

University of Mannheim / Department of Economics

Working Paper Series

***Market Transparency, Adverse Selection,
and Moral Hazard***

Tobias J. Klein Christian Lambertz Konrad O. Stahl

Working Paper 14-25

July 2014

Market Transparency, Adverse Selection, and Moral Hazard*

Tobias J. Klein

Christian Lambertz

Konrad O. Stahl[†]

July 2014

Abstract

We study how seller exit and continuing sellers' behavior on eBay are affected by an improvement in market transparency. 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 improving behavior—a reduction in moral hazard—, they preferred the latter because of lower cost. Increasing market transparency improved market outcomes.

JEL classification: D47, D83, L15.

Keywords: Anonymous markets, adverse selection, moral hazard, reputation mechanisms, market transparency, market design.

*We would like to thank the participants of the 2013 ICT Workshop in Evora, the 2nd MaCCI day in Mannheim, the 11th ZEW ICT Conference, the 2013 EARIE conference in Evora, the Fifth Annual Conference on Internet Search and Innovation at the Searle Center in Chicago, the Eighth bi-annual Postal Economics conference on E-commerce, Digital Economy and Delivery services in Toulouse, the Fifteenth CEPR/JIE Conference on Applied Industrial Organization in Athens, seminar participants at Hebrew, Tel Aviv and Tilburg universities, the 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. We are grateful to the editor, Ali Hortaçsu, and two referees for their detailed comments, as well as to Michael Anderson and Yossi Spiegel for unusually detailed and insightful suggestions on how to improve an earlier version of this paper. Financial support from the Deutsche Forschungsgemeinschaft through SFB TR-15 and Microsoft through a grant to TILEC is gratefully acknowledged. The latter was provided in accordance with the [KNAW Declaration of Scientific Independence](#).

[†]T. J. Klein: CentER, TILEC, Tilburg University; C. Lambertz: University of Mannheim; K. O. Stahl: University of Mannheim, C.E.P.R., CESifo, and ZEW. E-Mail for correspondence: T.J.Klein@uvt.nl.

1 Introduction

Informational asymmetries abound in anonymous markets, such as those opening in the internet on a daily basis. In particular, before trading takes place, the typical buyer does not know whether her anonymous counterpart, the seller she is confronted with, appropriately describes and prices the trading item, and whether he conducts the transaction conscientiously, so she receives the item in time and in good condition.

Without remedies, these informational asymmetries invite adverse selection and moral hazard. *Adverse selection* may arise along Akerlof's (1970) classical argument. Conscientious sellers leave—or may not even enter—the market, as long as their behavioral trait, and their effort, are *ex ante* unobservable to the buyers, and thus the buyers' willingness to pay or even to trade is hampered. For complementary reasons, opportunistically exploitative and careless sellers tend to self-select into such a market, because they can cheat on buyers by incorrectly claiming to offer high quality products and good delivery service. *Moral hazard* may arise because the provision of effort on both sides of the market is costly. Therefore, sellers may package goods badly; or delay, or default, on delivery. Likewise, buyers may delay, or default, on payments.

While the consequences of adverse selection and moral hazard are well understood conceptually, empirical tests on the direction of the effects as predicted by theory, and evidence on their magnitude are still scarce, and centered around insurance markets. In this paper, we use data on buyer satisfaction with seller behavior in an anonymous product market to show that an improvement in market transparency led to a significant increase in buyer satisfaction, but did *not* trigger a change in the exit rate of sellers from the market. We interpret the improvement in buyer satisfaction as reflecting an enhancement in seller behavior, and thus as a *reduction in seller moral hazard*. By contrast, we interpret the unchanged exit rate as *no change in seller adverse selection*.

Online markets provide a useful environment for collecting such evidence and conducting tests. Faced with adverse selection and moral hazard in combination with relatively high costs of legal enforcement, the market organizers designed remedies early on. In particular, they constructed mechanisms under which buyers and sellers mutually evaluate their performance; and published these ratings, so that agents on both sides of the market could build reputation capital. The reporting mechanisms were adjusted over time in reaction to opportunistic reporting

behavior on one or both sides of the market. These changes in the reporting mechanism are largely unexpected by the market participants, and thus can be perceived as natural experiments.

We collected data from eBay's web site before and after such a change. For four reasons, these data are particularly well-suited for studying the effects of increases in market transparency on moral hazard and adverse selection. First, eBay is one of the biggest markets ever to exist. Second, adverse selection and moral hazard are particularly important threats to the functioning of this market. Third, the change to the so-called classic reputation mechanism in eBay is a large-scale natural experiment and had a substantial impact on market transparency. And fourth, we are able to measure seller performance using an independent rating system that was not changed at the same time.

eBay's so-called classic reputation mechanism allows buyers and sellers to mutually evaluate their performance in just completed transactions. In May 2007, eBay added a new, second rating system, called Detailed Seller Ratings (DSR), that allows buyers to rate seller performance in detail—but not *vice versa*. The DSRs are reported as moving averages over the last 12 months, so unlike under the non-anonymous classic feedback scheme, the buyer's individual evaluation cannot be identified by the seller under the DSR scheme. One year later, in May 2008, eBay also changed the symmetry between buyer and seller rating in its classic feedback scheme, by forbidding a negative rating of the buyer by the seller and with it, removing buyer fear of seller retaliation to a bad rating by the buyer that was likely to have had an influence on buyer ratings before the change was enacted. This is the change whose effect we analyze, using the DSRs introduced a year before.

DSRs are suitable measures of buyer satisfaction, because on one hand they are transaction-based and on the other hand the buyer's evaluation of a particular transaction is not identifiable by the seller. Arguing that non-response bias may have been present but was not affected by that change, we show that the change in eBay's classic feedback mechanism led to a *significant and quantitatively important increase in buyer satisfaction with the incumbent sellers*, which, as we will claim, must have been due to an improvement in seller behavior. We also show that the May 2008 change in the classical feedback mechanism did *not lead to an increase in the exit rate of sellers*.

Adverse selection is a relevant issue in the world considered here. The reason is that we find sellers to be naturally and consistently heterogeneous in their behavior, as reflected in the

strategically unbiased buyer evaluations. In particular, sellers rated relatively poorly before the change tend to be rated relatively poorly thereafter. In view of this, the exit of poorly performing sellers constitutes a relevant option, as these sellers could expect to be rated even more poorly after the May 2008 change. Yet, rather than exiting from the market at an increasing rate, these sellers improve on all DSR dimensions, and this substantively and significantly more than the average incumbent sellers. This is surprising in view of the fact that the very introduction of DSR one year before the change analyzed by us may already have had a substantive impact on seller behavior—yet unobservable and much harder to quantify in the absence of strategically unbiased measures of buyer satisfaction.

