functioning of online markets. Given this documented relationship between price and user ratings in the literature, it is reasonable to suggest that sellers are affected by their reputation. Our paper builds on this research by examining whether a seller's sensitivity to ratings depends on the riskiness of a drug.

Moreover, another study by Bhaskar et al. examines the economic functioning of online drug markets by analyzing ratings, prices and exit decisions from transaction data retrieved from several darknet markets. The authors find that the reputation mechanism is efficient at mitigating the effects of moral hazard. Negative ratings are highly detrimental towards price. Consequently, the sellers who receive bad ratings prefer to exit the market instead of facing lower prices. As a result, only a small portion of sales receive poor ratings and the large majority of drugs are of high quality compared to street sales. Thus, online drug markets are associated with higher seller reputation and decreases moral hazard problems.

3.2 The Importance of Reputation

Several papers have examined the significance of reputation as a factor in determining the final bid price of common items traded on online auction sites, such as eBay and Amazon. Notably, Standifird (2002) studies the importance of an auction site's reputation as a determinant of the final bid price.[9] Auction sites serve as an intermediary between buyers and sellers to exchange personal products. The host of the auction site does not usually act as a seller on this electronic marketplace, but it does receive a fee for each auction. The reputation of the e-commerce company is determined by its ability to effectively match buyers and sellers. In order to accomplish this, the site must be able to simultaneously attract enough buyers and sellers to the electronic trading system. This is achieved if the auction site is very well-known for trading a variety of goods and

draws-in a large number of buyers and sellers in general, or if the site is able to attract many buyers and sellers for a specific product type. According to Standifird, if reputation type does matter, then an auction site with a reputation for trading a certain type of product, or attracting many buyers and sellers in general, will be able to generate a higher final bid price for this product compared to other auction sites which do not specialize in this type of product.

To test the accuracy of this hypothesis, the author compares the success of auctioning a specific good on three online auction venues: eBay, CNET and Amazon. Standifird selected these auction venues because they are each very distinct; eBay is well-known for trading a wide variety of goods, CNET is distinguished for the exchange of computer-related items, and Amazon is mainly used for the exchange of books, but one can also find various other goods on this site. Multiple Iomega Zip Disks are auctioned on all three sites; Standifird chose this item since it is widely used by the general public and it is also of specific interest to those who desire computer products. The author compares the average final bid price on all three sites and finds that it is significantly higher for eBay and CNET, where it sold for 5.25 USD and 7.00 USD respectively, than for Amazon, where it was only auctioned off for 1.32 USD.[9] Thus, these results confirm Standifird's hypothesis that the reputation of an online auction site does play a significant role in the determination of the final bid price for products.

As mentioned in the previous section, darknet markets, also called cryptomarkets, enable users to trade items while guaranteeing their anonymity through the use of untraceable cryptocurrencies. As a result, they have acquired a specific reputation for trading high-risk items, primarily illicit drugs, porn and weapons, due to the reduced risk imposed by the anonymous nature of the transactions. Thus, following Standifird's conclusions, we expect high-risk drugs to be more successful than low-risk drugs traded on the Silk Road, given the latter's reputation for trading illegal, high-risk items.

To continue, a study by Houser and Wooders (2006) examines the impact of seller and buyer reputations on price for items traded on eBay. They find that the seller's reputation has a significant effect on price, but the buyer's reputation is not significant.[5] This result justifies why our paper focuses solely on the seller's reputation.

Moreover, in an earlier study, Standifird (2001) examines the impact of an individual seller's reputation on the final bid price of items traded on eBay. According to his results, negative ratings

are strongly influential and harmful, while positive ratings only have a slight impact on price.[8] These two studies by Standifird suggest that reputation, both that of the online auction site and the individual seller, has a significant impact on the final bid price of items traded on online auction venues. However, to the best of our knowledge, no study has explored how sensitivity to ratings changes with the riskiness of the product. Hence, our paper seeks to answer this question. More generally, we also test the sensitivity of revenues to ratings, although this is not the focus of our paper.

4 Data

We use data from the original Silk Road (SR1) obtained by Christin through "crawls" of the item, user and category web pages on the site (from February to July of 2012), using computer code to gather the information.[4] To measure the relative safety of different drug types, we retrieve data on the number of overdose deaths by drug type from the Disease Control and Prevention's *WONDER* database.[1] The safety of the drug supply within a jurisdiction has been shown to be correlated with public policy decisions.[2] Thus, we dropped observations where the seller was not from the US and used American statistics on overdoses as the basis for our measure of drug riskiness.

