Market Transparency, Adverse Selection, and Moral Hazard

Tobias J. Klein

Tilburg University

Christian Lambertz

University of Mannheim

Konrad O. Stahl

University of Mannheim, Centre for Economic Policy Research, Center for Economic Studies—IFO, and Centre for European Economic Research (ZEW)

We study how an improvement in market transparency affects seller exit and continuing sellers' behavior in a market setting that involves informational asymmetries. The improvement was achieved by reducing strategic bias in buyer ratings. It led to a significant increase in buyer satisfaction with seller performance, but not to an increase in seller exit. When sellers had the choice between exiting—a reduction in adverse selection—and staying but improving behavior—a reduction in moral hazard—they preferred the latter. Increasing market transparency led to better market outcomes.

I. Introduction

The emergence of the Internet has led to an enormous increase in transactions taking place under informational asymmetry. Examples are on-

We would like to thank the participants of the 2013 Information and Communications Technology Workshop in Evora, the second Centre for Competition and Innovation day in

Electronically published October 31, 2016
[*Journal of Political Economy*, 2016, vol. 124, no. 6]

line markets for goods, as well as hotel, restaurant, and travel services. Without remedies, moral hazard arises, for instance, as a result of insufficient provision of costly effort. Likewise, sellers may inappropriately describe and price a good—in particular when used—and may not conscientiously conduct the transaction once it has been bought. Conversely, buyers may attempt to renegotiate deals in their favor, delay payments, or default on them. Adverse selection arises when exploitative and careless buyers and sellers enter into the market and conscientious ones exit.

In this paper, we use data on buyer satisfaction with seller behavior in an anonymous product market to show that an exogenous change in market transparency led to a significant increase in buyer satisfaction but did not lead to a change in the exit rate of sellers from the market. We interpret these results through the lens of the stage game of an infinite-period model involving one seller and one buyer. In the model, sellers differ by type—they are either conscientious or exploitative—and by their disutility of providing effort. A change in the feedback system that lowers the buyer's cost of reporting a negative experience leads the seller to change his behavior alternatively in two ways: first, the exploitative seller with high disutility of effort may leave the market, thus ameliorating adverse selection; second, both seller types, if remaining in the market, may engage in more effort toward improving buyer satisfaction, thus ameliorating moral hazard. On the basis of our model and in light of our empirical results, we reason that even for the badly performing exploitative seller types, the costs of changing their behavior are small relative to the benefits of stay-

Mannheim, the eleventh Centre for European Economic Research (ZEW) Information and Communications Technology Conference, the 2013 European Association for Research in Industrial Economics conference in Evora, the Fifth Annual Conference on Internet Search and Innovation at the Searle Center in Chicago, the eighth biannual Postal Economics conference on E-commerce, the conference on Digital Economy and Delivery services in Toulouse, the Fifteenth Centre for Economic Policy Research/Journal of Industrial Economics Conference on Applied Industrial Organization in Athens, the 2015 European Summer Symposium in Economic Theory Conference in Gerzensee, seminar participants at Copenhagen, Hebrew, Tel-Aviv and Tilburg Universities, the German Institute for Economic Research (DIW), Copenhagen Business School, and in particular, Ying Fan, Jin-Hyuk Kim, Shaul Lach, John List, Volker Nocke, Martin Peitz, David Reiley, Martin Salm, Steve Tadelis, John Thanassoulis, and Stefan Weiergräber for helpful comments. Fabian Kosse provided excellent research assistance. We are grateful to the editor Ali Hortaçsu for his insightful suggestions and guidance, two anonymous referees for their comments, as well as Michael Anderson and Yossi Spiegel for unusually detailed and insightful suggestions on how to improve this paper. Finally, we would like to thank Yufeng Huang for pointing us to the eBay price data. Financial support from the Deutsche Forschungsgemeinschaft through SFB TR-15 and Microsoft through a grant to the Tilburg Law and Economics Center is gratefully acknowledged. The latter was provided in accordance with the Royal Netherlands Academy of Arts and Sciences (KNAW) Declaration of Scientific Independence (http://www.knaw.nl/Content/Internet_KNAW/Actueel/bestanden/Scientific_Independence.pdf). Data are provided as supplementary material online.

ing in the market. Otherwise, the increase in market transparency due to more accurate reports on seller behavior should have resulted in an increased rate of exit. We support this by empirical evidence on the reasons why dissatisfied buyers left negative ratings and conclude that increasing market transparency improves market outcomes primarily via a reduction in moral hazard.

The data we use are from eBay, one of the first and biggest online trading platforms to exist to date. They are well suited for studying the desired effects. First, eBay faces strong threats in terms of moral hazard and adverse selection. We document that sellers are consistently heterogeneous in their behavior. Those rated more poorly than others before the change tend to be rated better thereafter, but still more poorly than others. Second, the institutional design allows us to measure the change in seller behavior before and after an exogenous shock to market transparency. For historic reasons, there are two transaction-based measures of buyer satisfaction on eBay, classic feedbacks and detailed seller ratings (DSRs). DSRs are reported as moving averages only, so in contrast to the classic feedback scheme, the seller cannot identify an individual buyer rating. In 2008, while leaving the DSR system unchanged, eBay lowered the opportunity costs to buyers to share negative experiences via the still more prominent classic feedback system.[1] This led to an increase in market transparency whose effect we measure using the DSRs. Toward this, our key assumption, supported by the data, is that the mapping between seller behavior and DSRs was not affected by the change to the classic feedback mechanism.

Our paper is closely related to three strands of the literature. The first is on the effects of quality disclosure. In the context of restaurants, Jin and Leslie (2003) show that requiring them to display quality grade cards in their windows causes them to make hygiene quality improvements. Anderson and Magruder (2012) relate online ratings of restaurants to restaurant reservation availability and find that an extra half star on the popular platform Yelp.com causes restaurants to sell out 19 percentage points more frequently. Greenstone, Oyer, and Vissing-Jorgensen (2006) show that financial investors valued an extension of disclosure requirements by documenting abnormal returns for firms most affected by these.

Second, our paper is related to the literature on the relationship between moral hazard and adverse selection in the context of health insur-

[1] The possibility for the seller to rate the buyer negatively was removed and, with this, the threat of negative retaliation by the seller to a negative buyer rating. Klein et al. (2006) show that under the old regime, the probability of buyers leaving a negative rating increased substantially toward the end of the period in which feedback could be left at all. The explanation they provide is that buyers leave a rating late because then it is less likely that sellers retaliate by leaving a negative feedback in return.

ance. Einav et al. (2013) show that some individuals select insurance coverage in part on the basis of their anticipated behavioral response to the insurance contract and term it "selection on moral hazard." For this, they exploit variation in the health insurance options, choices, and subsequent medical utilization across different groups of workers at different points in time. Bajari, Hong, and Khwaja (2014) also study individual selection of insurance contracts. As we do in a very different context, they provide evidence of moral hazard, but not of adverse selection. Their result is based on a structural model of demand for health insurance, in which, in order to isolate selectivity ex ante and lacking observable exogenous variation, they control for individual risk and risk preference. We instead develop our results from a natural experiment, involving self-selection and adjustment of moral hazard ex post. Following sellers over time allows us to control for unobserved differences across sellers by means of fixed effects. We then study whether an improvement of the mechanism led to increased seller effort or exit from the market.

Third, our paper is related to the literature on online ratings. Bajari and Hortaçsu (2004), Dranove and Jin (2010), and Cabral (2012) provide reviews of the theoretical and empirical literature. For eBay, the general finding is that better ratings benefit sellers by an increase in the probability of selling a product and in its selling price. See, for example, Melnik and Alm (2002), Lucking-Reiley et al. (2007), and Jin and Kato (2008) for evidence using field data and Resnick et al. (2006) for experimental evidence. These results show that classic ratings on eBay convey information. Resnick and Zeckhauser (2002) provide reduced-form evidence that points toward underreporting of negative experiences. Klein et al. (2006) and Bolton, Greiner, and Ockenfels (2013) add to this evidence. Dellarocas and Wood (2008) estimate a model of rating behavior, assuming that ratings, once given, are truthful and estimate the true underlying distribution of satisfaction. This can be seen as controlling for the selection bias coming from traders that are more likely to leave a rating when satisfied. Cabral and Hortaçsu (2010) provide evidence that is consistent with seller moral hazard. They find that just before exiting, sellers on eBay receive more negative feedback than their lifetime average.

With our paper we complement the aforementioned studies by providing direct evidence on the relationship between the level of market transparency as critically influenced by the design of the feedback mechanism and the prevalence of moral hazard and adverse selection. Our key finding is that an increase in market transparency reduces buyer regret and thereby leads to higher-quality outcomes. At the same time, sellers do not leave the market. Given the small cost of implementing the observed change in the reporting mechanism discussed by us, given the significant increase in buyer satisfaction generated from that, and given that the sellers' material costs of changing their behavior are arguably small, our results suggest that this increase in market transparency had a beneficial welfare effect.[2]

Our results give guidance on how to discipline seller behavior in markets other than eBay, most notably many of the "new" online markets for goods and services. In general, we show that small changes in institutional rules may have large effects on market outcomes and, with this, on the attractiveness of online platforms from the viewpoint of potential users.

We proceed as follows. In Section II, we describe the eBay feedback mechanism and in particular the change we focus on. Section III contains the description of our data. In Section IV, we present our main results. In Section V, we develop our model, from which we derive our preferred explanation and interpretation of the results. In Section VI, we support this interpretation of our main empirical findings with additional evidence. We conclude the paper with Section VII.

II. eBay's Feedback Mechanism

In February 1996, just a few months after the first auction had taken place on its website, eBay introduced its feedback mechanism. In its earliest form, the system allowed any eBay user to leave feedback on the performance of any other user independently of any transaction, in the form of a "positive," "neutral," or "negative" rating, possibly accompanied by a textual comment. This feedback was immediately observable on his or her feedback profile page, together with all ratings and comments that a user had ever received by other users.[3]

In February 2000, 4 years after its institution, the mechanism was changed into one with only transaction-based feedback. Since then, all new ratings must relate to a particular transaction; that is, only the seller and the buyer in a particular transaction can rate each other regarding their performance in that transaction.

From early on, there have been conflicts and heated discussions about unfairly biased reports. As a consequence, eBay repeatedly modified the system. In May 2007, eBay introduced a new form of unilateral buyer ratings: detailed seller ratings (DSRs). In addition to the original bilateral rating, buyers could now separately rate, with one to five stars, the accuracy of the item description, communication, shipping speed, and shipping charges. They are made anonymous by being published in aggre-

[2] Our results strongly suggest that consumer surplus increased. However, a rigorous welfare analysis is beyond the scope of this paper, as it would require us to observe, or infer, sellers' costs as well as buyers' preferences. Only then could we compare the increase in consumer surplus to the decrease in seller rents.

[3] An early description of the basic mechanism and an analysis of rating behavior are given in Resnick and Zeckhauser (2002).

gate form only, provided that at least 10 ratings have been left in the last 12 months, so that the seller cannot identify the individual rating.[4]

This change addressed what was felt to be a substantial flaw in eBay's original bilateral feedback mechanism, namely, that the buyer had to fear retaliation when leaving a negative rating before the seller—a problem well known to many eBay users and well discussed among scholars for some time. An important detail is that DSRs can be left only when a classic rating is left. The two ratings need not be consistent, however. That is, for the very same transaction, a buyer could leave a positive classic rating identifiable by the seller and a negative, truthful set of DSRs not identifiable by him. At the same time, the two ratings are not perfect substitutes. In particular, the reported DSRs give an evaluation of the seller's behavior on average, and the classical ratings show how the seller behaved at the margin, that is, in the most recent transactions. Moreover, the most recent classic ratings are linked to the auction listings and contain a textual comment.