Towards our interpretation of these results, we develop a toy stage game of an infinite period model, from which we predict effects of eBay’s removing negative buyer rating by the sellers on seller adverse selection and moral hazard. By our conceptualization, sellers differ by type—they are either conscientious or exploitative—and by the dis-utility they suffer from providing effort towards satisfying a buyer. Removing buyer fear of adverse retaliation by the seller incentivizes the buyer to report truthfully rather than opportunistically; and in particular to report bad experiences. That, in turn, leads the seller to change his behavior, in alternatively two ways: first, the exploitative seller with high dis-utility of effort may leave the market, thus ameliorating adverse selection; second, both seller types, if remaining in the market, may engage in more effort towards improving on buyer satisfaction, thus ameliorating moral hazard.

We assess in detail the robustness of our empirical results against alternative interpretations, and exclude the possibility that other developments have led to the observed significant increase in buyer satisfaction. This is important because our key identifying assumption is that average ratings would not have changed without the change to the feedback system. In particular, we document empirical patterns supporting our assumption that buyer ratings of individual transactions are systematically related to seller behavior. We also establish that there was no time trend in average ratings before the change; that effects of other changes in eBay’s allocation mechanism—most importantly the introduction of Best Match—can be isolated and did not affect our results; and that changes in other factors such as the decreased popularity of the classical auction format, or macroeconomic factors, could all not have led to the observed increase in buyer ratings after the change to the feedback system.

In all, this allows us to conclude that the interpretation based on our toy model fits best,

that is: *the reduction of the informational asymmetry due to a reduction of buyers' reporting bias disciplines sellers, and results in a reduction of seller moral hazard rather than a reduction of seller adverse selection.* On the basis of our toy model, the reason is that the additional costs sellers incur when they change their behavior are actually small relative to the benefits associated with being active, even for the badly performing exploitative seller types. Otherwise, we should have observed an increased rate of exit after the increase in transparency.

An increase in transparency reduces buyer regret and thereby leads to higher quality outcomes. Given the small cost of implementing the observed change in the reporting mechanism, 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.¹

As online markets gain in importance from day to day, these results strike us as important *per se*. They should also apply to the emerging situations in which reporting mechanisms can discipline seller behavior in other markets, most notably markets for hotel, restaurant, and travel services.

Towards detailing procedure and results, we proceed as follows. In the next Section 2, we relate what we do to the pertinent literature. In Section 3, we describe the eBay Feedback Mechanism and in particular the change we focus on. Section 4 contains the description of our data. In Section 5, we present our central results. In Section 6 we develop our toy model, from which we derive our preferred explanation and interpretation of the results. In Section 7 we provide additional support especially of our assumption that seller performance is well reflected in the buyers' evaluations. We also defend in much detail our preferred interpretation against competing ones. We conclude with Section 8.

2 Literature

Most empirical studies on adverse selection and moral hazard are on insurance markets. Take a car insurance company and an individual insurance taker. Adverse selection arises, for instance, if the individual, knowing to be an unsafe driver, self selects into buying (high) insurance cover-

¹A 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 we could compare the increase in consumer surplus to the decrease in seller rents.

age, rather than staying out of the insurance market (or buying low insurance coverage). Moral hazard is present if the insuree pursues lesser accident-preventing effort in the face of the insurance coverage, and is therefore more likely to have an accident. In our case, adverse selection arises if a seller endowed with exploitative preferences enters the market. Moral hazard arises if a seller does not wage effort into a good's appropriate and timely delivery.

Returning to the car insurance example, it is difficult to disentangle adverse selection and moral hazard by just observing the incidence of accidents conditional on coverage. Therefore, authors have first focused on showing that asymmetric information affects economic outcomes. This is done by relating individual choices to *ex post* outcomes (see [Chiappori, 2000](#), for an early review). For example, [Chiappori and Salanié \(2000\)](#) use data on contracts and accidents in the French market for automobile insurance to test whether insurance contracts with more comprehensive coverage are chosen by individuals who then have higher claim probabilities. If this is the case, then this can either be explained by moral hazard, or adverse selection, or both, without possibility of further discrimination.²

From the theorist's point of view, the inability to disentangle adverse selection and moral hazard effects does not come as surprise: the analyst typically cannot observe self selection *ex ante* by type, because the type is largely private information. In addition, with an endogenous change of effort, that type can modify the outcome.³ In our case, the generically poor seller can mimic a conscientious seller with such an endogenous change of effort.

Following up on [Chiappori and Salanié \(2000\)](#), [Abbring, Chiappori, and Pinquet \(2003\)](#) and [Abbring, Heckman, Chiappori, and Pinquet \(2003\)](#) show that dynamic insurance data allow researchers to isolate moral hazard effects, by looking at insurance contracts in which the financial loss associated with a second claim in a year is bigger, so that exercising moral hazard becomes more costly, and therefore the incentive to do so decreases. One can isolate moral hazard effects in this context because one naturally follows an individual over time, and therefore the factors influencing adverse selection stay the same, while incentives to exert moral hazard change. This is also the approach we take. At any rate, in the context of deductibles in health insurance [Aron-Dine, Einav, Finkelstein, and Cullen \(2012\)](#) also follow-up on this idea and investigate whether individuals exhibit forward looking behavior, and reject the hypothesis

²See also [Finkelstein and Poterba \(2004\)](#) for a similar approach in the context of annuitization and mortality and [Fang, Keane, and Silverman \(2006\)](#) in the context of health insurance.

³See, for instance, [Laffont and Martimort \(2002\)](#), Ch. 7.

of myopic behavior.

Focusing on adverse selection, [Einav, Finkelstein, Ryan, Schrimpf, and Cullen \(2011\)](#) show that some individuals select insurance coverage in part based on 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 \(2006\)](#) also study individual selection of insurance contracts. They provide, as we do in our very different context, 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 exogenous variation, they need to control in an elaborate way for individual risk and risk preference.

We instead develop our results from a natural experiment, involving, in our interpretation, self selection and adjustment of moral hazard *ex post*. We follow sellers over time, which allows us to control for unobserved differences across sellers by means of fixed effects when studying moral hazard. We then study whether an improvement of the mechanism led to increased exit from the market on the one hand, and/or increased seller effort on the other hand.

