



Supplement to:

Macanovic, Ana, Wojtek Przepiorka, and Vincent Buskens. 2026. "Governing the Agora—Informal Communication in Reputation-Based Online Markets" *Sociological Science* 13: 1305-1328.

APPENDIX A. DATA COVERAGE AND FEEDBACK RATE ESTIMATION

As our data were obtained by scraping cryptomarket websites, we cannot exclude the possibility that some transaction data are missing from our dataset (e.g., due to changes in the webpages, technical issues, etc.). The Silk Road dataset we utilize was collected by Christin (2013), who also provides full details about data coverage. Christin (2013) finds that the number of feedback ratings mirrors the expected number of transactions.

We further parse the information from HTML pages of Silk Road forum (Branwen 2021) and, after cleaning and deduplicating these data, we obtain 850,879 forum posts written by 39,116 users in 79,070 forum threads in 17 sub-forums. The official forum statistics at the time of data collection listed 1.3 million posts, 85,145 topics and 148,381 members. We obtain our count after deduplicating posts (which could exclude, e.g., several identical posts by the same author posted within the same minute, as was frequently done by new users to gain a required number of posts before entering certain forum sections) and only counting members who have posted in the forum at some point. Further, at the time of data collection, the Silk Road marketplace was already seized by the law enforcement, which led to many members deleting their accounts and posts (Branwen 2021), which could explain the lower user and post counts we obtained. We obtain 64 percent of all posts and 93 percent of all threads ever posted in the forum. The AlphaBay data were scraped in April 2017 by Norbutas and colleagues (2020). To estimate the coverage of the AlphaBay data, we use the information on the number of sellers' feedback evaluations displayed on the seller pages. However, this information only contained feedback from the preceding 12 months, so we can only validate our coverage for data generated after April 2016. We obtain the feedback numbers from seller pages and compare them to observations in our dataset in the 12 months preceding the data collection, obtaining an estimate that our data contain approx. 85% of all feedback generated in the date frame between April 2016 and April 2017. This share is likely to be somewhat lower for earlier dates.

To evaluate the feedback rates in AlphaBay, we use the abovementioned feedback numbers combined with the total number of transactions displayed on each seller's page. However, the latter number was not restricted to the last 12 months only. We thus calculate the feedback share on a subset of sellers who joined the market within the 12 months preceding the data collection. Over the whole sample, the feedback rate is only 27%. However, we observe 34 sellers who have more than 500 sales, but feedback rates below 10% and find that they sold very low value items (e.g., hacking guides) with value below 5 USD. Items below this price were not eligible for feedback on AlphaBay. Thus, we re-estimate the feedback provision rate

without taking these sellers into account, obtaining the feedback rate of 51% on (likely) eligible transactions. The weakness of this approach is that it does not include the profiles of many older sellers in the market, for whom the feedback provision rates could be different.

When it comes to the forum data, we obtain, after deduplicating our parsed data, 1,105,393 forum posts written by 55,605 users in 97,486 forum threads in AlphaBay. The official forum statistics at the time of data collection listed 1.2 million posts, 131,444 topics and 187,996 members. In Silk Road, we obtain around 90 percent of all posts and 74 percent of all forum threads ever posted.

APPENDIX B. RELEVANT FORUM POST IDENTIFICATION, DATA CODING, AND TEXT MINING

We start by parsing the content of Silk Road (Branwen 2021) and AlphaBay (Norbutas 2020) forum HTML pages. To do so, we write custom code that extracts, for each forum thread, the contents of the forum thread, including: the title, the original author, the content of the original post, and the data of posting. We then collect the date of posting, the author, and the text of all posts that were replies to this thread. We exclude text that was quoted by the authors, to only consider the content of each individual post. We then remove any duplicate posts. In Silk Road, we ran into an additional challenge, as our data on seller transactions only contained encrypted seller usernames, but not their full usernames that the forum users would have to use to discuss them. We manually match encrypted usernames (contained in URLs leading to seller profile pages) with mentions of their usernames (using posts that mention both the seller username and link to their URL) and recover the usernames of a large share of sellers (we recover usernames of 816 out of 1017 sellers in our data). Further, we manually remove additional 33 seller usernames which were too ambiguous to reliably match to seller mentions (e.g., “sms”, “oz”, “basic”, “cc”, “only”, “mastercard”, “Europe”). In AlphaBay, we also remove any seller usernames that are too ambiguous to be reliably identified in mentions. Given the large number of sellers, we use a general English dictionary (Moby Thesaurus List by Grady Ward as per Rinker (2020)) to exclude sellers with usernames matching words in this dictionary (e.g., “optimum”, “outlet”, “jumbo”, “stripe”). However, we do manually screen excluded sellers and retain usernames of 74 who were prominent in the market, whereas their usernames were not frequently used in contexts other than seller discussions (e.g., “Frenchmen”, “chest”); in the end, we exclude 157 sellers from our dataset.

We seek mentions of usernames of sellers for whom we have transaction data for in both markets in the forum texts. We seek forum posts that mention a seller; but also look for forum threads that mention a seller’s name in their title using string search commands. To sample mentions for manual annotation, we also consider posts in which a seller has replied in the same thread they were mentioned. We sample 700 mentions and seller replies from Silk Road and 300 mentions and seller replies from AlphaBay. To ensure we sample a sufficient number of posts representative of different types of positive and negative forum discussions (Japkowicz 2000; Krippendorff 2004), we sample in the following way. In Silk Road, we sample 75 posts from the “Product offers” subforum where sellers often advertised their goods, and 535 posts from “Rumor mill”, a subforum dedicated to evaluations of sellers and their goods. As we

aimed to code a sample representative of both positive and negative evaluations of sellers, we first use string search within forum posts to determine whether the post was likely to be non-negative or negative (the latter if we identify the presence of any of the following words: beware, warning, scammer, scammed, scam, away, avoid, avoiding, warn, ignore, report, waste, shit, shitty, reporting, reported, bad). We then sample 335 texts from the former group and 200 texts from the latter group. Finally, we sample 43 seller replies from the “Product offers” section and 47 from the “Rumor mill” section.

In AlphaBay, we sample 24 seller mentions from the “Drug discussions” subforum where buyers sometimes discussed sellers trading in these goods, 115 mentions from the “Listing reviews” subforum where traders discussed how good different item listings were (33 of which were taken from threads opened by sellers, and 82 opened by other users), and 68 mentions from the “Scam reports” section of the forum, where buyers could report issues with sellers to administrators. When it comes to seller replies in threads containing mentions, we sample 24 posts from “Drug discussions”, 54 from “Listing reviews”, and 15 from “Scam reports”. The exact numbers of sampled posts result from our goal of sampling mentions/replies in each subforum proportional to the number of posts in that subforum as a share of total forum posts. Later in the analysis, we found that our original mention detection code failed to identify some sellers whose usernames contained multiple words or special characters; and we have re-extracted the mentions upon correcting the code. However, we did not resample the mentions, as our text annotation was completed by this stage. This has no systematic effect on the representativeness of our coded sample given the size of our sample and the number of sellers whose usernames were originally not identified when filtering for mentions. In our main text, we do not discuss (theoretically or otherwise) the category of seller replies to their mentions. Still, as these texts might help our machine learning algorithms correctly identify the tone in forum posts, we keep these texts in the training and hold-out validation sets.