Unfortunately, we do not have data on when the purchase was made to construct a perfect measure of revenues. To approximate sales volumes (which we use to calculate revenues), we follow Christin's method of counting the amount of feedback left for an item. However, as Christin notes, buyers typically only leave one piece of feedback per order.[4] This means that our estimates for revenues are best viewed as a lower bound.

Additionally, since ratings are left after the customer buys and receives the product, the date of sale lags behind the date of feedback. However, we do not know by how much. In fact, the amount of lag will vary by transaction. This presents a modeling issue. Figures two and three below compare average rating and average revenues per bi-weekly period with and without a lead in the count of ratings (which is equivalent to lagging prices). A bi-weekly period was chosen because 77.6 percent of the items were shipped solely to the United States (where standard domestic shipping speeds are usually under a week).[3] After leaving the buyer time to use and review the product, we believe that choosing a bi-weekly period for our panel is a sensible modeling decision. Additionally,

in our panel data model, it has the added benefit of increasing the number of observations in our sample.

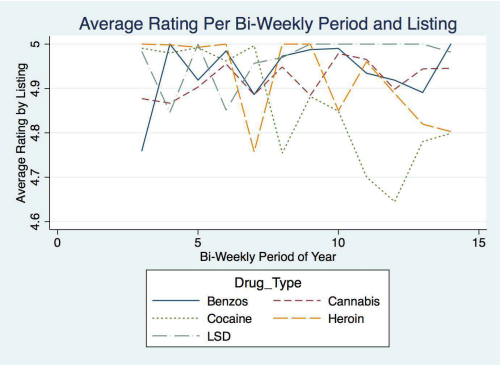

Figure 2

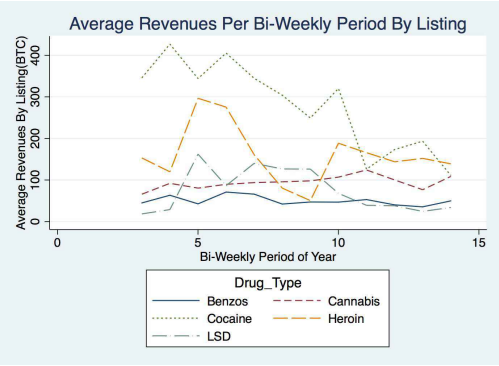

Figure 3: No Leads

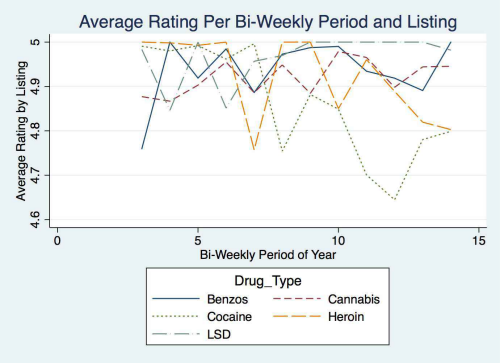

Figure 4

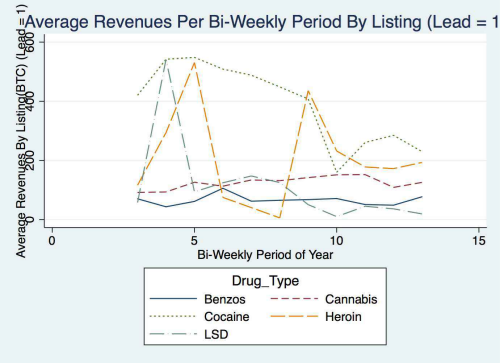

Figure 5: Lead = 1

Unfortunately, these graphs neither make the case for or against leading ratings by one period. There is not a strong visible relationship between average rating and average revenues with or without the lead. Thus, we will test the model with and without a lead for the count of ratings. The estimates using a lead for the count of ratings are presented in Appendix A.

Some summary statistics from our data are listed in Table 1 below. As expected, the most risky drug we studied was heroin, and the most safe was cannabis. Average feedback was very high for all drugs studied, and revenues were highest for cocaine.