The institutional change the effects of which we study led to an increase in market transparency. Market transparency also plays an important role in several other, distinct literatures. In the context of restaurants, [Jin and Leslie \(2003\)](#) show that quality disclosure for restaurants, by means of 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. Also in finance, there is a literature on the effects of mandatory disclosure. For instance, [Greenstone, Oyer, and Vissing-Jorgensen \(2006\)](#) show that financial investors valued an extension of disclosure requirements by documenting abnormal returns for firms that were most affected by this. In the context of competition policy, [Henze, Schuett, and Sluijs \(forthcoming\)](#) conduct an experiment in which they vary the extent to which consumers are informed about quality. They find effects of this on the quality firms provide in equilibrium, and conclude that information disclosure is a more effective tool to raise welfare and consumer surplus than theory would lead one to expect. At the same time, market transparency is not always easy to achieve. [Mayzlin,](#)

[Dover, and Chevalier \(forthcoming\)](#) show that firms that are being reviewed online, in their case hotels, actively manipulate those reviews if they have a possibility to do so.

There is also a literature on quality disclosure in electronic markets, which in turn is related to [Avery, Resnick, and Zeckhauser's \(1999\)](#) somewhat sweeping general hypothesis that the internet has greatly reduced the cost of distributing information and that there is an efficient provision of evaluations by users. [Dranove and Jin \(2010\)](#), [Bajari and Hortasçsu \(2004\)](#) and [Cabral \(2012\)](#) provide reviews of the theoretical and empirical literature on quality disclosure on the internet. For eBay, the general finding is that better ratings benefit sellers by an increase in the probability to sell a product, and in its selling price. See, e.g., [Melnik and Alm \(2002\)](#), [Lucking-Reiley, Bryan, Prasad, and Reeves \(2007\)](#) and [Jin and Kato \(2008\)](#) for evidence using field data, and [Resnick, Zeckhauser, Swanson, and Lockwood \(2006\)](#) for experimental evidence.

These results show that ratings on eBay convey information, but it is unclear how much. The reason is that, due to the design of the reputation mechanism, ratings were biased before the implementation of DSR, and the removal of symmetric classic feedback. [Resnick and Zeckhauser \(2002\)](#) provide reduced-form evidence that points towards underreporting of negative experiences, and [Klein, Lambertz, Spagnolo, and Stahl \(2006\)](#) complement this by showing that the probability to leave a negative rating increases substantially towards the end of the period in which feedback can be left.

[Klein, Lambertz, Spagnolo, and Stahl \(2009\)](#) provide detailed information on the actual structure of the feedback mechanism and provide first descriptive evidence on the newly introduced DSRs. [Bolton, Greiner, and Ockenfels \(2013\)](#) also provide such evidence and complement it with an experimental study. Focusing on why classic ratings are left at all, [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 that comes from traders being much more likely to leave a rating when satisfied.

[Cabral and Hortasçsu \(2010\)](#) provide evidence that is consistent with seller moral hazard. They find that just before exiting, sellers receive more negative feedback than their lifetime average. With our paper we complement the aforementioned studies by providing direct evidence on one of the most policy-relevant questions, namely the relationship between the design of the feedback mechanism and the presence of moral hazard or adverse selection.

Finally, [Nosko and Tadelis \(2014\)](#) suggest that platforms should more actively screen sellers and promote listings of better quality sellers. They develop a measure of seller quality and demonstrate its usefulness through a controlled experiment on eBay that prioritizes better quality sellers to a random subset of buyers.

3 eBay’s Feedback Mechanism

eBay’s feedback mechanism by which sellers and buyers could evaluate the performance of their trading partners was introduced in February 1996, just a few months after the first auction had taken place on its website.⁴ In its earliest form, the system allowed any eBay user to leave feedback on the performance of any other user in the form of a “positive,” “neutral,” or “negative” rating 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.

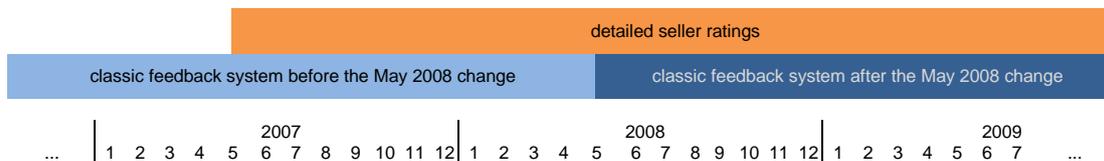
In February 2000, four years after its institution, the mechanism was changed to transaction-specific feedback. Since then, all new ratings must relate to a particular transaction, i.e. only the seller and the buyer in a particular transaction can rate each other regarding their performance in that transaction.

From early on, the feedback mechanism has led to conflicts and heated discussions about unfairly biased reporting. As a consequence, eBay repeatedly modified the system. In May 2007, eBay introduced a new form of unilateral rating by buyers: Detailed Seller Ratings (DSR). In addition to the original bilateral rating available heretofore, buyers could now separately rate, with one to five stars, the *accuracy of the item description*, *communication*, *shipping speed*, and *shipping charges*. These detailed ratings are left unilaterally by the buyer. They are anonymized by being published in aggregate 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.

This change addresses what was felt to be a substantial flaw in eBay’s original bilateral feedback mechanism, namely the buyer’s fear of retaliation when leaving a negative rating *before* the seller—a problem well known to many eBay users and well discussed among scholars for some

⁴An early description of the basic mechanism and an analysis of rating behavior are given in [Resnick and Zeckhauser \(2002\)](#).

Figure 1: Changes to the feedback mechanism



time.⁵ An important detail is that DSRs can only be left 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. Note that the two ratings are *not* perfect substitutes. In particular, the DSRs give an evaluation of the seller’s behavior on average, and the classical ratings show how the seller behaved at the margin, i.e. 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 as well: sellers could only leave 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.⁶ The two changes are summarized 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.

Anonymity ensures that buyers can leave a DSR without threat of retaliatory feedback by the seller. The buyer’s evaluation is nevertheless subjective. But buyer specificity strikes us as immaterial here, because sellers receive ratings from a large number of buyers and we use the average rating as a measure of seller behavior. Indeed, this would be the close-to-ideal measure for the purpose of this study, if rating standards could be ensured to stay the same

⁵Klein, Lambertz, Spagnolo, and Stahl (2006) gave an early account of this.

⁶In fact, eBay stated the reasons for this step in a public announcement in January 2008: *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).* (Taken from <http://announcements.ebay.com/2008/01/a-message-from-bill-cobb-new-pricing-and-other-news/>, last accessed in June 2013.) 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.

over time, and every buyer would leave a rating. In favor of the former, eBay displays a verbal meaning to every star rating in every category when ratings are given, which makes it more likely that the typical buyer’s rating behavior does indeed not systematically change over time. For instance, a rating of 4 stars in the rating category shipping speed means that the seller shipped the item “quickly”. As for the latter, non-response 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 7.1, 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.

Based on 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.