In May 2008, the classic bilateral feedback mechanism was transformed to effectively a unilateral one: sellers could leave only positive ratings on buyers or none at all. With this, eBay removed the possibility that the seller would strategically postpone his rating in order to implicitly or explicitly threaten the buyer with retaliation to a negative rating.[5] The timing of these two changes is depicted in figure 1. In this paper, we investigate the effect of the May 2008 change on seller behavior as measured by the DSR ratings introduced in May 2007.

It is worth noting that while anonymity of DSRs ensures that buyers can leave a DSR without threat of retaliatory feedback by the seller, the buyer's evaluation is nevertheless subjective. At the same time, however, sellers receive ratings from a large number of buyers so that the reported average DSR scores are good measures of seller behavior. Indeed, they

[4] Klein et al. (2009) provide detailed information on the actual structure of the feedback mechanism and provide first descriptive evidence on DSRs.

[5] In fact, eBay stated the reasons for this step in a public announcement in January 2008 (taken from http://announcements.ebay.com/2008/01/a-message-from-bill-cobb-new-pricing-and-other-news/): "Today, the biggest issue with the system is that buyers are more afraid than ever to leave honest, accurate feedback because of the threat of retaliation. In fact, when buyers have a bad experience on eBay, the final straw for many of them is getting a negative feedback, especially of a retaliatory nature. Now, we realize that feedback has been a two-way street, but our data shows a disturbing trend, which is that sellers leave retaliatory feedback eight times more frequently than buyers do and this figure is up dramatically from only a few years ago. So we have to put a stop to this and put trust back into the system. . . . Here's the biggest change, starting in May: Sellers may only leave positive feedback for buyers (at the seller's option)." Additional changes aiming at alleviating seller concerns about buyers' strategic abuse of feedback giving were implemented at several points in time, but not within our window of observations. For instance, in order to remove bargaining about good ratings, eBay abandoned earlier options to mutually withdraw feedback.

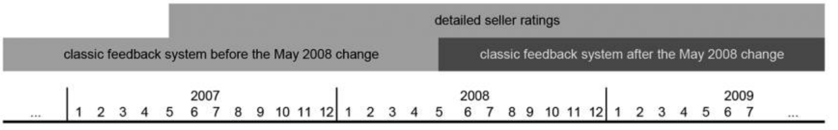

FIG. 1.—Changes to the feedback mechanism. Color version available as an online enhancement.

would be close-to-ideal measures for the purpose of this study if rating standards could be ensured to stay the same over time and if every buyer would leave a rating.

In favor of the former, eBay displays a verbal meaning next to every star rating in every category when ratings are given, which makes it more likely that the mapping from buyer satisfaction to buyer rating does indeed not systematically change over time. For instance, a rating of four stars in the rating category shipping speed means that the seller shipped the item "quickly." As for the latter, nonresponse in combination with selection bias is a threat to any survey-based empirical study. Selection bias is present if the observed average rating systematically deviates from the average report everybody has or would have given. Our approach is to follow sellers over time. Therefore, this is not a problem in our analysis, as long as the bias is the same before and after the change. In Section VI.A, we provide empirical support for this assumption. In particular, we show that the average number of ratings received and the ratio of DSRs relative to classic feedbacks received did not change substantially over time.

On the basis of these considerations, we interpret changes in the average DSR scores as unbiased measures of changes in the underlying transaction quality. We use them to investigate how individual seller performance reacts to the May 2008 change, when all ratings were effectively made unilateral, while the DSR system was left unchanged.

III. Data

Our data contain monthly information on feedback received by about 15,000 eBay users over a period of 3 years, between July 2006 and July 2009. The data were collected from eBay's US website using automated download routines and scripts to parse the retrieved web pages.

In May 2007, we drew a random sample of, respectively, 3,000 users who offered an item in one of five different categories. The categories were (1) laptops and notebooks, (2) Apple iPods and other MP3 players, (3) model railroads and trains, (4) trading cards, and (5) food and wine.[6]

[6] See table B1 in online app. B for the exact categories.

We chose these categories because they were popular enough to provide us with a large list of active sellers. Moreover, they appeared reasonably different from one another, and none of them was dominated by the listings of a few sellers.

Our sample is not representative either for the populations of sellers active on eBay or for the listings because we oversample big (semi-)professional sellers with many listings. It is, however, informative about the expected buyer satisfaction with seller behavior for buyers who pick a listing in one of those five categories in May 2007.

From June 1, 2007, onward we downloaded these users' feedback profile pages on 18 occasions, always on the first day of the month. The last data collection took place on July 1, 2009. The information dating back from May 2007 to July 2006 was inferred from the data drawn in June 2007 and later.[7]

Toward capturing changes in sellers' exit rates, we define the date of exit as the date after which a user did not receive any new classic feedbacks during our observation window.[8] This is a proxy, as it may also apply to users not receiving classic feedbacks but completing transactions, or not completing any transaction for a period of time beyond our observation window, but being active thereafter. In Appendix A we show that this is a valid concern for the last 3 months of our observation period but that our main results are not affected by this because they are related to changes in the exit rate around the time of the 2008 change to the classic feedback mechanism that took place sufficiently long before that. In Section IV.B, we therefore report results using data until April 2009.

Sample attrition is low in our sample. Out of the 15,000 user names we drew in May 2007, we were able to download feedback profiles for 14,937

[7] See fig. B1 in online app. B for a graphical representation of the times at which we collected data. We were unable to collect data in November and December 2007; January, February, September, and December 2008; and January and May 2009. As we explain in Sec. IV and online app. B, DSR scores in other months are informative about the ratings received in a month with missing data, because DSR scores are moving averages, and we are interested in the effect of the change on the flow of ratings. Notice that we follow sellers over time and that therefore our data are not informative about seller entry.

[8] This criterion captures the activity of users when active as a buyer or a seller, as classic ratings can be received when acting in either role. We based our definition on classic ratings because they are more informative about the exact time after which no more ratings were received, as described in online app. B. If, before and after the change to the classic feedback mechanism, users are equally likely to stop being active as a buyer, then finding an increase in the probability of becoming inactive according to this criterion would indicate that adverse selection was affected by the feedback change. To remedy this and to make the sample of potential exiters comparable to the one of users for whom DSRs are available, we report below results for the subsample of users for whom a DSR is available at some point in our data. This means that they must have been active enough in their role as a seller to receive at least 10 DSRs in a 12-month period.

unique users in our first data collection effort on June 1, 2007.[9] One year later, we could still download data for 14,683 users and 2 years later for 14,360 users.[10]

Table 1 contains summary statistics. In the first wave of data collected on June 1, 2007, the average user in our sample was active on eBay for almost 4 years. Proxying user experience by the length of time a user has been registered, the most experienced user in our sample had registered with eBay more than 11 years before we collected our first data and the least experienced user just a few days before our observation window opened. About 2,000 of our users had registered their accounts before the turn of the millennium and about 3,000 users only within 2 years before the May 2008 changes.

At that time, the feedback score was given by the number of distinct users who have left more positive classic ratings than negative ones, minus the number of users who have left more negative ratings than positive ones. On June 1, 2007, the mean feedback score of our users was 564, the median score was 88, and 769 users had a feedback score of 0. The average share of positive feedback that users had received over the last 12 months was 99.09 percent, which corresponds well to findings in other studies. The median number of feedbacks received during the year before that was 43.

In the following year, users received roughly as many classic ratings as in the year before, and also the percentage of positive ratings was very similar. On June 1, 2008, statistics for the DSRs are available for the 4,429 users who by then had received more than 10 DSRs. Otherwise, anonymity of the reporting agent would not be guaranteed, as a seller could infer the rating from the change in the DSR. The DSR score we report on here and used in our analysis is the average reported score across the four rating dimensions. DSR scores are available for about 15 percent of the users 1 month after their introduction in May 2007 and for about 30 percent of users 1 year later. Yet another year later, the picture again looks similar, except for the number of classic ratings received, which has decreased.

At this point, it is useful to recall the objective of our analysis: it is to study sellers' reactions to the May 2008 system change, on the basis of

[9] There were download errors for 11 users, and we decided to drop three users from our panel for which eBay apparently reported wrong statistics. Moreover, there were 48 users in our sample who had listings in two categories (and therefore were not unique) and two users who had listings in three of our five categories. We dropped the duplicate observations.

[10] We waged substantive effort to following users when they changed their user names. This is important because, otherwise, we would not be able to follow those users anymore and would also wrongly classify them as having exited.

TABLE 1
SUMMARY STATISTICS

	Observations	Mean	Standard Deviation	Percentile 5th	25th	50th	75th	95th
				June 1, 2007				
Duration membership in years	14,937	3.83	2.76	.09	1.33	3.54	6.12	8.46
Feedback score	14,937	563.66	2,704.53	.00	18.00	88.00	339.00	2,099.00
Percentage positive classic ratings	14,189[a]	99.09	5.67	97.10	99.70	100.00	100.00	100.00
Member is power seller	14,937	.07	. . .	. . .	. . .	. . .	. . .	. . .
Number classic ratings previous 12 months	14,937	273.10	1,351.22	.00	10.00	43.00	161.00	975.00
Percentage positive classic ratings previous 12 months	13,943[b]	98.95	6.51	96.49	100.00	100.00	100.00	100.00
				June 1, 2008				
Number classic ratings previous 12 months	14,683	282.16	1,247.49	.00	10.00	45.00	164.00	1,042.00
Percentage positive classic ratings previous 12 months	13,811[c]	97.95	9.95	93.10	99.54	100.00	100.00	100.00
Number DSR previous 12 months	4,429[d]	378.78	1,240.91	12.00	28.00	78.25	265.50	1,378.25
DSR score	4,429[d]	4.71	.19	4.35	4.65	4.75	4.83	4.90
Number DSR relative to number classic feedbacks	4,429[d]	.42	.19	.10	.27	.44	.59	.70

	June 1, 2009							
Number classic ratings previous 12 months	14,360	200.47	1,039.00	.00	2.00	20.00	97.00	761.50
Percentage positive classic ratings previous 12 months	11,524[e]	99.48	4.19	98.18	100.00	100.00	100.00	100.00
Number DSR previous 12 months	3,272[f]	376.41	1,249.90	12.00	26.38	72.00	255.75	1,378.00
DSR score	3,272[f]	4.78	.16	4.53	4.73	4.82	4.88	4.95
Number DSR relative to number classic feedbacks	3,272[f]	.46	.20	.11	.29	.48	.63	.74

NOTE.—The table shows summary statistics for our sample of sellers, for three points in time. These are the day at which we first collected data, as well as 1 and 2 years after that. DSRs were introduced in May 2007, so the first point in time is the beginning of the first month after this. The change in the classic feedback mechanism whose effect we analyze occurred in May 2008, i.e., in the month prior to the second point in time for which we report summary statistics. The third point in time is 1 year after that. The feedback score (at that time) is the number of users who have mostly left positive feedback in the classic system, minus the number of users who have mostly left negative feedback. The power seller status is awarded by eBay if a seller has a particularly high transaction volume and generally a good track record. The percentage positive ratings is calculated as the number of positive classic feedbacks divided by the total number of feedbacks received, including the neutral ones. The DSR score is the average DSR score, per user, across the four rating dimensions.