We recruited coders from the student body of University XY. The authors also took part in coding. Upon recruitment, coders were informed that will be working on an online task that includes answering questions about short texts written in an online forum accompanying an illegal online market. Coders were then given some information about illegal online marketplaces and the purpose of forum texts in these marketplaces. We rewarded the coders with a reward equivalent to around 16-USD in vouchers for 60-minute sessions. All sessions were conducted online in a custom-made oTree environment (Chen, Schonger, and Wickens 2016) hosted on the server at University XY. Coders would first call into an online meeting to

receive a briefing about the task; participants could then proceed to work on the task at their own pace, but at least one of the researchers was always available for questions – either through email or through online calls. Before proceeding onto the task, coders had to first complete an English test to evaluate their command of colloquial English. To proceed, they had to answer at least 8 of 10 questions correctly and all our coders passed this test. The coders were then introduced to the task and given five textual examples coded by the researchers, including the explanations.

The coding scheme was simple – it only required the coders to answer a question regarding the polarity of the forum text in question:

- **How negative/positive do you feel this text is?**
 - *Instructions: Negative texts express a bad experience or a negative, unfavorable attitude. Positive texts express a good experience or a positive, favorable attitude.*
- 1 – negative 2 – neutral 3 – positive

Below we provide one of the examples given to coders in the instructions for the coding:

Text: "Received my package last friday. Shipping and stealth were both pretty good. I don't use coke but my customers liked the product. It came rocked & shimmering. No complaints. also seemed to be a tad over weight"

Coding decision: "This text is expressing a positive experience (shipping was good, so was the packaging - stealth; the order was overweight). We thus evaluate this text as positive."

In this manner, we obtain three independent evaluations of whether each of the 1000 texts was positive, neutral, or negative. We measure intercoder agreement using Krippendorff's alpha which measures the degree of coder agreement adjusted for chance (Krippendorff 2004). Additionally, we use the alpha that controls for the ordinal character of our coding categories. We achieve Krippendorff's alpha of 0.75 for text polarity, which is satisfactory (Landis and Koch 1977). We obtain the final manually coded set by applying the majority rule – that is, we code each text for the polarity value assigned by at least two out of three coders.

To expand our manual coding onto the remaining mentions, we use deep learning transformer language models (Do, Ollion, and Shen 2022). Transformer models represent words as vectors in a high-dimensional space, positioning words that occur in similar contexts closer to each other in the vector space (Mikolov, Chen, et al. 2013; Mikolov, Sutskever, et al. 2013; Vaswani et al. 2018). The particular models we use represent these word vectors so that they also capture the fact that the same word can have different meanings across different contexts (e.g., bank as a financial institution is not the same as a riverbank). We rely on the RoBERTa model which was first pre-trained on the corpus of English Wikipedia and the BookCorpus to obtain the

word vectors (Devlin et al. 2019; Liu et al. 2019). This pretraining step is done using self-supervised learning, a method in which a small portion of the text in the training corpus (e.g., a word) is hidden from the model, and the model then attempts to predict the missing portion while minimizing the prediction error (Do et al. 2022)

We “fine-tune” RoBERTa model to let it learn the associations between word vectors in our texts and the coding decisions (Sun et al. 2019). This is done by adjusting the model layers, including its vector representations and the final layer called the classification head, to the task at hand. RoBERTa is able to combine individual word vectors into a single text-level vector; the classification head of the model finally outputs a probability that each text is positive, neutral, or negative (Merchant et al. 2020). To do this, we use the *huggingface* python library (Wolf et al. 2020) and feed the data into the models in batches of eight (Masters and Luschi 2018); that is, we let the model update all its parameters after learning from eight texts. We train our model for five epochs, that is, have it pass through all of the coded texts five times (Vaswani et al. 2018). We set the learning rate to $2e-5$ (Sun et al. 2019) and try to maximize model accuracy. We train the model on a random subsample containing 80% of our data; and then evaluate its performance on the remaining 20% of the data. Our fine-tuned RoBERTa model achieves accuracy of .82 on average. In Table B1 below, we show additional performance measures.

Table B1. Detailed performance of the fine-tuned RoBERTa model

Measure	Performance
Accuracy / Micro-F	0.820
Macro-F	0.818
Weighted F	0.821

RoBERTa model processes text sequences of up to 512 tokens (i.e., words or parts of words); any text beyond this cutoff will be discarded upon processing. In Table B2 we show that this does not present an issue in our case, as most of the forum texts that mention sellers contain less than 512 tokens.

Table B2. Details on token counts in Silk Road and AlphaBay forum texts with seller mentions

Measure	Silk Road	AlphaBay
Total texts	59 510	198 531
Mean tokens per text	141.431	82.932
SD	276.824	244.359
<i>Token distributions: percentiles</i>		
25%	32	15
50%	69	34
75%	145	79
90%	296	171
95%	470	273
99%	1318	768

APPENDIX C. DIFFERENT INDEPENDENT VARIABLE SPECIFICATIONS

It was rather common for longtime trustworthy sellers to suddenly start abusing buyers' trust in cryptomarkets (performing the so-called "exit scam" and disappearing with buyer funds without dispatching the goods; see Moeller and colleagues (2017). To capture the possibility that buyers were aware of this danger and preferred to rely on more recent ratings, in Table C1 we evaluate models including seller's ratings and mentions in the week preceding the current week (rather than including total ratings as in Table 3).

In Silk Road, we find that last week's ratings and mentions show a lower correlation coefficient ($r = 0.39$ for positive ratings and mentions and $r = 0.16$ for nonpositive ratings and negative mentions, $p < 0.001$ for both) compared to sellers' total ratings and mentions ($r = 0.56$ for positive ratings and mentions and $r = 0.40$ for nonpositive ratings and negative mentions, $p < 0.001$ for both). In AlphaBay, we similarly find that last week's ratings and mentions show a lower correlation coefficient ($r = 0.14$ for positive ratings and mentions and $r = 0.17$ for nonpositive ratings and negative mentions, $p < 0.001$ for both) compared to sellers' total ratings and mentions ($r = 0.22$ for positive ratings and mentions and $r = 0.22$ for nonpositive ratings and negative mentions, $p < 0.001$ for both).

Table C1 shows that results remain robust even if we consider more recent ratings in our models, with a few exceptions: in Model 3, we find that the coefficient on the effect of positive ratings on the likelihood of market exit is statistically significant in the model with only last week's ratings included; whereas this coefficient is not statistically significant in our main model with total ratings. Additionally, in Model 2 in Table C1, last week's nonpositive ratings have a significant negative association with revenues, which is not the case for total nonpositive ratings.

In Table C2, we show comparable models, but now considering total positive and negative mentions, rather than only the last week's ones. In these models, we consider the possibility that buyers consider histories of seller mentions rather than last week's discussions only. Note that, in these models, we also consider sellers' total number of forum posts – rather than the last week's as in our main models – for consistency. We see several differences compared to models in Table 3 and Table C1. First, we see that, unlike in Model 3 in Table 3, the association between the total number of positive mentions and seller likelihood of market exit in Silk Road is not statistically significant in this mode. Further, the association between the total number of negative mentions and the hazard of being banned in AlphaBay is not statistically significant

in models in Table C2. This differs from our conclusions from models in Tables 3 and C1, where we find statistically significant association between last week's negative mentions and market exclusion. These results suggest that it is not the total negative mentions, but only the recent reports, that increase the likelihood that the administrator will sanction the seller. This is not surprising, as negative mentions often functioned as reports of seller misdeeds that triggered investigation and, when necessary, rather quick market exclusion. We similarly fail to find a statistically significant association between total negative mentions and sales in Silk Road; while we do find such an association in Table 3. Finally, unlike last week's post counts, the seller's total posts are not significantly associated with several outcomes in models in Table C2. That is, it appears that sellers needed to keep being active in the forum in order to attract sales and remain in the market.