4 Data

Our data contain monthly information on feedback received by about 15,000 eBay users over a period of three years, between July 2006 and July 2009. The data were collected from eBay’s U.S. 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 & Notebooks, (2) Apple iPods & Other MP3 Players, (3) Model Railroads & Trains, (4) Trading Cards, and (5) Food & Wine.⁷ 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. From June 1, 2007 onwards 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

⁷See Table 5 in Appendix A for the exact categories.

2006 was inferred from the data drawn in June, 2007 and later.⁸

Towards 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. 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.^{9,10}

Out of the 15,000 user names we drew in May 2007, we were able to download feedback profiles for 14,937 unique users in our first data collection effort on June 1, 2007.¹¹ One year later, we could still download data for 14,683 users, and two years later for 14,360 users.¹²

Table 1 gives summary statistics. As described above, the first data collection took place on June 1, 2007. On that day, the average user in our sample was active on eBay for almost four years. Proxying user experience by the length of time a user has registered, the most experienced user in our sample had registered with eBay more than eleven 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 two years before the May 2008 changes.

On eBay, the feedback score is given by the number of distinct users who have left more

⁸See Figure 7 in Appendix A 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 of 2008; and January and May 2009. As we explain in Section 5 and Appendix A, 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 our data collection design is to follow sellers over time and that therefore, our data are not informative about seller entry.

⁹This means that we are more likely to misclassify infrequent sellers as inactive towards the very end of our observation period, as the time window is truncated in which we observe no change in the reputation record. We show below that this is only the case when we have less than three months of data on future ratings. Therefore, this is unlikely to affect our results, because we have collected data for more than a year after the feedback change, and we mostly use information around the May 2008 change to the system.

¹⁰This 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 Appendix A. If users are equally likely to stop being active as a buyer before and after the change to the classic feedback mechanism, 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 exitors 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.

¹¹There 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.

¹²We 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

	obs.	mean	std.	percentile				
				5	25	50	75	95
June 1, 2007								
duration membership in years	14,937	3.83	2.76	0.09	1.33	3.54	6.12	8.46
feedback score	14,937	563.66	2704.53	0.00	18.00	88.00	339.00	2099.00
percentage positive classic ratings	14,189 ^a	99.09	5.67	97.10	99.70	100.00	100.00	100.00
member is PowerSeller	14,937	0.07	-	-	-	-	-	-
number classic ratings previous 12 months	14,937	273.10	1351.22	0.00	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	1247.49	0.00	10.00	45.00	164.00	1042.00
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	1240.91	12.00	28.00	78.25	265.50	1378.25
DSR score	4,429 ^d	4.71	0.19	4.35	4.65	4.75	4.83	4.90
number DSR relative to number classic feedbacks	4,429 ^d	0.42	0.19	0.10	0.27	0.44	0.59	0.70
June 1, 2009								
number classic ratings previous 12 months	14,360	200.47	1039.00	0.00	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	1249.90	12.00	26.38	72.00	255.75	1378.00
DSR score	3,272 ^f	4.78	0.16	4.53	4.73	4.82	4.88	4.95
number DSR relative to number classic feedbacks	3,272 ^f	0.46	0.20	0.11	0.29	0.48	0.63	0.74

Notes: Table shows summary statistics for our sample of sellers. The three panels reflect consecutive points in time for which we report summary statistics: The day at which we first collected data, as well as one and two 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 one year after that. The feedback score 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 PowerSeller 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. ^aCalculated for those 14,189 users whose feedback score is positive. ^bCalculated for those 13,943 users who received classic feedbacks in the previous 12 months. ^cCalculated for those 13,811 users who received classic feedbacks in the previous 12 months. ^dCalculated for those 4,429 users who received enough DSRs so that the score was displayed. ^eCalculated for those 11,524 users who received classic feedbacks in the previous 12 months. ^fCalculated for those 3,272 users who received enough DSRs so that the score was displayed.

positive classic ratings than negative ones, minus the number of users who have left more negative ratings than positive ones. At the time the observation window opened, the mean feedback score of our users was 564, the median score was 88, and 769 users had a feedback score of zero. The average share of positive feedback users had received over the last twelve 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 positive ratings was very similar. On June 1, 2008, statistics for the DSRs are available for the 4,429 users who received more than 10 DSRs. The reason is that otherwise, anonymity of the reporting agent would not be guaranteed, as a seller could infer the rating from the change in the DSR. DSR scores are available for about 15 percent of the users one month after their introduction in May 2007, and for about 30 percent of users one year later. The DSR score we report on here and use in our analysis is the average reported score across the four rating dimensions. Yet another year later, the picture looks again 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 unbiased ratings by their buyers effective with the introduction of DSR one 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 only be received by users when acting as sellers. Hence, the average DSR score will reflect only how a user behaved in that very role.¹³ Still, it is also important to keep in mind that we will not be able to observe the reaction of sellers who receive less than 10 DSRs per year.¹⁴ However, looked at it in a different way, we capture behavior that is associated with most of the transactions on eBay, as those sellers who

¹³One may still wonder how often the users in our sample acted as buyers. On June 20th, 2008, eBay reveals in a statement that buyers leave DSR 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 it 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 only be left when a classic rating is left at the same time. It is a rather conservative lower bound because it assumes that a user receives a DSR every time he acts as a seller and it takes a bit more effort for the buyer to leave a DSR, as compared to leaving only a classic rating.

¹⁴This is one of the reasons why we will control for seller fixed effects. It is important to do so because sellers for whom DSRs are available may be different from those for whom DSRs are not available; and because sellers who exit at some point may be different from those who will not exit. See also the discussion below.

receive less than 10 DSRs per year are not involved in most of the sales on eBay.