[a] Calculated for those 14,189 users whose feedback score is positive.

[b] Calculated for those 13,943 users who received classic feedbacks in the previous 12 months.

[c] Calculated for those 13,811 users who received classic feedbacks in the previous 12 months.

[d] Calculated for those 4,429 users who received enough DSRs so that the score was displayed.

[e] Calculated for those 11,524 users who received classic feedbacks in the previous 12 months.

[f] Calculated for those 3,272 users who received enough DSRs so that the score was displayed.

unbiased ratings by their buyers effective with the introduction of DSR 1 year before. Users may sometimes act as sellers and sometimes as buyers. With our sampling rule, we ensure, however, that they were sellers in one of the five specified categories in May 2007. Moreover, DSRs can be received by users only when acting as sellers. Hence, the average DSR score reflects only how a user behaved in that very role.[11] When interpreting our results, it is important to keep in mind that we will not be able to observe the reaction of sellers who receive fewer than 10 DSRs per year. At the same time, we do capture behavior that is associated with most of the transactions on eBay, as those sellers who receive fewer than 10 DSRs per year are not involved in many sales on eBay.

IV. Results

A. *Staying Sellers' Reactions*

After the introduction of DSRs in May 2007, the May 2008 change to the classic feedback system lowered the cost to buyers to nonanonymously voice negative experiences by means of negative classic feedbacks. We document in Section VI.A below that this led at first to more negative classic feedbacks, which reflects reactions to transactions that had taken place prior to the change. At the same time, the DSR system remained unchanged. This allows us to attribute changes in DSR ratings over time to the May 2008 change to the classic feedback mechanism that led to higher market transparency. We expect such changes because the increased inclination to report negative experiences should increase the sellers' incentive to change their behavior toward satisfying buyers.

Figure 2 shows how the average DSRs evolved over time.[12] The dots are averages for the selected sample of sellers who conducted enough transactions so that a DSR score was available (recall that at least 10 DSR ratings have to be received for this). There are fewer sellers for whom this is the case in those first 2 months, and those even more selected sellers

[11] One may still wonder how often the users in our sample acted as buyers. On June 20, 2008, eBay revealed in a statement that buyers leave DSRs 76 percent of the time when leaving "classic" feedback. In our data collection just before this statement, the mean overall "DSR to classic" ratio of users for whom a DSR is displayed is about 43 percent. The difference between those 76 percent, where users acted as sellers, and the 43 percent, where they acted as buyers or sellers, comes about because they may also have acted as buyers. Looked at in a different way, the 43 percent in our sample is a lower bound on the probability that a user has acted as a seller in a given transaction, because DSRs can be left only when a classic rating is left at the same time.

[12] Recall that at any point in time, DSR indices are published in four categories, for every seller that has received more than 10 DSRs up to that point, with ratings aggregated over the respective preceding 12 months. Appendix fig. A1 shows that the patterns by category resemble one another closely. Therefore, from now on we will use the average DSR across rating categories.

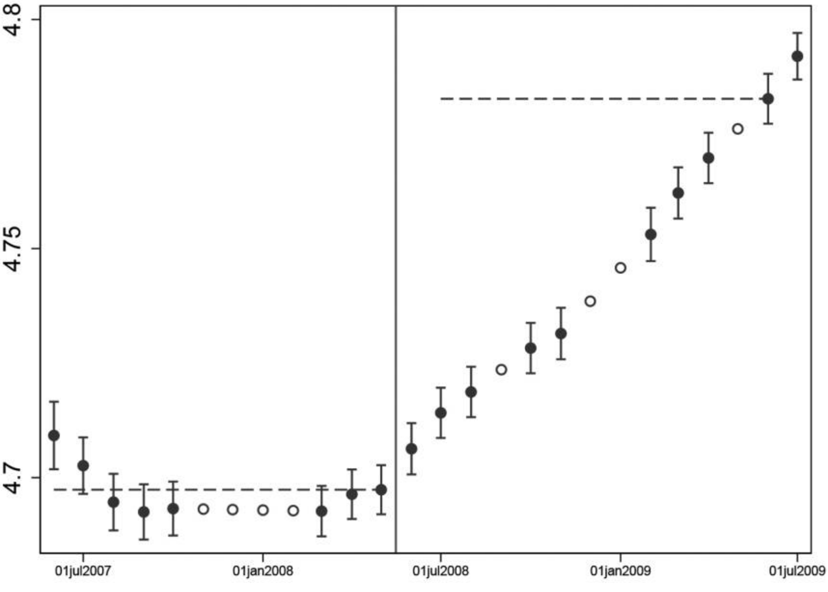

FIG. 2.—Evolution of detailed seller ratings. The figure shows how DSRs changed over time. The vertical line denotes the May 2008 change to the classic feedback mechanism. Dots are averages across users for whom DSRs are displayed; error bars depict corresponding 95 percent confidence intervals. Circles are linearly interpolated values for the periods in which we did not collect data. We substantially improve on the linear interpolation in our formal analysis. See footnote 7, online appendix B, and the discussion in the main text. Before averaging DSRs across users, we calculated the average DSR per user across the four categories. Horizontal dashed lines visualize that the dots are averages over the 12 months prior to the point in time at which the DSRs are displayed. Color version available as an online enhancement.

receive higher DSRs on average. For that reason, these first two dots in the figure cannot be compared to the remaining ones. In our regression analysis below, we take selectivity into account by controlling for fixed effects.

When interpreting figure 2 it is important to keep in mind that DSR scores show the average of all DSR ratings given in the previous 12 months. Therefore, if, on average, all ratings received after the change were higher by the same amount in all months after the change and there was no time trend before and after the change, respectively, and the same number of ratings was received in each month, then one would observe a flat curve before the change, a linear increase in the 12 months after the change, and thereafter again a flat curve (at a higher level). The full effect of the change equals the difference between the DSR score 1 year after the change and the DSR score right before the change. It is depicted in the horizontal lines in figure 2.[13] The figure clearly shows that the DSRs have increased after the May 2008 change.[14]

We performed regressions to quantify the effect shown in figure 2, controlling for fixed effects. Denote by DSR_{it} the average score across the four DSR rating dimensions reported for seller i in period t. Recall that our data are always drawn on the first day of the month and that DSR_{it} is the average of all ratings seller i has received over the previous 12 months. Let $w^t_{i\tau}$ be the weight put in the construction of the index on $\mathrm{dsr}_{i\tau}$, the average of all ratings given in month τ. This weight is zero for $\tau < t - 12$ and $\tau \geq t$. Otherwise, it is given by the observed number of ratings received in τ, divided by the total number of ratings received between periods $t - 12$ and $t - 1$. Hence $\Sigma^{t-1}_{\tau=t-12} w^t_{i\tau} = 1$ and

$$\mathrm{DSR}_{it} = \sum_{\tau=t-12}^{t-1} w^t_{i\tau} \cdot \mathrm{dsr}_{i\tau}. \tag{1}$$

We wish to estimate how $\mathrm{dsr}_{i\tau}$ changed after May 2008. That is, we are interested in estimating the parameter β in

$$\mathrm{dsr}_{i\tau} = \alpha + \beta \cdot \mathrm{POST}_{i\tau} + \alpha_i + \varepsilon_{i\tau},$$

where $\mathrm{POST}_{i\tau}$ takes on the value one after the change and zero otherwise. The change occurred between May 1 and June 1, 2008, and therefore we code $\mathrm{POST}_{i\tau} = 1$ if τ is equal to July 2008 or later, and $\mathrm{POST}_{i\tau} = 0.5$ if τ is equal to June 2008. With this we assume that half of the ratings received in May 2008 correspond to transactions taking place after the change.[15] The term α_i is an individual fixed effect with mean zero, and $\varepsilon_{i\tau}$ is an individual- and time-specific error term. We cannot estimate β directly by regressing $\mathrm{dsr}_{i\tau}$ on $\mathrm{POST}_{i\tau}$ because $\mathrm{dsr}_{i\tau}$ is not observed. However, by (1), the reported DSR score is the weighted average rating received in the preceding 12 months, so that

[13] The change occurred in mid-May 2008. Hence, the DSR score at the beginning of June 2009 contains no DSRs left before the change because it is calculated from the ratings received in the preceding 12 months. Conversely, the DSR score at the beginning of May 2008 contains no ratings received after the change. Figure B1 in online app. B shows at which points in time data were collected and depicts over which periods, respectively, the DSR scores were calculated.

[14] Unfortunately, we were not able to collect data for more than 1 year after the change, because eBay started to ask users to manually enter words that were hidden in pictures when more than a small number of pages were downloaded from their server. Otherwise, we would be able to assess whether the curve indeed flattens out 1 year after the change. The remarkable fact, however, is that the scores start increasing rapidly and immediately after the change.

[15] This is conservative. If anything, it biases our results downward because we partly attribute a positive effect to the time prior to the change. Then, we would (slightly) underestimate the effect of the change. See also the discussion in Sec. VI.B on competing explanations and the robustness check in online app. D.

$$\mathrm{DSR}_{it} = \alpha + \beta \cdot \left(\sum_{\tau=t-12}^{t-1} w_{i\tau}^{t} \cdot \mathrm{POST}_{i\tau} \right) + \alpha_i + \left(\sum_{\tau=t-12}^{t-1} w_{i\tau}^{t} \cdot \varepsilon_{i\tau} \right), \qquad (2)$$

and $\Sigma_{\tau=t-12}^{t-1} w_{i\tau}^{t} \cdot \mathrm{POST}_{i\tau}$ is the fraction of DSRs received after the 2008 change of the system. Hence, we can estimate α and β by performing a fixed-effects regression of the reported DSR score on a constant term and that fraction.[16] We can control for time trends in a similar way.[17]

It is important to control here for fixed effects because at any point in time the DSR score is observable for only a selected sample of sellers, namely, those involved in enough transactions so that the DSR score was displayed. Otherwise, the results may be biased; for example, the DSR score of poorly rated sellers with lower α_i's may be less likely to be observed before the change because by then they would not have received enough ratings. At the same time, we also control for seller exit when studying effects on staying sellers' behavior. In both cases, controlling for fixed effects is akin to following sellers over time and seeing how their DSR score changed, knowing the fraction of the ratings that were received after the feedback change. This is generally important because we are interested in the change in the flow of DSRs that is due to the May 2008 change of the feedback mechanism.

Table 2 contains the regression results using DSR scores averaged over the four detailed scores of all sellers. In specification 1, we use the whole sample and find an effect of 0.0581. In specification 2, we restrict the data set to the time from March 1 to October 1, 2008; hence there are only 30,488 observations. We do so to estimate the effect locally, because this allows us to see how much of the overall effect is due to an immediate response by sellers. The estimated effect is equal to 0.0414, which suggests that most of the effect occurs from mid-May to October 1, 2008. In specification 3, we instead allow for a piecewise linear time trend over the en-

[16] One might object that in (2) the weights enter both the regressor and the error term and therefore the estimates will be biased. This, however, is not a problem as long as $\mathrm{POST}_{i\tau}$ is uncorrelated with $\varepsilon_{i\tau'}$ conditional on the weights and for all τ, τ', which is plausible because the change to the system was exogenous. To see this, suppose that there are two observations for each individual, consisting of the DSR score and the fraction of DSR received after the change, respectively. Then one can regress the change in the DSR score on the change of that fraction, constraining the intercept to be zero. This will estimate the change in the mean of received DSR before vs. after the change, which is our object of interest. Alternatively, one can show that under the above-mentioned condition, the covariance between the regressor and the error term is zero.