In Table C3, we include an additional control into the models from Table 3: the total number of seller mentions up until the week preceding the week in question. We deduct the last week's mentions (which we include as our covariates of interest) from the total count. This variable captures sellers' overall salience in the forums. Controlling for this might be important if sellers who were more discussed in the forums in the past might be easier to recall for buyers when making a purchasing decision at the moment. Our main results remain robust to the inclusion of this control, with the exception of the association between total positive ratings and revenues in Silk Road, which ceases to be statistically significant upon the inclusion of this control in Table C3. Including the total number of mentions into our models improves the model fit slightly in models with last week's average revenues as the dependent variable but results in a poorer fit in models with last week's sales as the dependent variable. Including this control variable results in a lower log likelihood in the Silk Road survival model; but a higher one in AlphaBay survival models. Table C3 shows that a higher number of total mentions up to the week preceding the week in question is associated with an increase in average weekly revenue in both markets, and a decrease in market exit in AlphaBay. However, we also see that a higher total number of mentions up to the week preceding the current week is associated with a decrease in sales in AlphaBay.

Table C1. Regression models of weekly sales (1, 4) and average weekly revenue (2, 5) with time and seller FE and Cox Proportional Hazards survival models of seller hazard of market exit (3, 6) and market exclusion (7) for two markets; all ratings and mentions are last weeks' only

	(1) Silk Road Sales	(2) Silk Road Revenues	(3) Silk Road Exit Hazard	(4) AlphaBay Sales	(5) AlphaBay Revenues	(6) AlphaBay Exit Hazard ^c	(7) AlphaBay Excl. Hazard
Avg. weekly revenue ^a	0.239*** (0.018)		0.149* (0.063)	0.107*** (0.003)		0.127*** (0.011)	0.365*** (0.065)
Positive ratings ^{a, b}	0.333*** (0.018)	0.038** (0.013)	-0.295*** (0.076)	0.504*** (0.005)	0.043*** (0.006)	-0.550*** (0.022)	-0.591*** (0.075)
Nonpositive ratings ^{a, b}	0.030 (0.021)	-0.037* (0.016)	0.271 (0.142)	0.067*** (0.006)	-0.065*** (0.007)	0.673*** (0.035)	1.666*** (0.079)
Positive forum mentions ^{a, b}	0.093*** (0.020)	0.060*** (0.017)	0.100 (0.141)	0.045*** (0.008)	0.032** (0.011)	-0.238** (0.073)	-0.493 (0.201)
Negative forum mentions ^{a, b}	-0.032 (0.026)	-0.011 (0.022)	0.393* (0.178)	-0.058*** (0.008)	-0.016 (0.010)	0.449*** (0.048)	0.756*** (0.112)
Total unique items sold ^{a, b}	0.088*** (0.026)	0.271*** (0.026)	0.0001 (0.099)	-0.020*** (0.006)	0.521*** (0.010)	-0.019 (0.026)	0.363* (0.125)
Seller's forum posts ^{a, b}	0.080*** (0.016)	0.086*** (0.016)	-0.127 (0.089)	0.063*** (0.005)	0.091*** (0.009)	-0.093** (0.030)	-0.202 (0.130)
Seller's age in weeks	-0.011*** (0.002)	-0.021*** (0.002)		-0.004 (0.003)	0.004 (0.004)		
Seller FE		Y	Y	Y	Y	Y	Y
Time FE	Y	Y	N/A	Y	Y	N/A	N/A
N	9,555	9,555	9,555	151,448	151,448	151,448	151,448
Within R2	0.321	0.136		0.389	0.154		
Log Likelihood			-1,365.305			-24,722.000	-798.902

^a The variable has been log-transformed; ^b The variable has been lagged by one week; ^c Excluding the cases where sellers were forcefully excluded
Standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Table C2. Regression models of weekly sales (1, 4) and average weekly revenue (2, 5) with time and seller FE and Cox Proportional Hazards survival models of seller hazard of market exit (3, 6) and market exclusion (7) for two markets; all ratings and mentions are the totals up until the current week

	(1) Silk Road Sales	(2) Silk Road Revenues	(3) Silk Road Exit Hazard	(4) AlphaBay Sales	(5) AlphaBay Revenues	(6) AlphaBay Exit Hazard ^c	(7) AlphaBay Excl. Hazard
Avg. weekly revenue ^a	0.256*** (0.021)		0.145* (0.066)	0.111*** (0.004)		0.128*** (0.012)	0.566*** (0.066)
Total positive ratings ^{a, b}	0.088** (0.032)	0.016 (0.020)	0.023 (0.082)	0.286*** (0.011)	0.147*** (0.011)	-0.296*** (0.020)	-0.521*** (0.071)
Total nonpositive ratings ^{a, b}	-0.006 (0.030)	-0.033 (0.025)	0.141 (0.086)	-0.029** (0.010)	-0.143*** (0.010)	0.296*** (0.025)	1.898*** (0.088)
Total positive forum mentions ^{a, b}	0.097** (0.036)	0.117*** (0.031)	-0.002 (0.093)	0.056*** (0.015)	0.005 (0.014)	-0.118*** (0.030)	-0.310* (0.110)
Total negative forum mentions ^{a, b}	-0.105* (0.043)	-0.015 (0.035)	0.077 (0.105)	-0.078*** (0.014)	-0.003 (0.012)	0.115*** (0.029)	0.247 (0.098)
Total unique items sold ^{a, b}	0.312*** (0.052)	0.223*** (0.036)	-0.009 (0.046)	-0.046* (0.018)	0.360*** (0.018)	0.003 (0.032)	0.036 (0.129)
Total seller's forum posts ^{a, b}	0.037 (0.025)	0.078*** (0.023)	-0.279* (0.121)	0.011 (0.009)	0.093*** (0.011)	-0.024 (0.014)	-0.077 (0.060)
Seller's age in weeks	-0.024*** (0.003)	-0.023*** (0.003)		-0.011*** (0.003)	0.002 (0.004)		
Seller FE			Y	Y	Y	Y	Y
Time FE	Y	Y	N/A	Y	Y	N/A	N/A
N	9,555	9,555	9,555	151,448	151,448	151,448	151,448
Within R2	0.238	0.141		0.201	0.162		
Log Likelihood			-1,373.896			-25,119.300	-836.678

^a The variable has been log-transformed; ^b The variable has been lagged by one week; ^c Excluding the cases where sellers were forcefully excluded
Standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.00

Table C3. Regression models of weekly sales (1, 4) and average weekly revenue (2, 5) with time and seller FE and Cox Proportional Hazards survival models of seller hazard of market exit (3, 6) and market exclusion (7) for two markets; including mentions control variable.