5 Results

5.1 Staying Sellers' Reactions

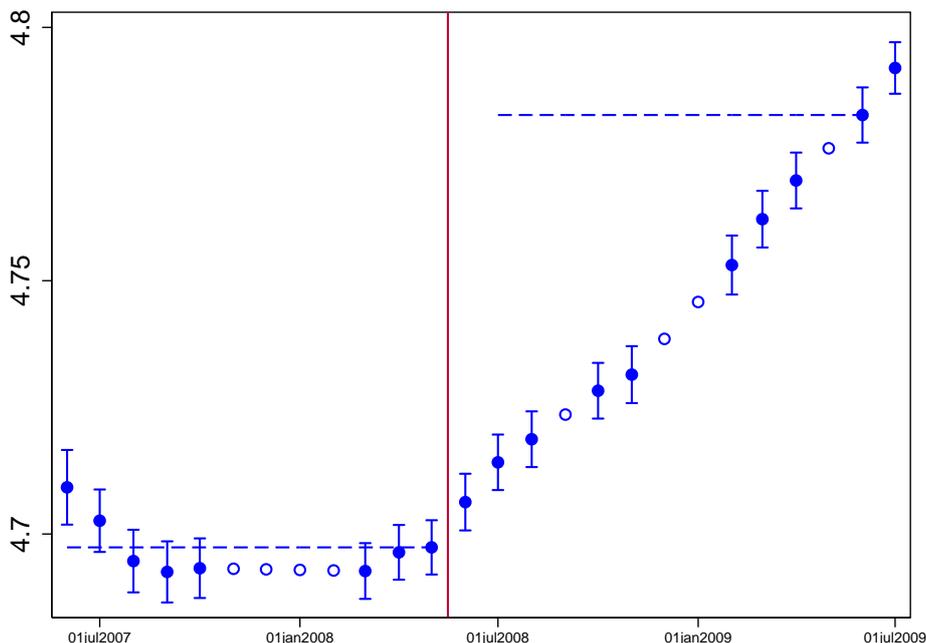
After the introduction of DSRs in May 2007, the May 2008 change to the classic feedback system provided additional means to buyers to non-anonymously voice negative experiences without fear of negative seller reaction. In the first instance, this should have led buyers to voice more critical reactions to sellers' activities by means of classic ratings. We document this in Section 7.1 below. At the same time, the DSR system remained unchanged in May 2008. This allows us to attribute changes in DSR ratings over time to changes to the classic feedback mechanism that led to higher market transparency.

In particular, the newly established possibility to buyers' freely voicing critique by means of negative classic ratings should have incentivized continuing sellers to prevent negative buyer ratings by significantly reducing shirking, i.e. not describing and pricing goods as of higher than the true quality; and increasing their effort in the other dimensions towards satisfying buyers. Therefore, we expect a significant increase in buyer satisfaction, as measured by the DSR scores that, however, expresses itself in a fashion dampened by the way the DSRs are displayed—as moving averages.

Figure 2 shows how the average DSRs in the relevant two 12-month intervals before and after the May 2008 change evolved over time.¹⁵ In the figure, each dot reflects the overall average across sellers and categories using only observations with complete data. For that reason, especially the first two dots in the figure cannot be compared to the remaining ones. All dots are averages for the selected sample of sellers who conducted enough transactions so that a DSR score was already 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 two months, and those even more selected sellers receive higher DSRs on average. In our regression analysis below, we take this into account by controlling for fixed effects.

¹⁵Recall 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. Figure 9 in Appendix B shows that the patterns by category resemble one another closely. Therefore, we will from now on use the average DSR across rating categories.

Figure 2: Evolution of Detailed Seller Ratings



Notes: Figure shows how DSRs changed over time. 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 8, Appendix A, 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. The capped spikes are 95 percent confidence intervals.

When interpreting Figure 2 it is important to once again 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 one year after the change and the DSR score right before the change. It is depicted in the horizontal lines in

Figure 2.¹⁶ The figure clearly shows that the DSRs have increased after the May 2008 change.¹⁷

We performed regressions to quantify the effect shown in Figure 2, controlling for fixed effects.¹⁸ Denote by DSR_{it} the average score across the four DSR rating dimensions reported for seller i in period t . Recall that our data is 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 $w_{i\tau}^t$ be the weight put in the construction of the index on $d sr_{i\tau}$, 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 number of ratings received in τ divided by the total number of ratings received between period $t - 12$ and $t - 1$. Hence $\sum_{\tau=t-12}^{t-1} w_{i\tau}^t = 1$ and

$$DSR_{it} = \sum_{\tau=t-12}^{t-1} w_{i\tau}^t \cdot d sr_{i\tau}. \quad (1)$$

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

$$d sr_{i\tau} = \alpha + \beta \cdot POST_{i\tau} + \alpha_i + \varepsilon_{i\tau},$$

where $POST_{i\tau}$ takes on the value 1 after the change, and zero otherwise. The change occurred between the 1st of May and the 1st of June, 2008, and therefore we code $POST_{i\tau} = 1$ if τ is equal to July 2008, or later, and $POST_{i\tau} = 0.5$ 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.¹⁹ 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 $d sr_{i\tau}$ on $POST_{i\tau}$ because

¹⁶The 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 7 in Appendix A shows at which points in time data were collected and depicts over which periods, respectively, the DSR scores were calculated.

¹⁷Unfortunately, we were not able to collect data for more than one 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 one year after the change. The remarkable fact, however, is that the scores start increasing rapidly and immediately after the change.

¹⁸In Section 7.2 on competing explanations, we also control for other relevant changes implemented by eBay in the observation window

¹⁹This is conservative in the sense that, if anything, it would bias our results downwards because we would 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 Section 7.2 on competing explanations, and the robustness check in Section C.

$dsr_{i\tau}$ is not observed. However, by (1), the reported DSR score is the weighted average rating received in the preceding 12 months, so that

$$DSR_{it} = \alpha + \beta \cdot \left(\sum_{\tau=t-12}^{t-1} w_{i\tau}^t \cdot POST_{i\tau} \right) + \alpha_i + \left(\sum_{\tau=t-12}^{t-1} w_{i\tau}^t \cdot \varepsilon_{i\tau} \right). \quad (2)$$

$\sum_{\tau=t-12}^{t-1} w_{i\tau}^t \cdot POST_{i\tau}$ 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.²⁰ We can control for time trends in a similar way.²¹

It is important to control for fixed effects in this context because at any point in time the DSR score is only observable for a selected sample of sellers, namely those who were 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 the 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 shows 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 this global 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

²⁰One 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 $POST_{i\tau}$ 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 abovementioned condition the covariance between the regressor and the error term is zero.

²¹For 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

	(1) full sample	(2) small window	(3) time trend	(4) DSR < 4.75	(5) DSR ≥ 4.75
average DSR before change	4.7061*** (0.0007)	4.7030*** (0.0005)	4.7149*** (0.0034)	4.5912*** (0.0011)	4.8138*** (0.0006)
effect of feedback change	0.0581*** (0.0024)	0.0414*** (0.0047)		0.0904*** (0.0044)	0.0316*** (0.0021)
effect of feedback change until September 2008			0.0168** (0.0083)		
effect of feedback change after September 2008			0.0589*** (0.0184)		
linear time trend before change			0.0009** (0.0004)		
linear time trend after change			0.0007 (0.0019)		
fixed effects	yes	yes	yes	yes	yes
R^2	0.0580	0.0131	0.0605	0.0809	0.0466
number sellers	5,224	4,919	5,224	2,337	2,337
number observations	67,373	30,488	67,373	31,260	33,508

Notes: 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 do 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. Specification (4) includes only those sellers who had a DSR score below the median of 4.75 in May 2008 and (5) only those above the median. One observation is a seller-wave combination. Throughout, we control for fixed effects. R^2 is the within- R^2 . Standard errors are cluster-robust at the seller level, and significance at the 5 and 1 percent level is indicated by ** and ***, respectively.