[17] For two separate time trends, the regressors are weighted average times before and after the change. When we subtract the time of the change from those, respectively, then the coefficient on the indicator for the time after the change is still the immediate effect of the change. The change in the trend can be seen as part of the effect. We will also make a distinction between a short-run effect and a long-run effect when we report the results. For this, the regressors will be the fraction of ratings received until the end of September 2008 and thereafter.

TABLE 2
Effect of the May 2008 Change on DSR Ratings

	Full Sample (1)	Small Window (2)	Time Trend (3)	DSR <4.75 (4)	DSR ≥4.75 (5)
Average DSR before change	4.7061***	4.7030***	4.7149***	4.5912***	4.8138***
	(.0007)	(.0005)	(.0034)	(.0011)	(.0006)
Effect of feedback change	.0581***	.0414***		.0904***	.0316***
	(.0024)	(.0047)		(.0044)	(.0021)
Effect of feedback change until September 2008			.0168**		
			(.0083)		
Effect of feedback change after September 2008			.0589***		
			(.0184)		
Linear time trend before change			.0009**		
			(.0004)		
Linear time trend after change			.0007		
			(.0019)		
Fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	.0580	.0131	.0605	.0809	.0466
Number sellers	5,224	4,919	5,224	2,337	2,337
Number observations	67,373	30,488	67,373	31,260	33,508

Note.—The table shows results of regressions of the average DSR score, averaged over the four categories, on a constant term and the fraction of feedbacks received after May 2008. For May 2008, we assume that half of the feedbacks were received before the change and the other half after the change. In specification 2, we exclude observations before March and after October 2008. In specification 3 we distinguish between the effect until the end of September 2008 and after that date and also account for a piecewise linear time trend. See the main text for details. Specification 4 includes only those sellers who had a DSR score below the median of 4.75 in May 2008 and specification 5 only those above the median. One observation is a seller-wave combination. Throughout, we control for fixed effects. The R^2 is the within R^2.

* Standard errors are cluster-robust at the seller level and significant at the 10 percent level.

** Standard errors are cluster-robust at the seller level and significant at the 5 percent level.

*** Standard errors are cluster-robust at the seller level and significant at the 1 percent level.

tire observation window. We find that the time trend before the change is very small and not significantly different from zero after the change. In light of figure 2 this is not surprising, as it already shows that there was no time trend in the reported DSR scores before May 2008. The effect of the change is estimated to be a short-run effect of 0.0168, until the end of September 2008, and a bigger effect of 0.0589 after that.[18]

[18] Without the piecewise linear time trend the short-run effect is estimated to be equal to 0.0325 and the long-run effect is estimated to be 0.0711, with standard errors 0.0057 and 0.0028, respectively. The magnitude of this estimate of the short-run effect is comparable to the one of the effect using the smaller sample that is reported in col. 2. We obtain similar estimates when we define the short run to last longer or shorter than 3 months.

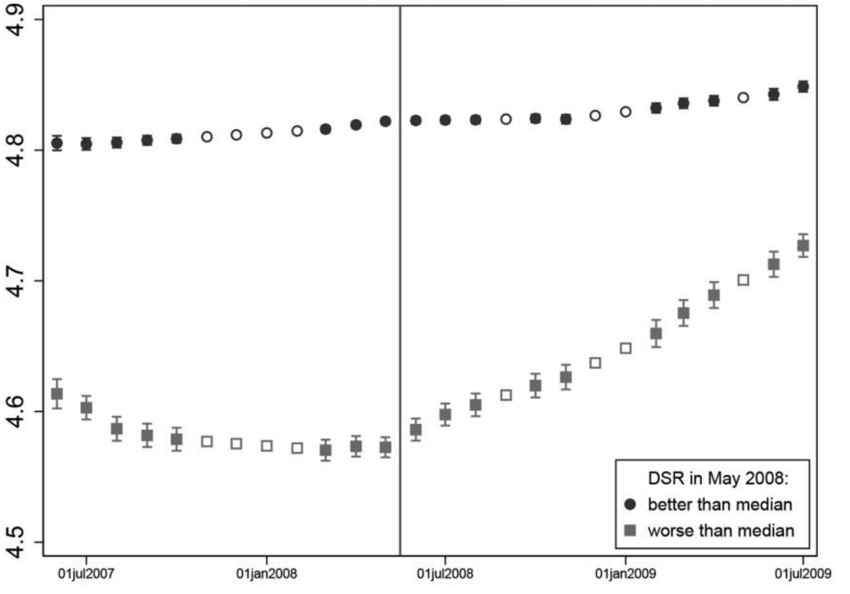

FIG. 3.—Evolution for two different groups. The figure shows how average DSR score changed over time, with sellers split into those with a DSR score above the median of 4.75 prior to the May 2008 change and those with a score below that. See also the notes to figure 2. Color version available as an online enhancement.

To assess the magnitude of the effect, it is useful to express the numbers in terms of quantiles of the distribution of DSR scores among sellers prior to the May 2008 change. According to the results in the first column, the average DSR before the change is 4.7061, and after the change, it is $4.7061 + 0.0581 = 4.7642$. This corresponds to roughly the 40 and 60 percent quantiles of the distribution of ratings prior to the change, respectively. Hence, the May 2008 change has led to a significant and sizable increase in the buyers' evaluations. This is also remarkable because buyers have been able to express their satisfaction with seller performance without the threat of retaliation already by means of the DSR ratings introduced in May 2007.

We also looked at how this increase differed between sellers with low and high DSR scores before the change. Toward that, we split our sample at the median DSR of 4.75 between high- and low-ranked sellers just before the May 2008 change. Figure 3 gives the picture. The increase in DSR score is stronger for sellers with below-median scores ex ante. Columns 4 and 5 of table 2 report the corresponding estimates, again controlling for seller fixed effects. The difference between the effect for above- and below-median sellers is significantly different from zero.[19] We obtain similar results when we perform regressions for those two different groups only for a smaller time window, as in specification 2, or control for time trends, as in specification 3. In the second part of table C2 in online appendix C, discussed later in the context of the robustness checks, we show the effects of the feedback change by decile of sellers' DSR rating. We find a decline in the magnitude and significance of the effect, with increasing decile.

In all, the empirical evidence provides support for our hypothesis that abandoning negative buyer rating by sellers—and thereby reducing impediments against negative seller rating by buyers in the classic rating system—has led to significant and substantive increases in the buyers' evaluations as measured by the independently measured DSRs.

Finally, toward exploring what is behind this increase, we used our Klein et al. (2006) data on individual classic feedbacks and sampled 470 negative classic feedbacks given by buyers. They all correspond to transactions that have taken place before May 2008. We used cases in which buyers left feedback first and coded the text comments. The results are reported in table 3. Arguably, most of these buyer complaints are related to forms of seller behavior that can be changed at relatively low cost.

B. *Seller Exit*

The results shown above suggest that the May 2008 change to the classic feedback mechanism led to a significant and substantial increase in buyer satisfaction. At the same time, the change could also have led to a selection process, with poorly performing sellers leaving the market. Figure 4 shows how the fraction of sellers who have become inactive, and the corresponding hazard rate into inactivity, changed over time.[20]

Many sellers leave over time, both before and after the change. By June 1, 2009, about 25 percent of the sellers active at the opening of our observation window had become inactive. The figure shows, however, that the May 2008 change did not lead to an increase in the exit rate of

[19] One concern may be that the increase for the sellers with low DSR before the change was driven by mean reversion. Indeed, we have divided sellers on the basis of their score. To check whether mean reversion has to be accounted for, we instead divided sellers according to the median score on August 1, 2007. With this, scores for the bad sellers also increase only after the change. This shows that mean reversion is not of concern here.

[20] In order to provide results that complement those for the evolution of DSRs, we restrict the sample to those users for whom a DSR rating is available at some point in time. In figs. 2 and 3, information on a particular seller at a given point in time is used if the DSR score is available at that particular point in time. This means that the composition of sellers over whom we average changes over time. To obtain the regression results in table 2, we therefore control for seller fixed effects. In these regressions, we use, as we do in the analysis of seller exit in this section, information on sellers for whom a DSR is available at some point in time. In that sense the results are comparable.

TABLE 3
Stated Reasons for Dissatisfaction When Buyer Leaves First Negative Feedback

Stated Reason	Percentage
Communication problem	43.0
No item received	38.3
Item arrived in poor condition	18.1
Slow shipping	11.5
Unfair shipping charges	4.7

Note.—Based on 470 text comments belonging to classic feedbacks left by buyers. See Klein et al. (2006) for data description.

sellers. To test this formally, we conducted ordinary least squares regressions of indicators for exiting sellers on an indicator for the time period after May 2008, controlling for a piecewise linear time trend and using only the observations in which sellers are at risk of exiting, that is, have not exited yet. The results are shown in columns 1–4 of table 4. There is no statistically significant increase in the exit rate after May 2008. Moreover, the time trend in the hazard rate after the change is not statistically

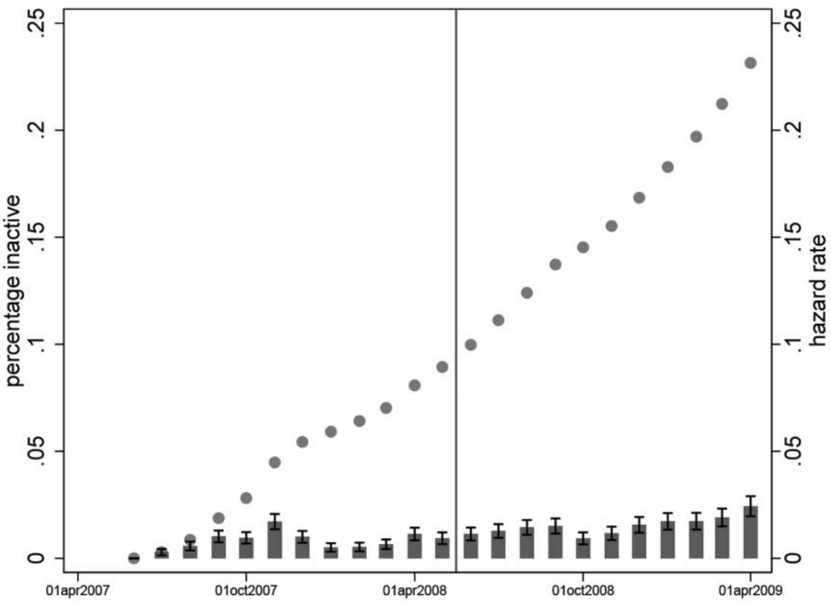

Fig. 4.—Exit from the market. The figure shows the evolution of the percentage of inactive sellers since June 1, 2007 (dots), and the corresponding hazard rates (bars, with 95 percent confidence intervals). Inactivity, or exit, is defined as not receiving any classic ratings anymore in the sample. Reported for the subsample of users for whom a DSR rating is available at least once until July 1, 2009. Color version available as an online enhancement.