Table 1: Mean Characteristics by Drug

	Revenues	Riskiness	Feedback Rating
Benzos $N = 513$	49.58	5.86	4.94
	(79.22)	(2.11)	(0.46)
Cannabis $N = 2621$	94.91	0.07	4.93
	(213.24)	(0.02)	(0.43)
Cocaine $N = 293$	281.26	3.37	4.87
	(468.35)	(4.11)	(0.62)
Heroin $N = 125$	156.78	8.74	4.90
	(175.60)	(17.54)	(0.47)
LSD $N = 157$	76.85	0.03	4.96
	(118.10)	(0.04)	(0.36)

Riskiness = Overdoses/Number of Listings. Standard deviation in parentheses

Additionally, central to our research question is whether or not feedback ratings actually reflect the quality of the drug (danger of the drug is one element of this) or if feedback primarily addresses shipping-related concerns. To answer this question, we looked at voluntary written feedback that was left on a darknet website similar to the Silk Road, Hansa Market, using data from a December 2016 crawl by Lewis.[7] We were unable to read written reviews from Christin's crawl as they were anonymized. We performed counts on several common words related to shipping, and several words related to the product. While we found more keywords related to shipping in the reviews, it is clear from Table 3 that a lot of reviews also commented on product quality. This lends credibility to the plausibility of our hypothesis.

Table 2: Commonly Used Words in Darknet Market Reviews

Shipping Related		**Product Related**	
Word	**Count**	**Word**	**Count**
ships	32	product	2852
shipping	1966	quality	1476
delivery	2535	stuff	898
fast	4637	as described	256
received	349	best	1093
quick	1597	amount	43
Total	11116	**Total**	6618

5 Model

In our model, revenues REV - for seller S from drug listing D for drug type T (e.g. marijuana, cocaine) during bi-weekly period B - are given by the following regression:

$$REV_{SDTB} = \beta_1 O_{TB} \cdot R_{SDTB} + \beta_2 R_{SDTB} + \beta_3 O_{TB} + \Gamma_S + \theta_B + e_{SDTB} \tag{1}$$

where revenues are a function of the number of bi-weekly feedback ratings NR and the bi-weekly average price for a particular listing:

$$REV_{SDTB} = \overline{Price_{SDTB}} * NR_{SDTB} \tag{2}$$

In our regression, O denotes the riskiness of the drug class (defined by number of overdoses standardized by the number of listings), R denotes the average rating of the drug listing over the bi-weekly period, while Γ denotes a fixed effects for seller. Θ denotes indicator variables for bi-weekly period that control for time trends in revenue. Additionally, we cluster our standard errors by bi-weekly period and seller. Thus the coefficient on the interaction between drug riskiness and average rating of the drug listing β_1 is an estimate of how sensitivity to ratings changes depending on the riskiness of the drug. We expect that if a drug poses more danger to a user, the seller's sensitivity to ratings will be higher. Based on the aforementioned findings of Janetos and Tilly, who found that there was only a positive relationship between price and ratings for established sellers, we run the regression separately for old and new listings on the website.[6]

6 Results

The results from running the regression specified in equation 1 are listed below:

Table 3: Sensitivity of Revenues to Average Ratings By Drug Riskiness

	New Listings	**Old Listings**
Rating*Riskiness	3.20*	-15.36
	(1.88)	(14.93)
Riskiness	-14.27	71.67
	(9.34)	77.91
Average Rating	-21.66**	62.32*
	(10.49)	(36.82)
N	3000	709

*** p <0.01 ** p <0.05 * p <0.1

The interaction term between average ratings and riskiness is positive and significant for new listings on the website and insignificant for older listings. New listings are those that are younger than the median age of listings in our sample. Furthermore, the difference between coefficients on the interaction term for new and old listings was shown to be statistically significant using the Hausman test at the 10 percent significance level. This means that for new listings, as a drug's riskiness increases, sellers are increasingly sensitive to changes in the average rating of their product. Newer listings may be perceived by users as more risky, thus making buyers heavily value the information provided by reviews for more dangerous items on the market.

7 Conclusions and Recommendations

Although our results counter Janetos and Tilly's findings, it is possible that users perceive newer listings as more risky, since there is less information available on these items. Consequently, buyers could attribute more importance to information provided by reviews for riskier drugs that are relatively new and less known to users.