1, 2008. In specification (3), we instead allow for a piecewise linear time trend over the entire 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. After that, DSR scores increase almost linearly over time, but this is driven by the fact that DSR scores are averages over DSRs received in the previous year, and the fraction of DSRs received after May 2008 increased gradually over time. Consequently, the DSR scores will also only increase gradually, even if the flow of DSRs jumps up and remains unchanged at a higher level after the change. 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.²²

²²Without 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. Then, the magnitude of the short run effect is comparable to the one of the effect using the smaller sample that is reported in column (2). We obtain similar estimates when we define the short run to last longer or shorter than three months.

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.

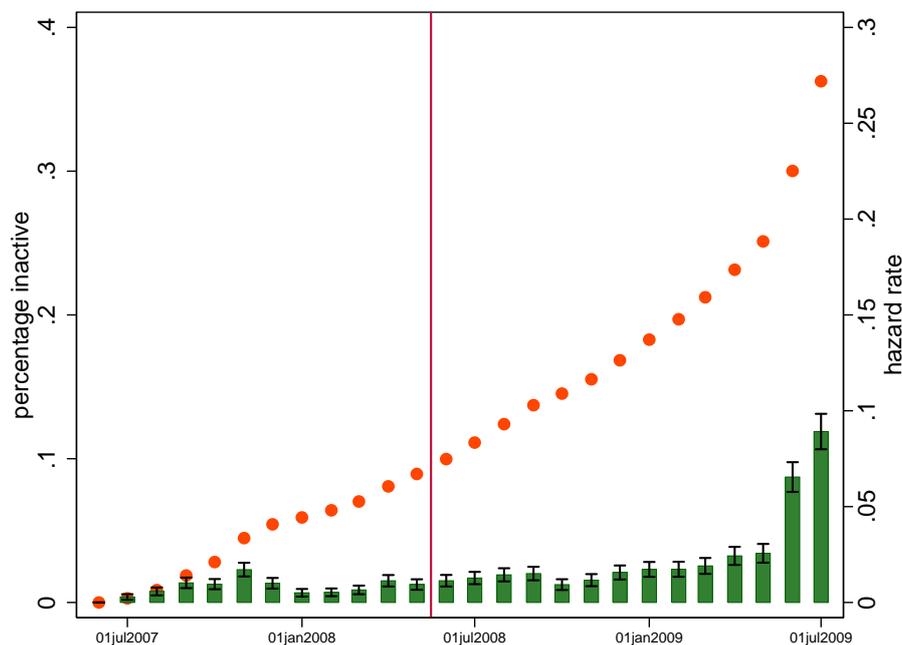
We also looked at how this increase is differentiated between sellers with low, and high DSR before the change. Towards that, we split our sample at the median DSR of 4.75 between high and low ranked sellers just before the May 2008 change. Figure 10 in Appendix B gives the picture. The increase in DSR score is stronger for sellers with below-median score *ex ante*. The last two columns 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.²³ 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 6 in Appendix B 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.

Recall again that the system change was not with respect to DSR, but with respect to the classic reporting mechanism. The anonymous and unilateral DSR were established one year before the May 2008 change whose consequences we consider here, and they remained anonymous and unilateral thereafter. Already with the DSR introduced in May 2007, buyers had been able to express their true valuation of seller performance without fear of retaliation by that seller. By looking at the effect of changing the non-anonymous established reporting mechanism, we pick up only an additional effect. It is remarkable that this effect shows up as clearly as documented above.

In all, the empirical evidence provides support of our hypothesis that abandoning negative buyer rating by sellers—and thereby reducing impediments against negative seller rating by

²³One concern concern may be that the increase for the sellers with low DSR before the change may be driven by mean reversion. Indeed, we have divided sellers based on 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 only increase after the change. This shows that mean reversion is not of concern here.

Figure 3: Exit from the market



Notes: 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.

buyers in the classic rating system—has led to significant and substantive increases in the buyers’ evaluations as measured by the independently measured DSRs. The DSR ratings improved significantly. As we will argue below, this results from an improvement in the behavior especially of the sellers rated poorly before the change, and with this, a reduction in moral hazard.

5.2 Seller Exit

The May 2008 change allowed buyers to non-anonymously voice negative experiences, without fear of negative seller reaction. The results shown Section 5.1 above suggest that this led to a significant and substantial increase in buyer satisfaction. At the same time, it could also have motivated poorly performing sellers to leave the market. Figure 3 shows how the fraction of individuals who have become inactive, and the corresponding hazard rate into inactivity, changed over time. 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.²⁴ The fraction of individuals becoming inactive increases more in the last months of our observation period and that the corresponding hazard rates are higher. In Figures 12 and 13 in Appendix B we show that the increase in that fraction towards the end of that window, and the corresponding increase in the hazard rate in the last three months, can be attributed to truncation bias. It arises because we define exit as the first point in time from which we do not observe a user to receive any classic ratings anymore. To see why this generates truncation bias suppose that a user is only active and receives a rating 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 only be observed to be active again in August. This example shows that we will be more likely to misclassify individuals towards the end of the sample and that the likelihood is related to how active a user is.

Coming back to the pattern in Figure 3, we see that overall, many sellers leave over time, both before and after the change. By June 1, 2009, about 25 percent of the sellers have become inactive. As before, we split the sample into sellers with above and below median DSR score prior to the May 2008 change.²⁵ By May 1, 2009, 33 percent of the below-median sellers have left the market, compared to 18 percent of the above-median ones. Figure 11 in Appendix B shows the corresponding hazard rates.

The most important finding for both good and bad sellers is that *the May 2008 change did not trigger any significant increase in the exit rate of sellers*. We also formally tested whether the hazard rate was different before and after the change. Towards this, we conducted OLS 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 where sellers are at risk of exiting, i.e. have not exited yet. The results are shown in Table 3. There is no statistically significant increase in the exit rate, with the baseline exit rate higher for below-

²⁴In Figures 2 and 10 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.

²⁵Unlike in our analysis of the evolution of DSR scores before, we use here a linearly extrapolated value if the DSR score is only available 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 Figure 3, for the reasons given in Footnote 24. 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 only happen if they exit after that point in time. This would then lead to an upward bias in the hazard rates after May 2008.