	(1) Silk Road Sales	(2) Silk Road Revenues	(3) Silk Road Exit Hazard	(4) AlphaBay Sales	(5) AlphaBay Revenues	(6) AlphaBay Exit Hazard ^c	(7) AlphaBay Excl. Hazard
Avg. weekly revenue ^a	0.249*** (0.021)		0.143 (0.066)	0.108*** (0.004)		0.124*** (0.012)	0.550*** (0.066)
Total positive ratings ^{a, b}	0.110*** (0.032)	0.031 (0.021)	0.022 (0.080)	0.296*** (0.011)	0.151*** (0.011)	-0.290*** (0.020)	-0.490*** (0.074)
Total nonpositive ratings ^{a, b}	-0.010 (0.029)	-0.039 (0.024)	0.121 (0.084)	-0.027** (0.010)	-0.144*** (0.010)	0.287*** (0.024)	1.761*** (0.093)
Positive forum mentions ^{a, b}	0.154*** (0.024)	0.059*** (0.017)	-0.066 (0.143)	0.130*** (0.012)	0.030** (0.011)	-0.276*** (0.074)	-0.469* (0.191)
Negative forum mentions ^{a, b}	-0.016 (0.028)	-0.020 (0.021)	0.377* (0.181)	-0.070*** (0.011)	-0.011 (0.010)	0.677*** (0.048)	0.783*** (0.112)
Total unique items sold ^{a, b}	0.292*** (0.051)	0.220*** (0.036)	-0.270* (0.120)	-0.052** (0.018)	0.369*** (0.018)	-0.002 (0.032)	-0.034 (0.132)
Seller's forum posts ^{a, b}	0.099*** (0.019)	0.086*** (0.016)	-0.122 (0.091)	0.123*** (0.007)	0.094*** (0.009)	-0.090** (0.030)	-0.282 (0.134)
Seller's total mentions ^{a, b, d}	-0.006 (0.023)	0.092*** (0.020)	0.024 (0.061)	-0.032*** (0.009)	0.025** (0.008)	-0.069*** (0.018)	-0.169* (0.061)
Seller's age in weeks	-0.022*** (0.003)	-0.024*** (0.002)		-0.010*** (0.003)	0.003 (0.004)		
Seller FE		Y	Y	Y	Y	Y	Y
Time FE	Y	Y	N/A	Y	Y	N/A	N/A
N	9,555	9,555	9,555	151,448	151,448	151,448	151,448
Within R2	0.248	0.141		0.208	0.161		
Log Likelihood			-1,371.405			-25,043.820	-814.396

^a The variable has been log-transformed; ^b The variable has been lagged by one week; ^c Excluding the cases where sellers were forcefully excluded; ^d Excluding last week's mentions; Standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

APPENDIX D. DIFFERENT MARKET EXIT CUT-OFF SPECIFICATIONS

In models in Table 3, we consider a seller to have exited the market if their last received feedback (or last item listing posted on the market website) was more than 30 days before the last observation collected in our data. We considered this a reasonable cutoff, given that sellers sometimes went on short holidays or breaks, but were usually not returning to the market if inactive for long time periods. Reading through Silk Road and AlphaBay forums, it was common for sellers to, at times, go into the “vacation mode”, taking breaks from their activities for several weeks, after which they would return.

Table D1 shows comparison of Cox PH models with different criteria for market exit in Silk Road; Tables D2 and D3 do the same for AlphaBay voluntary and forceful exits. All of these models account for time-varying nature of the independent variables. Table D1 shows that the Model 4 has the lowest log likelihood, indicating that modelling market exit with our data shows the best fit to seller exit having occurred at least 60 days before the last date in our data. We see the same for the models in tables D2 and D3 pertaining to seller exit in AlphaBay.

Our results are somewhat less robust to changes in the market exit cutoff in Silk Road (Table D1), primarily with regard to the effect of total positive ratings, but also negative mentions (see models (1) and (3)). Therefore, we are somewhat less confident about the robustness of the effects of last week’s negative forum mentions on the likelihood of seller market exit. In our additional checks (omitted for brevity), we find that the effect of negative forum mentions remains statistically significant in models where we include last week’s ratings (instead of totals as in Table D1) for all cutoffs, except the 15-day cutoff.

When it comes to AlphaBay, all results remain robust to changes in the market exit cutoff specification for both market exit (Table D2) and market exclusion (Table D3), albeit models with a 60-day cutoff appear to be a better fit in both cases.

Table D1. Comparison of Cox PH models with different market exit cutoffs in Silk Road

	(1) 15 days	(2) 30 days	(3) 45 days	(4) 60 days
Logged average weekly revenue	0.167** (0.053)	0.148* (0.065)	0.159* (0.072)	0.062 (0.079)
Total positive ratings ^{a, b}	-0.046 (0.065)	0.029 (0.078)	0.131 (0.086)	0.193* (0.094)
Total nonpositive ratings ^{a, b}	0.097 (0.069)	0.125 (0.083)	0.090 (0.093)	0.091 (0.101)
Positive forum mentions ^{a, b}	0.002 (0.114)	-0.055 (0.140)	0.019 (0.148)	0.018 (0.158)
Negative forum mentions ^{a, b}	0.235 (0.154)	0.389* (0.179)	0.358 (0.193)	0.490** (0.196)
Total unique items sold ^{a, b}	-0.102 (0.098)	-0.272* (0.120)	-0.370** (0.133)	-0.363* (0.143)
Seller's forum posts ^{a, b}	-0.058 (0.071)	-0.116 (0.090)	-0.103 (0.097)	-0.127 (0.105)
Observations	9,555	9,555	9,555	9,555
Exit events	344	235	192	164
Log Likelihood	-1,995.431	-1,371.482	-1,130.377	-972.408

Standard errors in parentheses; ⁺ p < 0.1 * p < 0.05, ** p < 0.01, *** p < 0.001

Table D2. Comparison of Cox PH models with different market exit cutoffs in AlphaBay – market exit excluding market exclusion

	(1) 15 days	(2) 30 days	(3) 45 days	(4) 60 days
Logged average weekly revenue	0.113*** (0.011)	0.121*** (0.012)	0.122*** (0.012)	0.126*** (0.013)
Total positive ratings ^{a, b}	-0.278*** (0.017)	-0.295*** (0.020)	-0.307*** (0.021)	-0.312*** (0.021)
Total nonpositive ratings ^{a, b}	0.194*** (0.020)	0.259*** (0.023)	0.297*** (0.024)	0.327*** (0.025)
Positive forum mentions ^{a, b}	-0.337*** (0.068)	-0.325*** (0.073)	-0.347*** (0.076)	-0.339*** (0.077)
Negative forum mentions ^{a, b}	0.572*** (0.044)	0.638*** (0.047)	0.649*** (0.048)	0.669*** (0.049)
Total unique items sold ^{a, b}	0.025 (0.028)	-0.002 (0.032)	0.019 (0.034)	0.011 (0.036)
Seller's forum posts ^{a, b}	-0.146*** (0.028)	-0.119*** (0.029)	-0.107** (0.031)	-0.107** (0.031)
Observations	151,448	151,448	151,448	151,448
Exit events	3901	3158	2872	2678
Log Likelihood	-30,524.150	-25,051.440	-22,887.710	-21,389.190

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

Table D3. Comparison of Cox PH models with different market exit cutoffs in AlphaBay – market exclusion

	(1) 15 days	(2) 30 days	(3) 45 days	(4) 60 days
Logged average weekly revenue	0.529*** (0.065)	0.529*** (0.065)	0.536*** (0.066)	0.536*** (0.066)
Total positive ratings ^{a, b}	-0.494*** (0.070)	-0.498*** (0.071)	-0.506*** (0.073)	-0.506*** (0.073)
Total nonpositive ratings ^{a, b}	1.694*** (0.091)	1.714*** (0.092)	1.732*** (0.093)	1.732*** (0.093)
Positive forum mentions ^{a, b}	-0.523* (0.184)	-0.510* (0.184)	-0.501* (0.184)	-0.501* (0.184)
Negative forum mentions ^{a, b}	0.752*** (0.110)	0.748*** (0.110)	0.733*** (0.111)	0.733*** (0.111)
Total unique items sold ^{a, b}	-0.010 (0.131)	-0.039 (0.132)	-0.025 (0.134)	-0.025 (0.134)
Seller's forum posts ^{a, b}	-0.363* (0.131)	-0.356* (0.131)	-0.359* (0.133)	-0.359* (0.133)
Observations	151,448	151,448	151,448	151,448
Exit events	158	156	154	154
Log Likelihood	-829.583	-818.143	-802.280	-802.280

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

APPENDIX E. ASSUMPTIONS ON THE INVARIANCE OF EFFECTS IN COX MODELS

The Cox Proportional Hazards models in Table 3 and Appendixes C and D rest on the assumption that the effects of covariates on the hazard of market exit remain invariant over time (Grambsch and Therneau 1994). In Table E1 we show that the proportional hazards assumption is violated for all of our models.