TABLE 4

EFFECT OF THE MAY 2008 CHANGE ON SELLER EXIT

	EXIT FROM THE MARKET				NUMBER DSRS RECEIVED PER MONTH			
	Full Sample (1)	Small Window (2)	DSR <4.75 (3)	DSR ≥4.75 (4)	Full Sample (5)	Small Window (6)	DSR <4.75 (7)	DSR ≥4.75 (8)
Before change	.0119***	.0104***	.0181***	.0065***	32.7129***	31.7242***	38.5660***	27.6063***
	(.0014)	(.0011)	(.0026)	(.0017)	(.0883)	(.0502)	(.1688)	(.0732)
Effect of feedback change	.0004	.0017	.0001	.0007	−.0651	−.0056	−.1631	.0334
	(.0018)	(.0015)	(.0032)	(.0020)	(.1293)	(.2115)	(.2582)	(.0914)
Linear time trend before change	.0015***		.0021***	.0009**	.3425***		−.0879	.7274***
	(.0004)		(.0007)	(.0005)	(.0931)		(.1699)	(.0890)
Linear time trend after change	.0002		−.0003	.0006*	−.2289***		−.5995***	.0685
	(.0003)		(.0005)	(.0003)	(.0795)		(.1521)	(.0734)
R^2	.0009	.0001	.0008	.0013	.0009	.0000	.0049	.0103
Observations	56,467	19,119	26,157	30,310	32,243	16,761	15,021	17,222

NOTE.—Columns 1–4 show the results of regressions of an indicator for exiting on a constant term, an indicator for after May 2008, as well as a piecewise linear time trend in specifications 1, 3, and 4. Columns 5–8 show the results of regressions of the number of DSRs received in the past 12 months, divided by the maximum of 12 and the number of months since their introduction in May 2007, on a constant term and an indicator for after May 2008. We control for a piecewise linear time trend in specifications 5, 7, and 8. In all specifications, one observation is a seller-wave combination provided that the seller has not left before. In specifications 2 and 6, we exclude observations before April and after July 2008. For all other specifications, we use data from January to December 2008. Specifications 4 and 7 include only those sellers who had a DSR score below the median of 4.75 in May 2008, and specifications 5 and 8 include only those above the median. We used an extrapolated value if the DSR score was available only at a later point in time. Robust standard errors are in parentheses.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

different from zero at the 5 percent level and is estimated to be lower than before. This suggests that the feedback change neither triggered immediate nor induced delayed exit.[21]

As before, we also split the sample into sellers with above- and below-median DSR scores prior to the May 2008 change.[22] Not surprisingly, 33 percent of the below-median sellers have left the market by May 1, 2009, as compared to 18 percent of the above-median ones. Figure 5 shows the corresponding hazard rates. Yet the results presented in columns 3 and 4 of table 4 confirm that the exit rate did not change significantly in May 2008.

We also assessed whether the change to the feedback system resulted in a lighter form of exit, namely, a reduction in the (poor) sellers' activity. The results are reported in columns 5–8 of table 4. Measuring the level of activity by the number of DSRs received and using specifications corresponding to the ones for our analysis of exit, we also find no effects of the change to the feedback system.[23]

Observe that in figures 4 and 5, we show exit rates and the fraction of inactive sellers only until April 2009 and omit the last 3 months of data. The reason for this is that because of the way we defined exit, simple estimates of exit rates for the last 3 months are plagued by a form of truncation bias. We define exit as the first point in time from which we do not observe a user receiving any classic ratings anymore, and therefore, we are more likely to misclassify infrequent sellers as inactive toward the end of our observation period. To see why this generates truncation bias,

[21] Another way to test for increased exit after the feedback change is to use the McCrary (2008) test for a discontinuity of the density of the time of exit among those whom we do classify as exiting at one point or another. We estimate the decrease to be 2.3 percent (of the density), with a standard error of 21.1 percent, which means—in line with the results presented above—that the density has no discontinuity at the time of the feedback change. See also App. fig. A4. In this figure, the decrease by 2.3 percent is given by the percentage difference between the nonlinearly extrapolated (to the vertical line) curve to the right and the one to the left.

[22] In contrast to our analysis of the evolution of DSR scores, we use here a linearly extrapolated value if the DSR score is available only at a later point in time. The reason for this is that, otherwise, we would obtain biased results. To see why, suppose that a user would not have a DSR score on May 1, 2008, but would have one at all future times. Then, we would have included him in the sample for fig. 4, for the reasons given in fn. 20. Not including him here as a below-median seller would lead to biased results in the sense that we would systematically exclude sellers for whom the DSR score becomes available only later, which can happen only if they exit after that point in time. This would then lead to an upward bias in the hazard rates after May 2008.

[23] As explained before, we have based our analysis of exit on classic feedbacks because the date of exit is defined as the earliest time after which a user has not been observed to receive any more feedbacks, and missing months are a problem for that. For the level of activity, this is not a concern, and therefore, we use the number of DSRs received as a measure of activity. Our results thereby correspond more closely to the ones on seller behavior, because we measure seller behavior by the average DSR ratings. The numbers of observations differ from the ones reported in table 2 because we restrict the sample to cover only the year 2008 and because we condition on not having exited yet, as we do in cols. 1–4.

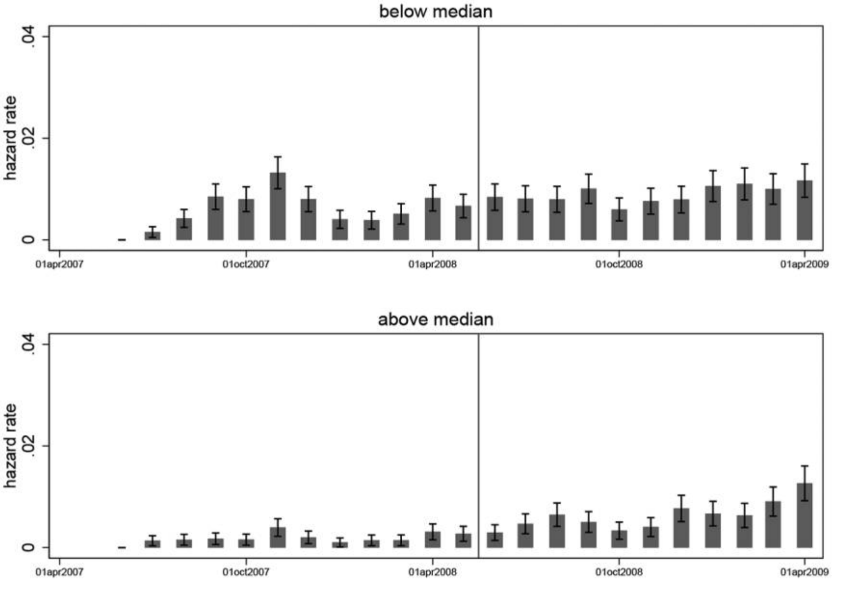

FIG. 5.—Exit for two different groups. See the notes to figure 4. Sellers are split into those who had a DSR score above the median of 4.75 prior to the May 2008 change and those who had a score below that. We used an extrapolated value for May 2008 if the DSR score was available only at a later point in time. Color version available as an online enhancement.

suppose that a user is active and receives a rating only in the second month of each quarter. Then, if we have data until July 2009, we will observe the user to be active in February and May and will incorrectly infer that he exited in June because he would be observed to be active again only in August. This example also shows that the likelihood that we misclassify individuals as inactive is related to how active a user is. In order to assess whether the presence of truncation bias likely affects our finding that the change to the feedback system did not trigger exit, we performed two robustness checks. Appendix figures A2 and A3 show the respective results. In the first robustness check, we redefine exit on the basis of data for all but the last months in our data set. In the second robustness check we simulate data and show that defining exit on the basis of receiving ratings can indeed produce a pattern similar to the one in the last 3 months of our data that are also shown in figure A2. Together, these robustness checks show that truncation bias affects only our estimates of the exit rates in the last 3 months of the observation period. It does not affect our conclusions, because they are related to changes in the exit rate long before that.

V. A Simple Explanatory Paradigm

In this section, we develop our preferred interpretation of the empirical findings. Our informal and highly stylized model also allows us to specify our notion of moral hazard and adverse selection for the present context.

We focus on one stage in an infinitely repeated game between one seller S and many buyers. Figure 6 provides an overview of the sequence of decisions the seller S and a randomly selected buyer B make in a given transaction, after nature has revealed the quality of the good to the seller. Buyers rate sellers first, because otherwise the buyer would be free to rate truthfully anyway and the change to the classic feedback system would not have an effect. One can also think of this sequence as arising in equilibrium.[24]

Sellers are either conscientious, indicated by C, or exploitative, indicated by E. Moreover, they differ by their cost of effort for delivering the good. When engaging in high effort, a seller of type j faces effort cost c^j, $j \in \{C, E\}$, with $0 < c^C < c^E$. The cost of providing low effort is normalized to zero for both types of sellers. The typical seller is endowed with publicly known reputation capital built from previous trading partners' reactions to his behavior, denoted by k^j, $j \in \{C, E\}$.

The good is either of high quality q_h or low quality q_ℓ. When offering it, a type j seller decides whether to announce it at its true quality q_i and ask for an appropriate price $p_i(k^j)$, $i \in \{\ell, h\}$, $j \in \{C, E\}$, which he always does if the good is of high quality, so $i = h$; or to shirk if $i = \ell$, by announcing the low-quality good as being of high quality, q_h, and ask for a high price $p_h(k^j)$. The buyer, not knowing the true quality of the good, observes the quality-price tuple as announced by the seller, denoted by $[\hat{q}_i, \hat{p}_i]$, $i \in \{\ell, h\}$, as well as the seller's reputation capital k^j. She is not able to infer the seller type from k^j.[25] She forms an expected utility $\mathbb{E}[u(\hat{q}_i, \hat{p}_i, k^j)]$, $j \in \{C, E\}$, related to buying the good. We assume this utility to increase in the quality as announced by the seller, decrease in the announced price, and increase in the seller's reputation, which is why sellers have reputational concerns. Buyer B decides to buy the item if $\mathbb{E}[u(\hat{q}_i, \hat{p}_i, k^j)] \geq \tilde{u}$, where $\tilde{u}$ is the utility associated with her exogenously specified outside option.

Once the good is sold, S decides whether to spend effort on its delivery. Buyer B receives the good, observes the accuracy of the item description and the shipping quality, and rates S. This results in an upward or downward revision of k^j. Decisions are made rationally, that is, with backward induction in that simple infinitely repeated stage game.

[24] In Klein et al. (2006), we found that the seller rated his counterpart in only 37 percent of all cases in which both left a rating, and that in this sequence, a positive rating by the seller was followed by a negative buyer rating in less than 1 percent of the cases, indicating that the holdup situation we consider to be at the root of the phenomenon analyzed here is not prevalent when the seller rates first.

[25] One can think of this as being related to a lack of buyer sophistication.

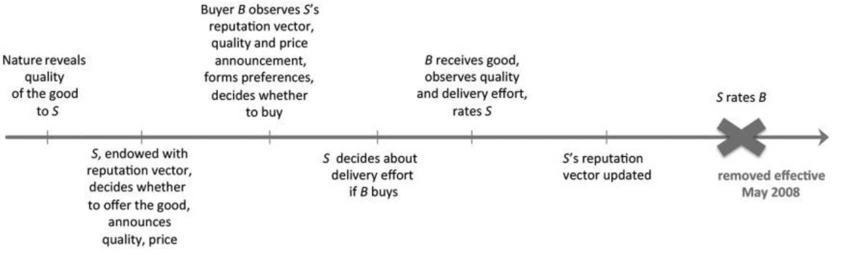

FIG. 6.—Sequence of decisions in a typical eBay transaction. Color version available as an online enhancement.