Table 3: Effect of the May 2008 change on seller exit

	(1) full sample	(2) small window	(3) DSR < 4.75	(4) DSR ≥ 4.75
exit rate before change	0.0119*** (0.0014)	0.0104*** (0.0011)	0.0181*** (0.0026)	0.0065*** (0.0017)
effect of feedback change	0.0004 (0.0018)	0.0017 (0.0015)	0.0001 (0.0032)	0.0007 (0.0020)
linear time trend before change	0.0015*** (0.0004)		0.0021*** (0.0007)	0.0009** (0.0005)
linear time trend after change	0.0002 (0.0003)		-0.0003 (0.0005)	0.0006* (0.0003)
R^2	0.0009	0.0001	0.0008	0.0013
number observations	56,467	19,119	26,157	30,310

Notes: Table shows 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 specification (1), (3) and (4). In specification (2), we exclude observations before April, and after July 2008. Specification (4) includes only those sellers who had a DSR score below the median of 4.75 in May 2008 and (5) only those above the median. We used an extrapolated value if the DSR score was only available at a later point in time. One observation is a seller-wave combination provided that the seller has not left before. Robust standard errors in parentheses. Significance at the 5 and 1 percent level is indicated by ** and ***, respectively.

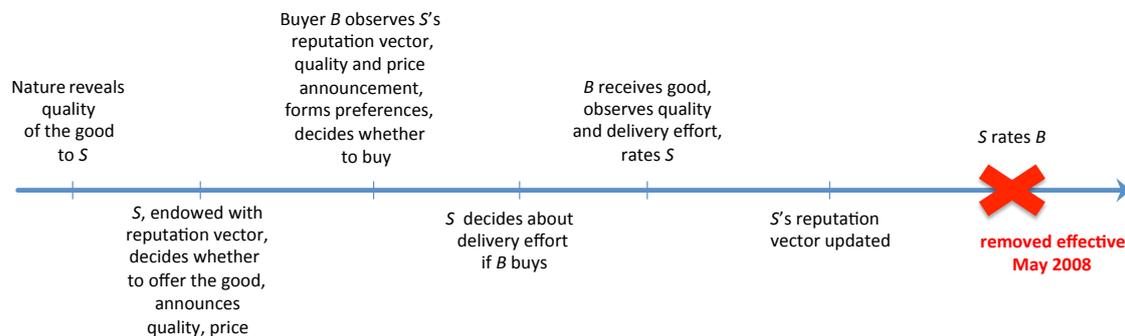
median, as compared to above-median sellers. Moreover, the time trend in the hazard rate after the the change is not statistically different from zero at the 5 percent level. Together with the finding that the increase in the hazard rate in the last three months can be attributed to truncation bias, this suggests that the feedback change did neither trigger immediate, nor induce delayed exit.²⁶

6 A Simple Explanatory Paradigm

In this section, we develop our preferred explanation of these results, and in the ensuing section, we defend it against alternative explanations. Our explanation is summarized in a toy model involving one stage in an infinitely repeated game, with one seller and many buyers, of which one randomly selected buyer arrives in the stage in question. The explanation concentrates on

²⁶Another 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 Figure 14 in Appendix B. In this figure, the decrease by 2.3 percent is given by the percentage difference between the non-linearly extrapolated (to the vertical line) curve to the right and the one to the left.

Figure 4: Sequence of decisions in a typical eBay transaction



the effects of removing the hold-up on the typical buyer's evaluation. Before May 2008, that hold-up had been caused in the classic feedback system by the fact that the seller could retaliate any negative rating by the buyer.

As we will argue that this change could have resulted in a decrease in seller adverse selection and/or moral hazard, we should clarify our understanding of the two concepts within the present context, before developing our toy model. As every so often, the essential information asymmetry relates to the seller type, and his services provided to the uninformed buyer. The seller may be of a conscientious or exploitative type. The conscientious type appropriately describes and prices the good offered by him. In particular, he behaves conscientiously by describing a poor good as poor, and offering it at an appropriately low price. The latter, exploitative seller type describes even a poor good as of high quality and quotes a high price. Furthermore, sellers may differ by their cost of effort spent in the delivery of the good. When buyers cannot identify seller types and behavior *ex ante*, adverse selection may arise via the entry of exploitative sellers into the market, and moral hazard via inefficiently low delivery effort.

Returning to our toy model, the sequence of decisions in a typical eBay transaction is condensed in the time line in Figure 4. We focus on a rating sequence involving the seller's rating of the buyer *after* the buyer's rating of the seller, with the following justification. In Klein, Lambertz, Spagnolo, and Stahl (2006), we found that the seller rated his counterpart before the buyer did so in only 37 percent of all cases in which the rating was mutual;²⁷ and that in this sequence, a positive rating by the seller was followed by a negative buyer rating in less than one percent, indicating that the hold up situation we consider to be at the root of the phenomenon

²⁷See also Bolton, Greiner, and Ockenfels (2013) for a similar finding.

analyzed here is not prevalent in that case.

Assume now that the good to be traded can take on one of two qualities, q_h and q_l with $q_h > q_l$, selected by nature and revealed to sellers at the beginning of the stage game. The good can be described as of high or poor quality, offered at prices $p_h > p_l > 0$ that are positive functions of the seller's reputation capital introduced below, and delivered at some dis-utility, or cost of effort.²⁸ Sellers are differentiated by two types: they may be conscientious, indexed by C , or exploitative, indexed by E .

Sellers also differ by their cost of providing effort towards the delivery of the good. For simplicity, we match effort cost differences into types. When engaging in high effort, seller type j faces effort cost c^j , $j \in \{C, E\}$, with $0 < c^C < c^E$. When engaging in low effort, that effort cost is normalized to zero for both types of sellers. The typical seller is endowed with publicly known reputation capital denoted by k^j , $j \in \{C, E\}$, taking on values on some closed interval on the positive real line. That reputation capital is built from buyer reactions to his behavior in previous transactions. Similarly, the buyer is endowed with publicly known reputation capital k^B . The buyer may, or may not have reputational concerns. In particular, she may derive utility from being rated well, for two reasons: first, when poorly rated, a seller may exclude her from further trades; and second, she may intend to use her ratings as a buyer when selling a good.²⁹

When offering the good (at production cost normalized to zero), the typical seller type j decides whether to announce it at its true quality q_i and an appropriate price $p_i(k^j)$, $i \in \{l, h\}$, $j \in \{C, E\}$, which he always does if the good is of high quality, so $i = h$; or to shirk if $i = l$, by announcing the low quality good as of high quality, q_h , at high price $p_h(k^j)$.