Table E1. Tests for the proportional hazards assumptions of our models

	Model 3 (Silk Road market exit)	Model 6 (AlphaBay market exit)	Model 7 (AlphaBay market exclusion)
chi-square	20.262	217.11	26.874
df	7	7	7
p-value	0.00500***	<0.00001***	0.00035***

The test for violations of the proportional hazards assumption is especially sensitive on our AlphaBay dataset that contains a large number of observations. To understand the extent to which the deviations from the assumption of the invariance of the effect of each covariate systematically affect our conclusions, we start with graphical diagnostics by plotting Schoenfeld residuals against market exit times (Grambsch and Therneau 1994). Figures E1-E2, E3-E4, and E5-E6 show these plots for Models 3, 6, and 7 from Table 3, respectively. We first plot full residual plots, and then, for each model, also zoom in on the smoothing spline to understand the nature of deviations more generally. These plots allow us to test for the independence of residuals and time: observing a significant pattern over time would imply a systematic violation of the proportional hazards assumption. Additionally, in Tables E2-E4 we fit a Cox models with a step function – i.e., fitting separate coefficients for covariate effects over different time intervals (Therneau, Crowson, and Atkinson 2023). We fit separate coefficients for different periods only for covariates that violate the assumption of invariance over time (as per Figures E1, E3, and E5).

In our Silk Road model, we observe rather small deviations from the invariance of the effect assumption for most covariates (see Figure E1 and the scale of the Y axis in Figure E2). It appears that total positive ratings have a smaller protective effect on seller's likelihood of market exit in the early stages of seller's market existence. Recall, however, that we do not find that their effect on the seller survival is statistically significant in Table 3. Negative ratings appear more harmful in early stages of seller tenure, but this effect is also not statistically

significant. More notably, we find that increases in negative mentions increase seller's likelihood of market exit in the earlier stages of seller's lifetime, but this effect then decreases as seller remains in the market longer.

Figure E1. Scaled Schoenfeld residuals and a solid smoothing spline to the plot for Silk Road market exit model (Model 3 in Table 3). The dashed lines represent a ± 2 -standard-error band.

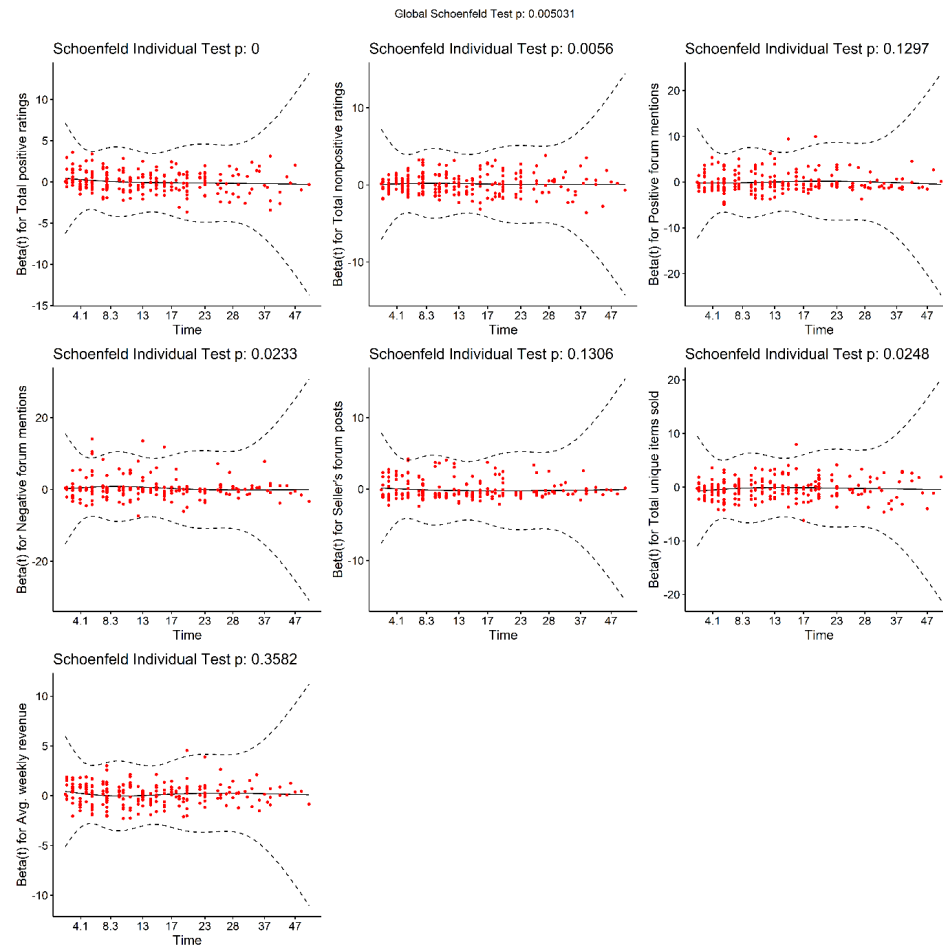
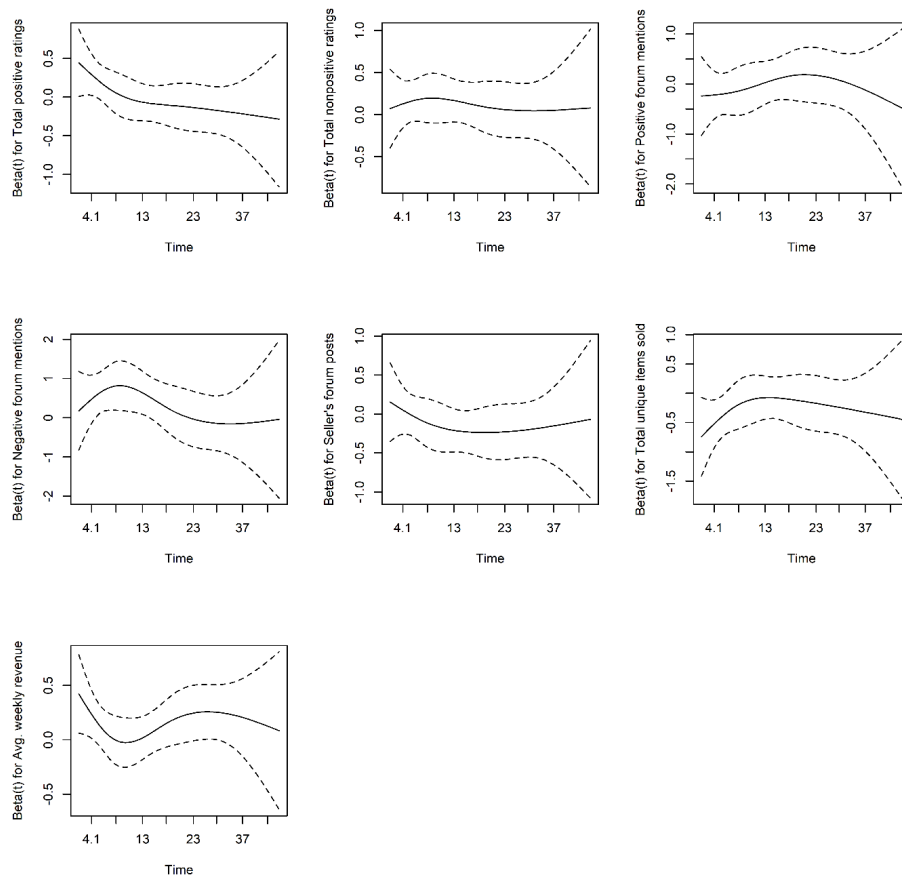


Figure E2. Solid smoothing spline to the plot for Silk Road market exit model (Model 3 in Table 3). The dashed lines represent a ± 2 -standard-error band.