Before May 2008, the sequence of decisions involving such a transaction was typically concluded by the additional step indicated in figure 6, in which the seller rated the buyer. This is important, because buyers who rate may or may not have reputational concerns. These could arise because they may intend to use their good ratings received as buyers when selling a good; also, with a bad rating, a future seller may exclude her from a trade.[26] Because of these reputational concerns, a seller could opportunistically condition his rating on the buyer's rating (and had an incentive to develop a reputation for that) and leave a negative mark in reaction to a negative mark given by the buyer (he could claim it was *B*'s fault that something went wrong). When nature selected q_ℓ, it therefore became less costly to him to announce the low-quality good at high price p_h and to not take any effort to deliver the good, as compared to a situation in which buyers would always rate truthfully: He could nevertheless receive a positive contribution to his reputation capital, as buyers with reputational concerns would still rate him positively. Only a buyer with no reputational concerns would rate the seller negatively, and this rating would be retaliated by the seller. In any case, negative experiences were underreported, an impediment to market transparency.

The fact that, before the May 2008 change, exploitative sellers would find it optimal to provide low effort and overstate the quality of the goods they sell was a form of moral hazard. Adverse selection would arise when these exploitative sellers would find it profitable to enter this market and exert moral hazard.

After the May 2008 change, even buyers with reputational concerns can give a strategically unbiased negative rating without having to fear retaliation costly to them. Consequently, market transparency increases. The seller now has a higher incentive to accurately describe the item even if of low quality and quote an appropriately low price. He also has a higher in-

[26] For the latter, eBay has established clear rules; see http://pages.ebay.com/help/sell/buyer-requirements.html.

centive to exert effort toward delivering the item in expectation of a good rating, and with it the possibility of selling at higher prices in the future.

This suggests that an increase in market transparency will discipline sellers if they decide to stay in the market. Staying in the market is especially costly to the exploitative sellers, who may therefore alternatively decide to exit the market—a *reduction in adverse selection*. They will be more likely to do so the higher their costs to changing behavior, and the less costly it is for them, in terms of forgone profits, to leave the platform instead of changing behavior. Conversely, if the costs to changing behavior are low relative to the forgone profit from trading on eBay, then they will decide in favor of changing their behavior to the benefit of the buyers—a *reduction in moral hazard*.

VI. Additional Empirical Support for Our Interpretation and Discussion of Competing Explanations

In Section IV, we showed that removing negative seller ratings of buyers in eBay's classic feedback system, and with it potential retaliation to negative buyer ratings, was associated with a significant improvement in DSRs especially for sellers that previously were rated poorly, and with no change in sellers' exit behavior, especially that of the poorly rated ones. In Section V, we gave an explanation that is consistent with these results. In this section, we first present additional evidence that supports the assumptions underlying our explanation and then work through a list of competing explanations to show that these are likely not to hold. Along the way, we conduct a number of robustness checks. We discuss additional competing explanations in online appendix C and conduct an additional robustness check in online appendix D.

A. Empirical Support for the Assumptions Underlying Our Explanation

The key assumption underlying our explanation is that buyer feedback reflects the quality and effort of the seller in question. Clearly, the ideal measure of seller type and effort would be independent observations of the conscientiousness of the seller when describing the good and the effort waged when delivering the good. Yet such direct measures are not available for eBay transactions and indeed tend not to exist for brick-and-mortar stores.

Our measure of seller effort is a reported average of buyers' ratings of seller performance. That report is not provided within the classic feedback system whose change we analyze, but in the DSR system with anonymous buyer reports introduced 1 year before the classic system was changed. Not that anonymity removes all biases. In particular, different buyers may rate the same buying experience differently. Yet, as long as these biases are (mean) independent of seller performance and time, subjective buyer ratings are useful for evaluating changes in seller performance—once all buyers leave a rating.

Another source of bias could be that not all buyers rate. For our analysis, however, it matters only whether any bias before the May 2008 change remains unaffected by that change. That bias could in principle even be seller specific. Econometrically, the bias would then be part of the seller fixed effect and thereby be controlled for.[27] Indirect evidence for this is provided by the fact that the number of DSRs received remains unchanged.[28] This is an indirect measure, because only the number of ratings, rather than the number of transactions, is recorded in our data. However, at the same time, the ratio of the number of DSRs relative to the number of classic ratings stayed the same, as documented in table 1 and formally tested in Section VI.B. This suggests that the decision whether or not buyers rate was not affected, so that changes in DSR ratings for a given seller can indeed be expected to reflect changes in buyer satisfaction.

The relationship between seller behavior and buyer rating should also be reflected in the classic feedbacks. Toward their analysis, we classified all users sampled as being foremost sellers or buyers on eBay on the basis of the ratio between the number of DSRs and (cumulative) classic feedbacks received by May 1, 2008. The 25 percent of users with the highest ratio are classified as foremost sellers and the 25 percent with the lowest ratio as foremost buyers.

In figure 7 we compare the percentage of positive feedbacks obtained for the two subpopulations in the observation window. As it is based on some 23,000 observations, it shows very clearly that effective May 2008, the percentage of positive feedbacks dropped for users identified as foremost sellers but remained unchanged for those identified as foremost buyers. Our explanation is as follows: Some proportion of the sellers did not anticipate the May 2008 change as indicated in eBay's earlier announcement (see fn. 5) and thus still behaved opportunistically right before the May 2008 change. Thereafter, buyers could still leave negative classic ratings of this opportunistic seller behavior in those transactions without the risk of seller retaliation, which they did. For the users classified as foremost sellers, we therefore expected and, indeed, observe a downward jump in buyer ratings right after the May 2008 change, that

[27] Formally, a sufficient condition for this to be true is that the propensity for a buyer leaving a rating is the same before and after the change. Thinking about it through the lens of a Heckman (1978) selection model, this would imply that the inverse Mills ratio term stays constant because the index that changes the probability would remain unchanged.

[28] Table 1 shows that the number of DSRs in the 12 months before June 2008 is roughly equal to the number of DSRs received in the 12 months before June 2009. A more formal test of whether the feedback change had an effect on the number of ratings is done in table 4.

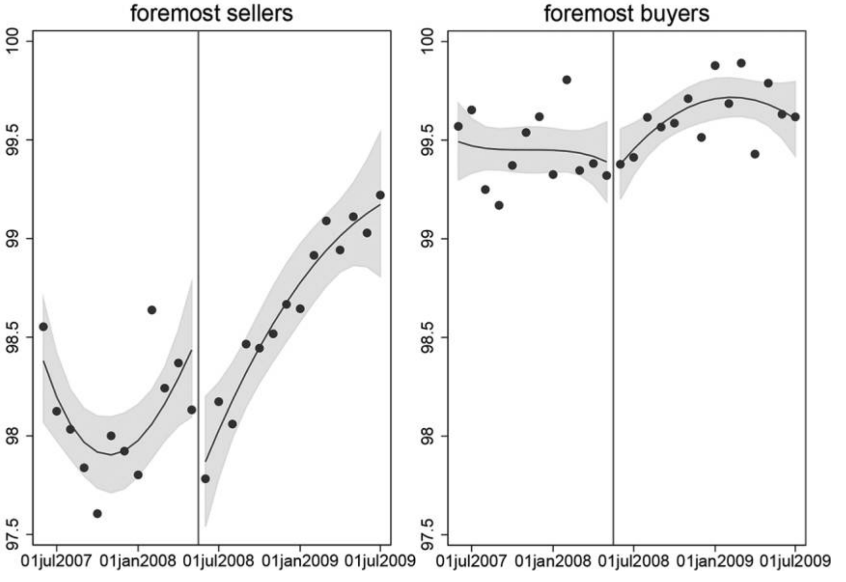

Fig. 7.—Effect on classic feedbacks. The figures show the percentages of positive feedbacks over time. The dots are averages per wave. The lines are fitted values of local quadratic regressions, and the shaded area shows pointwise asymptotic 95 percent confidence intervals, respectively. We used the Epanetchnikov kernel with a bandwidth of 200. The solid vertical line depicts the change to the classic feedback mechanism. Color version available as an online enhancement.

is, before these sellers could react to that change. This, in our view, shows that first, buyer ratings reflect seller behavior quite accurately and, second, that buyers have reputational concerns. The increasing ratings thereafter suggest that, as time goes on, sellers do react to buyer ratings by improving their behavior.[29]

Appendix table A1 contains the corresponding formal tests. Columns 1–4 contain results for foremost sellers and columns 5–8 results for foremost buyers, following our classification. Each column corresponds to a different bandwidth for the kernel regressions, and in the rows we show results for a local linear regression and a local quadratic regression. Figure 7 suggests that a bandwidth of 200 fits the data well when we use a local quadratic specification. The corresponding estimate for sellers is $-.727$. It is significant at the 1 percent level.

[29] As indicated before, the May 2008 change was announced by eBay already in January 2008. All sellers aware of this announcement should have strategically adjusted their behavior before the May 2008 change, reducing the observed jump in classical ratings. Hence the early announcement effect works to our disadvantage by reducing the effect we still observe. In that sense our estimates are lower bounds on the total effect to be expected from the change.

Next, one might wonder whether there are indeed generically different seller types, so that adverse selection can arise at all; otherwise, only moral hazard would play a role. Recall that we have included two parameters in the specification of our key regression, namely, a seller fixed effect α_i and a seller-specific time-varying effect ε_{it}. The fraction of the variance of $\alpha_i + (\Sigma_{\tau=t-12}^{t-1} w_i^{\tau} \cdot \varepsilon_{i\tau})$ at a given point in time and across sellers that is due to variation in α_i gives us an indication of the relative strength of the seller fixed effect. In the five specifications reported in table 2, this fraction ($\times$ 100) amounts to 84, 94, 84, 77, and 54 percent, respectively. Only the last fraction is low. But that reports on the above-median sellers. One sees that a substantive part of the heterogeneity across sellers is time invariant, so that differences across sellers over time must be at least as important as seller-specific differences in outcomes.[30] This is in line with our view that sellers differ by type.

Finally, Cabral and Hortaçsu (2010) argue that in an anonymous market such as the one under discussion, one should expect a correlation between ratings and exit, because rational sellers change their behavior just before leaving the market, or alternatively are more likely to leave the market after having received negative feedback, because of the lower value to staying in the market.[31] Evidence along those lines supports our claim that buyers correctly value the transaction via the DSR. In Appendix figure A5, we compare the continuing and the exiting sellers' DSR scores, relative to those obtained 3 months earlier. Whereas the DSR scores of the continuing sellers remain essentially unchanged in the time window considered, the exiting sellers' DSR scores are lower on average.

B. Competing Explanations for the Increases in DSRs

With our interpretation that the increase in DSRs after the May 2008 change is caused by reduced seller moral hazard, we abstract from possible other causes, such as other contemporaneous changes in eBay's rules or changes in the macroenvironment. Ideally, one would assess these alternative explanations using a "control group" from a market in which

[30] To be precise, this is the ratio between the variance of the fixed effect and the total variance in the reported DSR ratings. However, these are moving averages. To explore what the ratio would be if we could use the monthly DSR rating flows, we conducted a Monte Carlo study. Assuming that we observe sellers for 20 periods each, we find that the ratio we calculate here is approximately twice as big as the ratio that we would calculate had we access to the flows of monthly DSRs. This suggests that there is substantive persistence in seller performance, amounting to about 40–50 percent of the variance originating from the seller fixed effects.