Two possible implications can be derived from our results. To begin, disregarding any possible econometric flaws in our methodology, our statistically significant results for new listings suggest

that reputation does play an important role and helps regulate online markets to some extent. Hence, the presence of a ratings system is necessary to help mitigate the issues stemming from asymmetric information in online markets. Without these feedback features, online auction sites would have trouble matching buyers and sellers, and could eventually collapse.

However, despite the significance of our findings, they do not imply that the market can perfectly regulate itself and generate prices that wholly reflect the quality of products. A ratings system does help mitigate risk, but additional government regulations are necessary to ensure the safety of drugs if policy makers were to consider decriminalizing drugs.

Furthermore, considering the various challenges we encountered when estimating seller revenues, it is likely that our results would have been more significant had we had access to higher-quality data with information on quantity sold and the time of purchase, as this would have allowed us to estimate price per gram. While we tried implementing various lags in reviews, the truth remains that the number of reviews is by no means a perfect proxy for sales because of the uncertainty surrounding when the purchase was actually made. In addition, we had to drop all listings that were categorized as "Drugs" and "Prescription Drugs", which were the second and third most popular items on the Silk Road according to Christin (2012) [4]. Subsequently, by removing these items from our data set, we lost a considerable amount of information, and this could have weakened the significance of our results.

8 Appendix A: Results with Leading Revenues

As discussed in Section 4, it is possible that the number of reviews should be leaded by one period when constructing revenues. This gives us data on sellers, listings, prices, reviews, and drug type. In this appendix, we verify the validity of our previous results by testing an alternative model specification where the number of ratings is leaded by one period (NR_{B+1}). This leading variable is designed to account for the delay between the time when the consumer makes the purchase, and the time when the consumer leaves feedback (after receiving the product). In other words, if the purchase is made in bi-weekly period B, we assume that the feedback is left in period $B+1$, which is also when the seller receives the revenue.

The model is identical to that which was specified in equation 1, with the exception of adding

a lead for the count of ratings in our measure of revenues:

$$REV_{SDTB} = \overline{Price_{SDTB}} * NR_{SDT(B+1)} \tag{3}$$

We do not add a lead for the average rating variable R_{SDTB} in our original fixed effects regression, since we want to take into account the impact of ratings on revenues at the time when the consumer makes the purchase in period B. However, it is logical to add a lead in the count of ratings ($NR_{SDT(B+1)}$), because the latter is used as a proxy for the quantity purchased in period B, and thus is associated with the price at the time of purchase in period B.

By running this regression, we obtain the following results:

Table 4: Sensitivity of Revenues to Average Ratings By Drug Riskiness (Lead in Revenues = 1)

	New Listings	**Old Listings**
Rating*Riskiness	4.10	-44.78
	(11.59)	(27.89)
Riskiness	-16.53	27.88
	(57.89)	(161.05)
Average Rating	.83	108.92
	(6.23)	
N	1129	438

*** p <0.01 ** p <0.05 * p <0.1

With this modification, our results lose significance. This lends credibility to our theory that since most buyers reside in the United States, they received their packages within two weeks, eliminating the need for a lead in quantity. However, it is worth pointing out that one of the reasons our estimates lost significance is that the lead substantially reduced the sample size and thus our standard errors.

References

[1] Cdc wonder. *Centers for Disease Control and Prevention*.

[2] European drug report 2017. *European Monitoring Centre for Drugs and Drug Addiction*.

[3] How to send mail. *USPS*.

[4] Nicolas Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web*, pages 213–224. ACM, 2013.

[5] Daniel Houser and John Wooders. Reputation in auctions: Theory, and evidence from ebay. *Journal of Economics & Management Strategy*, 15(2):353–369, 2006.

[6] Nick Janetos and Jan Tilly. Reputation dynamics in a market for illicit drugs. *arXiv preprint arXiv:1703.01937*, 2017.

[7] Sarah Jamie Lewis. Dark web data dumps, Dec 2016.

[8] Stephen S Standifird. Reputation and e-commerce: ebay auctions and the asymmetrical impact of positive and negative ratings. *Journal of management*, 27(3):279–295, 2001.

[9] Stephen Scott Standifird. Online auctions and the importance of reputation type. *Electronic Markets*, 12(1):58–62, 2002.