Our descriptive analysis shows that most sellers exited Silk Road after up to 15 weeks since their entry. The risk then flattens out, before we observe another drop at 30 weeks, with only a small number of sellers surviving beyond this time step. Our results from a Cox model with a step function in Table E2 suggest that the statistically significant association between negative forum mentions and the likelihood of market exit in Silk Road persists only in the first 15 weeks of seller's market tenure, after which it ceases to be statistically significant. The remainder of our conclusions regarding the covariates of interest remains unchanged compared to the results from models in Table 3.

Table E2. Cox Proportional Hazards survival models of seller exit in Silk Road and a comparable model with a step function and cutoffs at 15 and 30 weeks since market entry

	(1) Model 3	(2) Model 3 with a step function
Logged average weekly revenue	0.148* (0.065)	0.141* (0.065)
Total positive ratings ^{a, b}	0.029 (0.078)	
Until week 15		0.112 (0.104)
Until week 30		-0.106 (0.133)
From week 30		-0.365 (0.243)
Total nonpositive ratings ^{a, b}	0.125 (0.083)	
Until week 15		0.161 (0.121)
Until week 30		0.222 (0.135)
From week 30		0.218 (0.256)
Positive forum mentions ^{a, b}	-0.055 (0.140)	-0.085 (0.141)
Negative forum mentions ^{a, b}	0.389* (0.179)	
Until week 15		0.578** (0.199)
Until week 30		0.027 (0.378)
From week 30		-0.207 (0.743)
Total unique items sold ^{a, b}	-0.272* (0.120)	
Until week 15		-0.283 (0.169)
Until week 30		-0.252 (0.189)
From week 30		-0.342 (0.313)
Seller's forum posts ^{a, b}	-0.116 (0.090)	-0.117 (0.090)
Observations	9,555	9,667
Exit events	235	235
Log Likelihood	-1,371.482	-1,363.650

^a The variable has been log-transformed; ^b The variable has been lagged by one week; ^c The increase comes from additional cutoffs in the model with a step function. Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In the AlphaBay model with voluntary market exit (Model 6 in Table 3), we observe deviations (albeit rather small) from the invariance of the effect assumption for all covariates except for

the seller's number of posts (see Figure E3 and the scale of the Y axis in Figure E4). Most notably, we see that the positive effect of total nonpositive ratings on the likelihood of market exit decreases over time, becoming protective as sellers' tenure in the market progresses. Other covariates of our interest (total positive ratings, positive and negative forum mentions) show less substantial variations over time. This might indicate that negative ratings were not as important for market survival of tender sellers; but rather helped increase their sale numbers. Buyer might not have found small increases in negative ratings to be important enough in the light of high reputation scores and large past transaction numbers.

Figure E3. Scaled Schoenfeld residuals and a solid smoothing spline to the plot for AlphaBay market exit model (Model 6 in Table 3). The dashed lines represent a ± 2 -standard-error band.

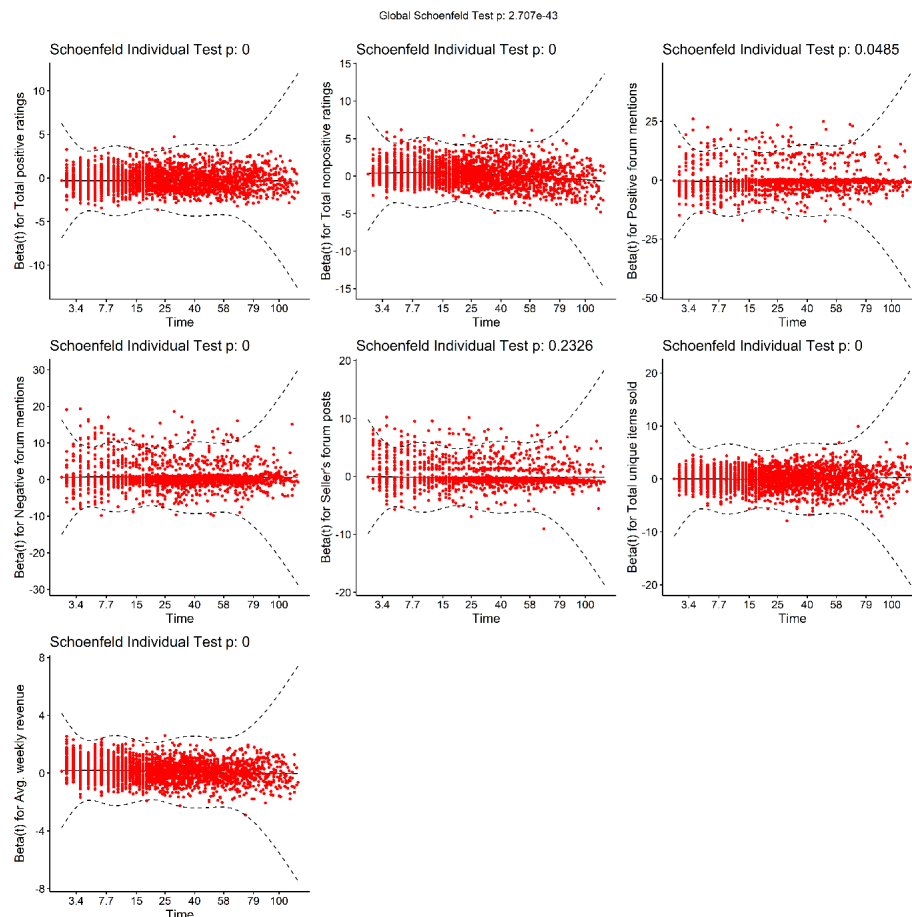
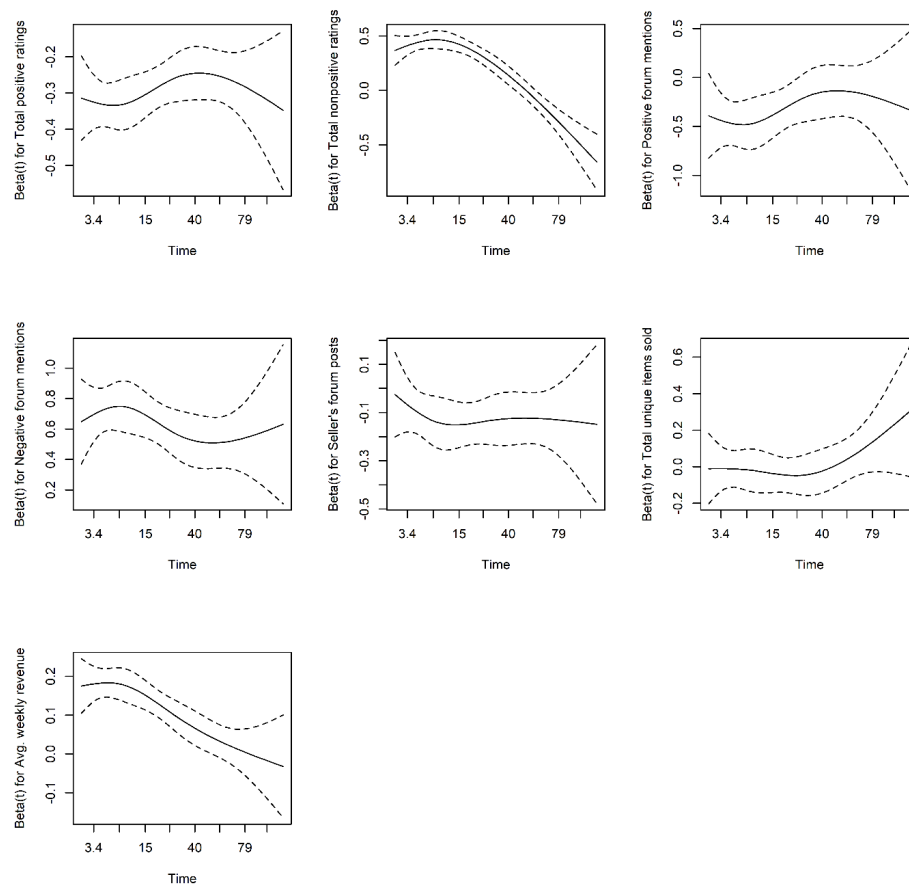


Figure E4. Solid smoothing spline to the plot for AlphaBay voluntary market exit model (Model 6 in Table 3). The dashed lines represent a ± 2 -standard-error band.