[31] In terms of our model in Sec. V, a seller who plans to exit will profitably deplete his reputation capital by shirking, i.e., selling the low-quality good at a high price and not providing costly effort toward delivery resulting in stage payoff $p_h(k^E) > 0$ that eventually converges to zero with the depletion of reputation capital.

comparable sellers and buyers interact exactly in the same way as they did on eBay, except that there was no change to the feedback mechanism. Unfortunately, such a market does not exist. In the following and in online appendix C, however, we go through a list of competing explanations and conclude that none of those is likely to have caused the observed increase in DSRs.

First, the results could have been generated simply by grade inflation rather than seller effort. Figure 2 speaks against that, as there is no grade increase before but a significant one after the May 2008 change. This is confirmed by the results reported in column 3 in table 2. There was only a very small time trend before the change and none thereafter.

Second, before the change, buyers intending to leave a negative rating without retaliation could have done so by leaving a negative DSR. After the change, they could simply leave a negative classic rating and abstain from leaving a negative DSR. This would also lead to a decrease in DSRs. One way to test this is to check whether the number of DSRs relative to the number of classic ratings has decreased after the change. The numbers in table 1 already suggest that this was not the case. Toward a formal test we ran a regression, controlling for fixed effects, to estimate the change between that ratio on May 1, 2008, and July 1, 2009. In both cases, the ratio is for the preceding 12 months. This regression uses only sellers for which DSR ratings were available at both points in time. The ratio on May 1, 2008, is 0.4211 and the estimated change in the ratio is 0.0384, with a standard error of 0.0025. This shows that, if anything, the number of DSRs per classic ratings has slightly increased, invalidating the aforementioned concern.

Third, the ratings could also have increased in equilibrium because of a composition effect: sellers previously ranked highly could have absorbed a larger share and sellers ranked poorly a smaller share of the transactions. Table 2 gives evidence to the contrary: the number of DSRs remained unaffected for both above- and below-median sellers. More importantly, our results are robust to composition effects because DSRs are first aggregated at the seller level and only then averaged when generating the figures or performing the regressions. On top of that, the panel structure allows us to follow sellers over time, which we do by means of controlling for fixed effects in the regressions, and therefore, we also control for seller exit.

In online appendix C, we discuss four additional competing explanations: a change to the way auction listings were displayed, the introduction of discounts to power sellers, a shift from auctions to fixed-price offers, and changes in the macroeconomic environment. We conclude, on the basis of additional empirical evidence, that also those have likely not caused the increase in DSR ratings that we attribute to the increase in market transparency.

VII. Conclusion

The functioning of markets crucially depends on the way market participants behave, which in turn is related to market design. In this paper, we exploit changes in the mechanism by which traders can report on each other's performance to estimate the effect of increased market transparency on seller adverse selection and seller moral hazard.

Specifically, in May 2008, eBay changed its established nonanonymous feedback system from bilateral to essentially unilateral ratings, by allowing sellers to evaluate buyer behavior only positively rather than also neutrally or negatively as before. With this, eBay eliminated buyer fear of seller strategic retaliation to negative feedback given by buyers, which—by eBay's own argument—had resulted in underreporting of negative experiences.

One year before, eBay had introduced unilateral anonymous detailed seller ratings that already allowed buyers to rate sellers without a bias generated by fear of seller retaliation to a negative rating but retained the classic rating that, because nonanonymous, could be opportunistically biased. This gives rise to the research design we exploit: we use the DSRs as measures of seller behavior and study the effect of increasing market transparency, induced by the removal of buyer reporting bias via the May 2008 change in the classic rating mechanism.

We find that increased market transparency resulted in improved buyer satisfaction with seller behavior but no increase in the exit rate of poorly rated sellers. In fact, the poorly rated sellers' ratings improved more than average. We develop a simple model that focuses on the effects of this natural experiment and use it to provide a definition of moral hazard and adverse selection in this context and to interpret our empirical findings. Supported by a wealth of auxiliary empirical evidence we conclude that the removal of information bias in consumer reports, that is, an increase in market transparency, has a significant disciplining effect on sellers because it provides an additional incentive to them to exert effort. In combination with our finding that seller exit was not affected, this suggests that incentives given to them in this way resulted in positive welfare effects.

From a business policy point of view, we consider our analysis an interesting example of how relatively small changes in the design of an information mechanism can have economically significant effects. From the point of view of academic research, our study is, to the best of our knowledge, the first in which, at least for classical product markets, the effects of reducing buyer-seller informational asymmetries on adverse selection and moral hazard are clearly separated and directly juxtaposed to one another.

eBay is an important example of a market form that increases in importance from day to day. Similar reputation mechanisms are used to address the challenges associated with informational asymmetries also in other markets—most notably markets for travel, restaurant, and hotel services. This paper provides guidance on how their design could be improved.

Appendix A

Additional Tables and Figures

This appendix contains tables and figures that are referred to in the main text.

TABLE A1
Effect on Classic Feedbacks

	Bandwidth Foremost Sellers				Bandwidth Foremost Buyers			
	50 (1)	100 (2)	200 (3)	300 (4)	50 (5)	100 (6)	200 (7)	300 (8)
Local linear	−.365	−.544*	−.337	−.266	.038	.061	.027	.058
	(.398)	(.307)	(.223)	(.211)	(.208)	(.165)	(.120)	(.119)
Local quadratic	−.502	−.395	−.730***	−.767***	−.052	.053	−.037	−.064
	(.609)	(.456)	(.345)	(.331)	(.365)	(.239)	(.172)	(.175)

Note.—The table shows estimated effects of the feedback change on the percentage positive classic ratings received by users who were either foremost sellers or foremost buyers. These were obtained by performing nonparametric kernel regressions. We used an Epanetchnikov kernel. The cells contain estimates for local linear and local quadratic regressions, and the respective standard errors are in parentheses. Each column corresponds to a different bandwidth. To classify users, we used the ratio between DSR and classic feedbacks for the last year, on May 1, 2008. In particular, we classify those 25 percent users with the highest ratio as foremost sellers and the 25 percent with the lowest ratio as foremost buyers. This leads to 22,717 observations for the first group and 26,215 for the second group, coming from 1,168 and 1,169 users, respectively. Bootstrapped standard errors are cluster-robust at the seller level.

* Significant at the 10 percent level.

*** Significant at the 1 percent level.

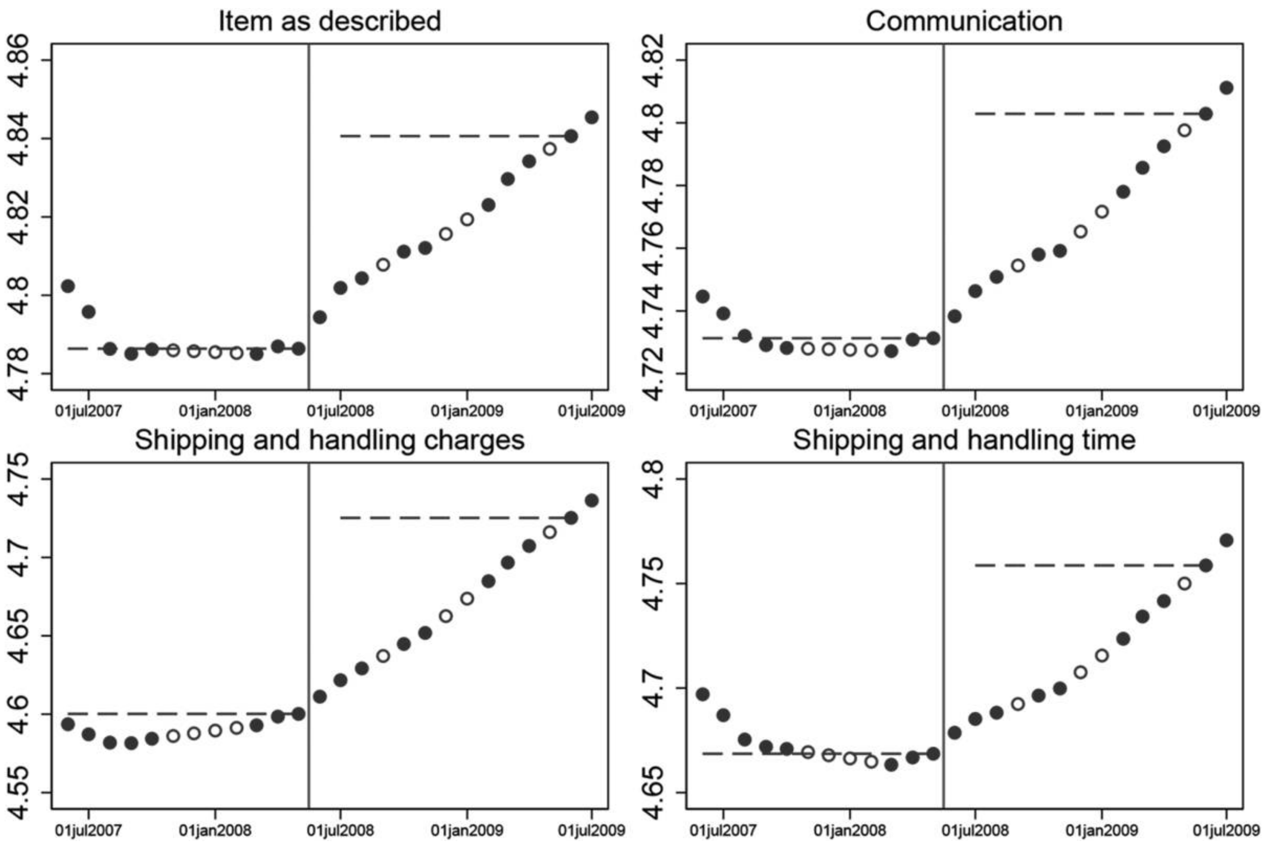

FIG. A1.—Average DSR score by category. The figure shows how the average of the four DSR rating categories changed over time. Color version available as an online enhancement.

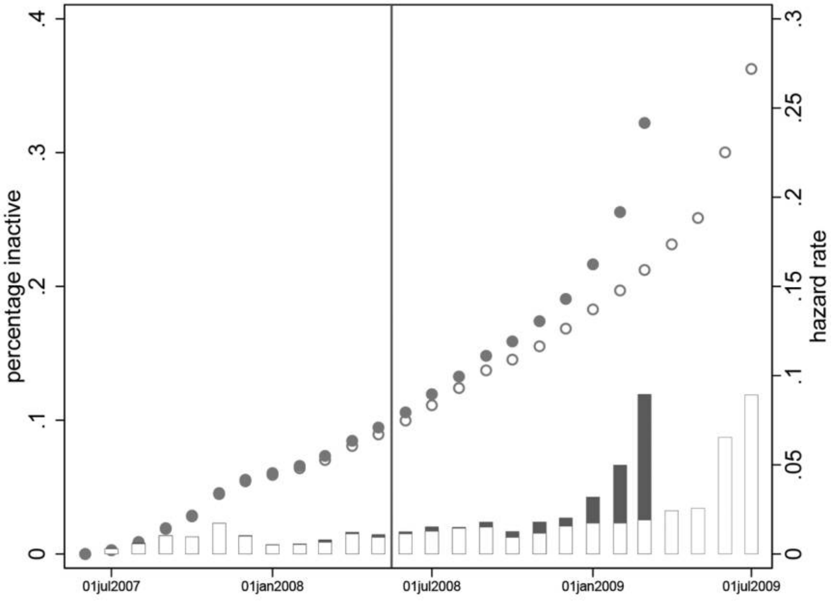

FIG. A2.—Truncation bias. The figure shows that truncation bias, as discussed in Section III, arises in the last 3 months. Nonfilled dots and bars in this figure correspond to the filled ones in figure 4. Filled dots are for the case in which we drop the last four waves of data and define inactivity as not observing any additional classic feedback until then. Filled bars are the resulting changes in the hazard rate. Only the last three estimates of the hazard rate, from January 2009 until March 2009, are affected by this. This suggests that the estimated hazard rates in figure 4 are not affected until April 2009. Color version available as an online enhancement.

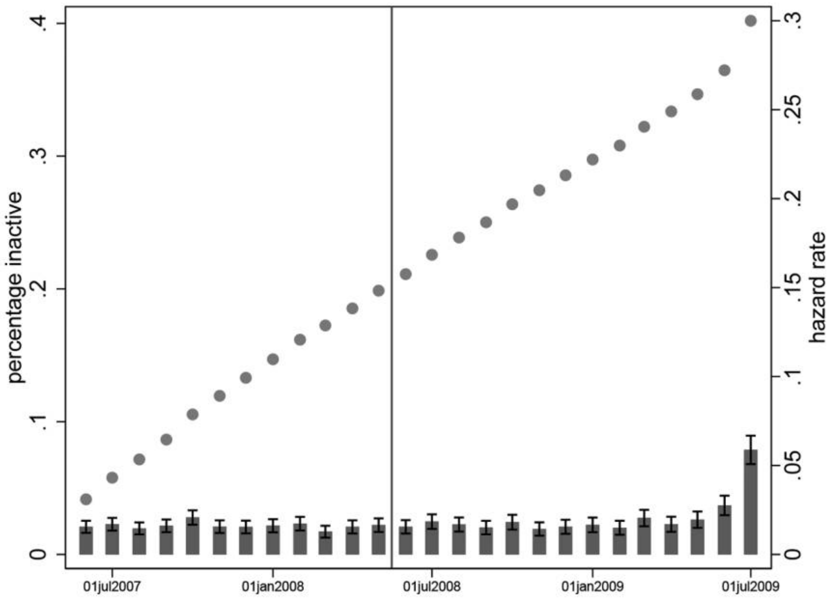

FIG. A3.—Simulated data. The figure shows the result of a Monte Carlo simulation based on the data used for figure 4. We reconstruct the classic feedbacks given for nine periods prior to the start of our data collection (see online app. B for details) and calculate the fraction of these periods in which a user had received classic feedback. We then simulate data using that rate, together with the assumption that at any point in time the probability of exiting is 1.5 percent. This generates an increase in the hazard rate in the last 3 months that solely arises because we misclassify users that are not active in every month as inactive. Color version available as an online enhancement.

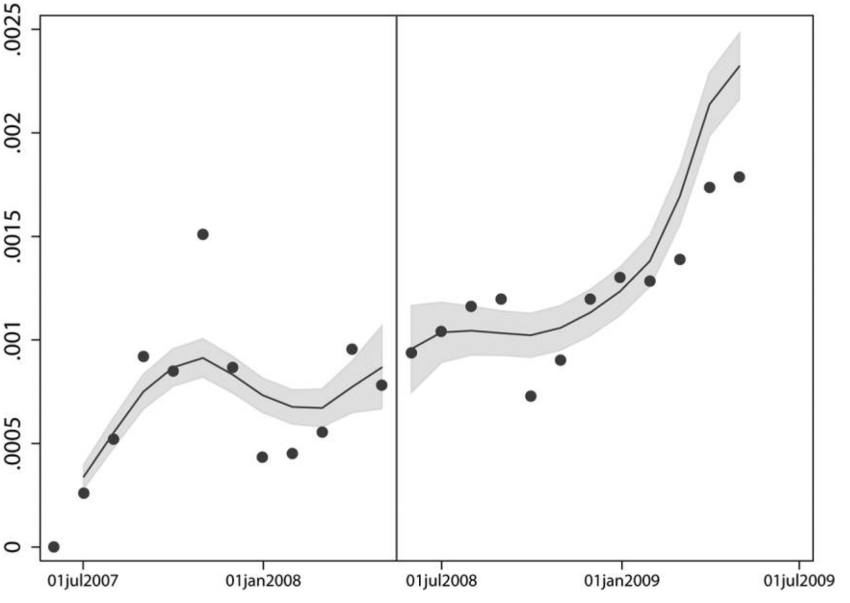

FIG. A4.—Density of the time of exit. The figure shows the density of the time of exit among those whom we classify as exiting until the end of the sample period. The dots are fractions of observations that exit in a 1-month time interval and correspond to bins in a histogram. See McCrary (2008) for details. Color version available as an online enhancement.

Figure A5 shows that exiting users behave worse in the 3 months prior to exiting, as compared to their performance before that. The error bars in this figure suggest that this difference is not significantly different from zero. However, they depict pointwise confidence intervals. The difference in the change of behavior over time between exiting and staying sellers is significantly different from zero when we pool over the time periods. The corresponding regression with standard errors clustered at the seller level shows that the point estimate of the intercept, which is the average over the dots for the stayers in the figure, is 1.001, with a standard error of 0.0001.[32] Statistically, the coefficient on an indicator for becoming inactive is −0.005 with a standard error of 0.0009. This means that the ratio is significantly lower for individuals who retire from the market, indicating that performance tends downward before retirement. The standard deviation of the ratio in a given wave, for example, May 2008, is 0.0077, so the effect is equal to 65 percent of this, which arguably is nonnegligible.

[32] The figure also shows that in the first 2 months, the difference between those turning inactive and those who do not is much bigger (the following squares are only linear interpolations). This is misleading, however, because the first two data points are based on fewer observations, as DSR scores are reported only if at least 10 DSR ratings were received.

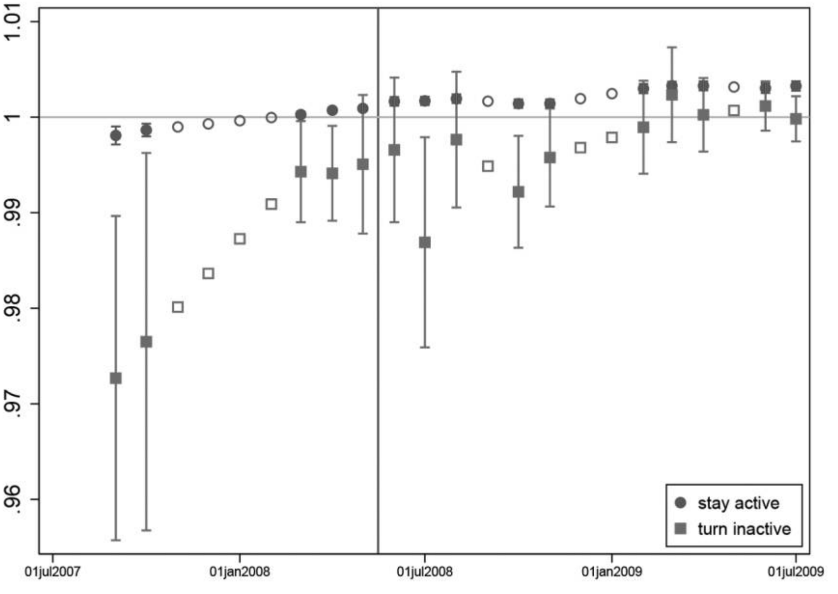

Fig. A5.—Behavior prior to inactivity. In this figure we compare the ratio between the current DSR and the DSR 3 months before for exiting users (depicted by the squares) to that of the stayers (depicted by the dots). We used linear interpolation in cases in which we did not collect data for the latter DSR score when calculating the ratio. Circles and squares are linearly interpolated values for the periods in which we did not collect data. The error bars depict pointwise 95 percent confidence intervals. Color version available as an online enhancement.

References

Anderson, M., and J. Magruder. 2012. "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database." *Econ. J.* 122 (563): 957–89.

Bajari, P., H. Hong, and A. Khwaja. 2014. "Moral Hazard, Adverse Selection and Health Expenditures: A Semiparametric Analysis." *RAND J. Econ.* 45 (4): 747–63.

Bajari, P., and A. Hortaçsu. 2004. "Economic Insights from Internet Auctions." *J. Econ. Literature* 42 (6): 457–89.

Bolton, G., B. Greiner, and A. Ockenfels. 2013. "Engineering Trust: Reciprocity in the Production of Reputation Information." *Management Sci.* 59 (2): 265–85.

Cabral, L. M. 2012. "Reputation on the Internet." In *The Oxford Handbook of the Digital Economy*, edited by M. Peitz and J. Waldfogel, 343–54. Oxford: Oxford Univ. Press.

Cabral, L. M., and A. Hortaçsu. 2010. "The Dynamics of Seller Reputation: Theory and Evidence from eBay." *J. Indus. Econ.* 58 (1): 54–78.

Dellarocas, C., and C. Wood. 2008. "The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias." *Management Sci.* 54 (3): 460–76.

Dranove, D., and G. Z. Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *J. Econ. Literature* 48 (4): 935–63.

Einav, L., A. Finkelstein, S. Ryan, P. Schrimpf, and M. Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *A.E.R.* 103 (1): 178–219.

Greenstone, M., P. Oyer, and A. Vissing-Jorgensen. 2006. "Mandated Disclosure, Stock Returns, and the 1964 Securities Acts Amendments." *Q.J.E.* 121 (2): 399–460.

Heckman, J. J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46 (4): 931–59.

Jin, G., and A. Kato. 2008. "Price, Quality, and Reputation: Evidence from an Online Field Experiment." *RAND J. Econ.* 37 (4): 983–1005.

Jin, G. Z., and P. Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *Q.J.E.* 118 (2): 409–51.

Klein, T. J., C. Lambertz, G. Spagnolo, and K. O. Stahl. 2006. "Last Minute Feedback." Discussion Paper no. 5693, Centre Econ. Policy Res., London.

———. 2009. "The Actual Structure of eBay's Feedback Mechanism and Early Evidence on the Effect of Recent Changes." *Internat. J. Electronic Bus.* 7 (3): 301–20.

Lucking-Reiley, D., D. Bryan, N. Prasad, and D. Reeves. 2007. "Pennies from eBay: The Determinants of Price in Online Auctions." *J. Indus. Econ.* 55 (2): 223–33.

McCrary, J. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *J. Econometrics* 142 (2): 698–714.

Melnik, M., and J. Alm. 2002. "Does a Seller's Reputation Matter? Evidence from eBay Auctions." *J. Indus. Econ.* 50 (3): 337–49.

Resnick, P., and R. Zeckhauser. 2002. "Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System." In *The Economics of the Internet and E-Commerce*, vol. 11, edited by M. Baye, 127–57. Amsterdam: Elsevier.

Resnick, P., R. Zeckhauser, J. Swanson, and K. Lockwood. 2006. "The Value of Reputation on eBay: A Controlled Experiment." *Experimental Econ.* 9 (2): 79–101.