In AlphaBay, we see that the rates of seller exit flatten out after 20 weeks, and then again after 40 weeks, with a rather limited number of sellers surviving past this mark. In Tables E3 and E4 we fit Cox models with step function paralleling models 6 and 7 in Table 3 with different coefficients in the period of the first 20 weeks of seller's lifetime, between 20 and 40 weeks, and from 40 weeks onward.

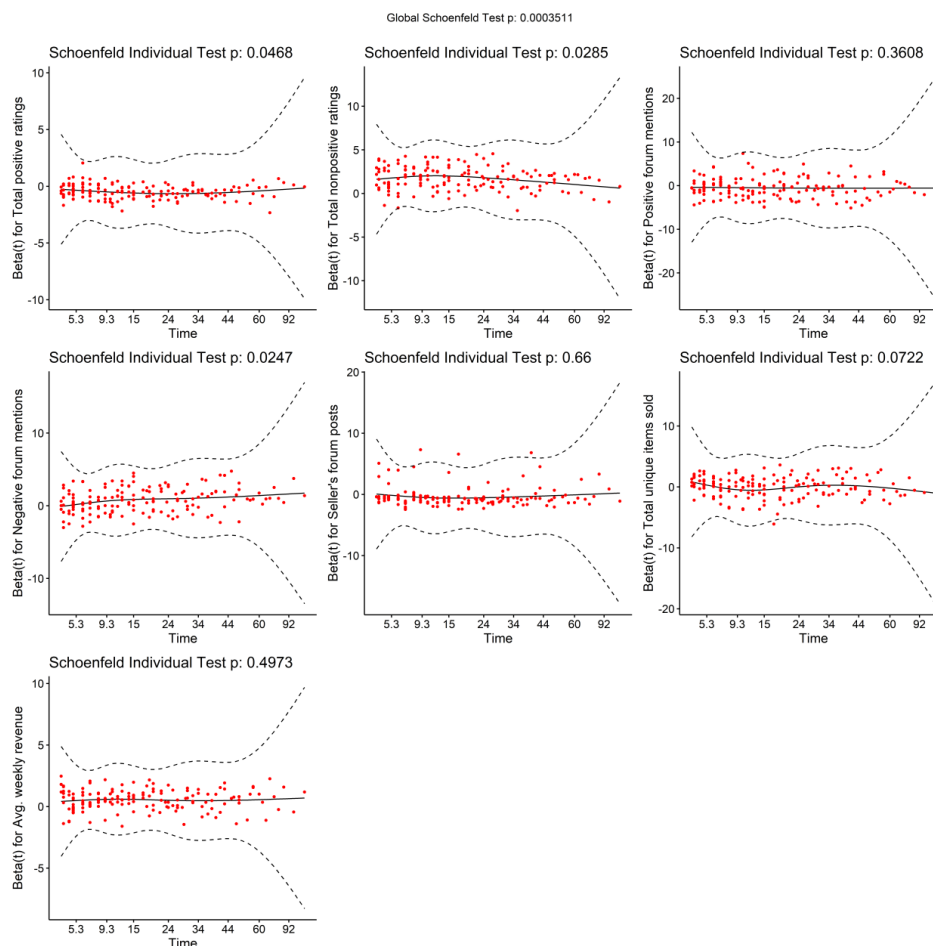
Table E3. Cox Proportional Hazards survival models of seller exit in AlphaBay and a comparable model with a step function and cutoffs at 20 and 40 weeks since market entry

	(1) Model 6	(2) Model 6 with a step function
Logged average weekly revenue	0.121*** (0.012)	
Until week 20		0.167*** (0.016)
Until week 40		0.071* (0.029)
From week 40		0.028 (0.025)
Total positive ratings ^{a, b}	-0.295*** (0.020)	
Until week 20		-0.310*** (0.033)
Until week 40		-0.262*** (0.043)
From week 40		-0.285*** (0.031)
Total nonpositive ratings ^{a, b}	0.259*** (0.023)	
Until week 20		0.543*** (0.039)
Until week 40		0.284*** (0.048)
From week 40		0.060 (0.036)
Positive forum mentions ^{a, b}	-0.325** (0.073)	
Until week 20		-0.408*** (0.101)
Until week 40		-0.192 (0.155)
From week 40		-0.189 (0.133)
Negative forum mentions ^{a, b}	0.638*** (0.047)	
Until week 20		0.628*** (0.065)
Until week 40		0.513*** (0.103)
From week 40		0.455*** (0.090)
Total unique items sold ^{a, b}	-0.1002 (0.032)	
Until week 20		-0.054 (0.056)
Until week 40		-0.080 (0.068)
From week 40		0.047 (0.050)
Seller's forum posts ^{a, b}	-0.119** (0.029)	-0.112*** (0.029)
Observations	151,448	153,268
Log Likelihood	-25,051.440	-24,972.600

^a The variable has been log-transformed; ^b The variable has been lagged by one week; ^c The increase comes from additional cutoffs in the model with a step function. Standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

When it comes to market exit (excluding forceful market exclusion) in AlphaBay, our results in Table E3 suggest that the protective effect of total positive ratings remains significant, but its' size decreases with market tenure. Last week's positive mentions have a statistically significant protective effect only within the first 20 weeks of a seller's lifetime. Total nonpositive ratings increase a seller's likelihood of exiting the market voluntarily, but this effect weakens after week 20 and stops being significant after week 40. The positive association between last week's negative mentions and market exit remains significant throughout and weakens only slightly in the last time period.

Figure E5. Scaled Schoenfeld residuals and a solid smoothing spline to the plot for AlphaBay market exclusion model (Model 7 in Table 3). The dashed lines represent a ± 2 -standard-error band.



Finally, when it comes to the AlphaBay model with market exclusion by the administrators (Model 7 in Table 3 in main text), in Figures E5 and E6 we see less significant deviations from the invariance assumption for most variables. We see that the protective effect of total positive ratings and the damaging effect of total negative ratings on the likelihood of seller exclusion decrease with seller tenure. The association between last week's negative mentions and the likelihood of exclusion from the market becomes stronger with seller's tenure in the market. This could indicate increased administrator attention to tenured sellers who showed signs of planning an exit scam or engaging in selective scamming. We fit a Cox model with a step function for market exclusion (Table E4) comparable to the model in Table E3 for voluntary market exit. Model in Table E4 confirms the conclusions based on Figures E4 and E5.

Figure E6. Solid smoothing spline to the plot for AlphaBay market exclusion model (Model 7 in Table 3). The dashed lines represent a ± 2 -standard-error band.

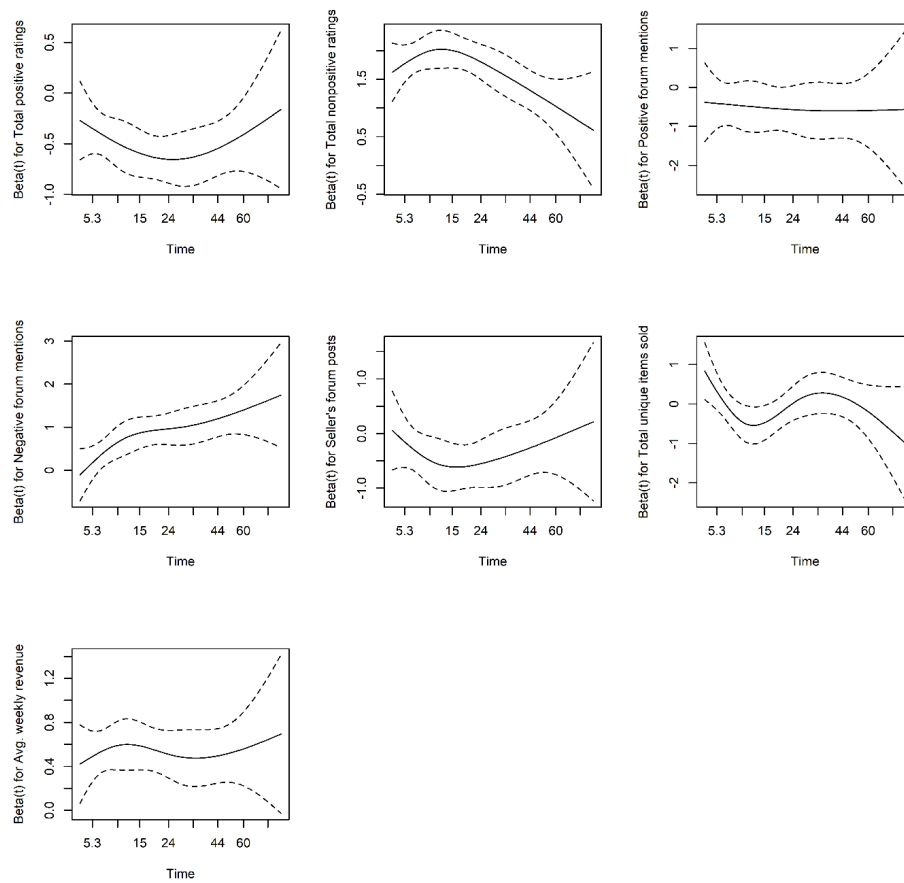


Table E4. Cox Proportional Hazards survival models of seller exclusion in AlphaBay and a comparable model with a step function and cutoffs at 20 and 40 weeks since market entry

	(1) Model 7	(2) Model 7 with a step function
Logged average weekly revenue	0.529*** (0.065)	
Until week 20		0.549*** (0.084)
Until week 40		0.402** (0.140)
From week 40		0.607** (0.165)
Total positive ratings ^{a, b}	-0.498*** (0.071)	
Until week 20		-0.447** (0.129)
Until week 40		-0.610*** (0.183)
From week 40		-0.537*** (0.105)
Total nonpositive ratings ^{a, b}	1.714*** (0.092)	
Until week 20		1.871*** (0.126)
Until week 40		1.668*** (0.199)
From week 40		1.293*** (0.198)
Positive forum mentions ^{a, b}	-0.510* (0.184)	
Until week 20		-0.482 (0.230)
Until week 40		-0.610 (0.407)
From week 40		-0.676 (0.467)
Negative forum mentions ^{a, b}	0.748*** (0.110)	
Until week 20		0.524** (0.139)
Until week 40		0.869** (0.236)
From week 40		1.328*** (0.281)
Total unique items sold ^{a, b}	-0.039 (0.132)	
Until week 20		-0.109 (0.183)
Until week 40		0.212 (0.283)
From week 40		-0.236 (0.293)
Seller's forum posts ^{a, b}	-0.356* (0.131)	-0.398** (0.134)
Observations	151,448	153,268
Log Likelihood	-818.143	-809.603

^a The variable has been log-transformed; ^b The variable has been lagged by one week; ^c The increase comes from additional cutoffs in the model with a step function. Standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

APPENDIX REFERENCES

- Branwen, Gwern. 2021. 'Darknet Market Archives (2013-2015)'. Retrieved (<https://www.gwern.net/DNM-archives>).
- Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. 'oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments'. *Journal of Behavioral and Experimental Finance* 9:88–97. doi: 10.1016/j.jbef.2015.12.001.
- Christin, Nicolas. 2013. 'Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace'. Pp. 213–23 in *WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web*.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 'BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding'. Pp. 4171–86 in *Proceedings of the 2019 Conference of the North*. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Do, Salomé, Etienne Ollion, and Rubing Shen. 2022. 'The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy'. *Sociological Methods & Research*. doi: <https://doi.org/10.31235/osf.io/3fkzc>.
- Grambsch, Patricia M., and Terry M. Therneau. 1994. 'Proportional Hazards Tests and Diagnostics Based on Weighted Residuals'. *Biometrika* 81(3):515–26. doi: 10.1093/biomet/81.3.515.
- Japkowicz, Nathalie. 2000. 'The Class Imbalance Problem: Significance and Strategies'. Pp. 111–17 in *Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI)*.
- Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology*. 2nd ed. Thousand Oaks, CA: SAGE Publications.
- Landis, J. Richard, and Gary G. Koch. 1977. 'The Measurement of Observer Agreement for Categorical Data'. *Biometrics* 33(1):159–74. doi: 10.2307/2529310.
- Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, et al. 2019. *Roberta: A Robustly Optimized Bert Pretraining Approach*. doi: <https://arxiv.org/abs/1907.11692>.
- Merchant, Amil, Elahe Rahimtoroghi, Ellie Pavlick, and Ian Tenney. 2020. 'What Happens To BERT Embeddings During Fine-Tuning?' doi: <https://doi.org/10.48550/arXiv.2004.14448>.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. 'Efficient Estimation of Word Representations in Vector Space'. in *ICLR Workshop Papers*.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeffrey Dean. 2013. 'Distributed Representations of Words and Phrases and Their Compositionality'. Pp. 3111–19 in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*.

- Moeller, Kim, Rasmus Munksgaard, and Jakob Demant. 2017. 'Flow My FE the Vendor Said: Exploring Violent and Fraudulent Resource Exchanges on Cryptomarkets for Illicit Drugs'. *American Behavioral Scientist* 61(11):1427–50. doi: 10.1177/0002764217734269.
- Norbutas, Lukas. 2020. 'TRUST ON THE DARK WEB: An Analysis of Illegal Online Drug Markets'. ICS-dissertation, Utrecht.
- Norbutas, Lukas, Stijn Ruiter, and Rense Corten. 2020. 'Reputation Transferability across Contexts: Maintaining Cooperation among Anonymous Cryptomarket Actors When Moving between Markets'. *International Journal of Drug Policy* 76:102635. doi: 10.1016/j.drugpo.2019.102635.
- Rinker, Tyler W. 2020. 'Qdap: Quantitative Discourse Analysis Package'.
- Sun, Chi, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. 'How to Fine-Tune BERT for Text Classification?' Pp. 194–206 in *Chinese Computational Linguistics*. Vol. 11856, *Lecture Notes in Computer Science*, edited by M. Sun, X. Huang, H. Ji, Z. Liu, and Y. Liu. Cham: Springer International Publishing.
- Therneau, Terry, Cynthia Crowson, and Elizabeth Atkinson. 2023. *Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model. R package Survival Vignettes*. Version 3.5-7.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, et al. 2018. 'Attention Is All You Need'. Pp. 5999–6010 in *Advances in Neural Information Processing Systems* 30.
- Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, et al. 2020. 'HuggingFace's Transformers: State-of-the-Art Natural Language Processing'. Pp. 38–45 in *Proceedings of the 2020 conference on empirical methods in natural language processing*.