Our typical buyer B , not knowing the true quality of the good, observes the quality-price tuple as announced by the seller, denoted by $[\hat{q}_i, \hat{p}_i]$, $i \in \{l, h\}$, as well as the seller's reputation capital value k^j . On their basis she forms an expected utility $\mathbb{E} \left[u(\hat{q}_i, \hat{p}_i, k^j; k^B) \right]$, $j \in \{C, E\}$. Natural assumptions on this utility are that it increases in the first and the third argument, i.e. the quality as announced by the seller and his reputation; and decreases in the second argument, i.e. the announced price. In the fourth argument, her own reputation, it increases only if she

²⁸Recently, fixed price announcements have become increasingly popular on eBay, as compared to the classic auction format (Einav, Farronato, Levin, and Sundaresan, 2013). In our toy model, we could replace the fixed price announcement by an auction. The winning bid would then reflect the quality of the good as (more or less incorrectly) described by the seller.

²⁹As to the first case, eBay has established clear rules, see <http://pages.ebay.com/help/sell/buyer-requirements.html>.

has reputational concerns. She decides to buy the item if $\mathbb{E} \left[u(\hat{q}_i, \hat{p}_i, k^j; k^B) \right] \geq \tilde{u}$, where \tilde{u} is the utility associated with her exogenously specified outside option.

In case the buyer orders the good, the seller decides whether to spend effort on its delivery. The seller is considered neoclassical, no matter the type: unless punished via a reduction in his reputation capital, he exploits on the anonymity in the market by providing low effort. In that case seller type j 's pay off is $\hat{p}_i(k^j)$, and positive, as long as his reputation capital is sufficiently high. If providing high effort, seller type j 's pay off is $\hat{p}_i(k^j) - c^j$, which is always positive if $i = h$ no matter j , positive if $i = l$ and $j = C$, but tends to be negative if $i = l$ and $j = E$. Hence, if the good is of low quality and announced this way by the exploitative high effort cost seller, his zero profit participation constraint in the stage game assumed here is violated when he intends to provide effort—especially when the reputation capital is low, implying that the seller could quote only a low price to entice the buyer into ordering the good.

Finally, buyer B receives the good, observes the accuracy of the item description and the shipping quality, and rates seller j . This results in a natural upwards, or downwards revision of k^j , that enters next period as the seller's reputation capital.

Before May 2008, the sequence of decisions involving such a transaction was typically concluded by the additional step indicated in Figure 4, in which the seller rated the buyer along the classic scale, resulting in a revision of her reputation capital k^B . By assumption, a negative rating of the buyer by the seller did affect the buyer only if she had reputational concerns. Decisions are supposed to be taken rationally, that is, with backward induction in that simple stage game, that is repeated infinitely often.

Towards results from this toy model, consider first the sequence of decisions before the May 2008 change. The typical seller j can opportunistically condition his rating on the buyer's rating observed by him, by giving a negative mark if the buyer does so. Retaliation by the seller implies that a buyer with reputational concerns is captive to the seller's rating, and thus forced to rate j positively, no matter the seller's decisions taken before—and observed by the buyer after the transaction has taken place. In this case, if nature selects q_l , the exploitative seller shirks with probability 1 on the buyer, by announcing the low quality good at high price p_h , and by not taking any effort to deliver the good—yet still receiving a positive contribution to his reputation capital. Alternatively, the buyer with no reputational concerns rates the seller badly, and this rating is retaliated by the seller, yet without consequence on the buyer's behavior.

The conscientious seller announces the good in the quality selected by nature, and wages effort, no matter the cost. Irrespective of all this, both sellers, if identical except of the type, end up with the same level of reputation capital, i.e. $k^E = k^C$, if confronted with a buyer with reputational concerns. If the buyer has no reputations concerns, however, the exploitative seller ends up with lower reputation capital, so $k^E < k^C$.

Consider now eBay's change in the rating mechanism effective May 2008. Even the buyer with reputational concerns can now give a strategically unbiased negative rating without fearing retaliation. There is abundant evidence that a seller, who intends to stay in the market, must be concerned about his reputation because he can sell more rapidly, and at higher price. The May 2008 change then implies that, in order to obtain a positive mark, such a seller must accurately describe the item even if of low quality, and quote an appropriately low price. He must also take effort in delivering the item. With the assumptions made above, this tends to also imply a positive stage payoff $\hat{p}_l - c^C$ for the conscientious, but a lower, if not negative stage payoff $\hat{p}_l - c^E$ for the exploitative seller type.

In this situation, a seller who consistently behaved poorly, and therefore had accumulated low reputation capital before the May 2008 change, faces two alternatives: either to exit the market—but before then profitably depleting his reputation capital, by shirking, i.e. selling the low quality good at high price and by not providing costly effort towards delivery, resulting in stage payoff $p_h(k^E) > 0$ that eventually converges to zero with the depletion of reputation capital; or alternatively to forgo that short run rent and to continue operating in the market—but then to provide goods in a way that his reputation capital increases even if his current stage pay off is negative, because this allows him to accumulate reputation capital, and with it to sell high quality goods at high price later on.

In all, on the basis of this toy model, the May 2008 change disciplines sellers, and thus results in two main effects: *a reduction in moral hazard exercised by the sellers intending to stay in the market; and/or, a reduction in adverse selection exercised by the exit of poorly rated sellers.* Moral hazard is reduced via an increased delivery effort of the sellers remaining in the market and results in an improved buyer evaluation; adverse selection is improved via an increase in the exit rate of exploitative sellers (especially if poorly rated before the May 2008 change) who, before that, typically deplete their reputation capital. Alternatively, if the poorly rated sellers continue to stay in the market, we should see an above average contribution to the reduction in

moral hazard, towards an improvement in their reputation capital.³⁰

Rather than observing both effects, we only observe a significant improvement in buyer ratings of continuing sellers, which we interpret as a factual improvement of seller behavior; and do not observe any increase in the exit rate of poorly rated sellers—yet a particularly strong increase in their ratings after the May 2008 change.

Why does the poorly performing sellers’ reaction to that change appear to be so asymmetric, against exit, and for improved performance in place? Along the lines of our toy model, the share of sellers with high opportunity cost of adjusting to the new rating regime appears to be small, so *improving on the performance*—as reflected in buyer evaluations—*is still profitable for sellers* in the long run, *even if nature selects a low quality/low price item for them*. Clearly, giving up on shirking, by correctly describing and selling a low quality good at low price, involves the opportunity cost of foregoing a possibly large rent of selling that good at a high price. Yet that rent must be held against the depletion of the reputation record.

To summarize, our preferred explanation of our empirical findings is that market transparency has a positive effect on seller effort provision and no effect on seller exit because the cost to providing that additional effort is lower than the foregone profit from not doing so or of exiting the market.

7 Additional Empirical Support, Competing Explanations, and Robustness

In Section 5, we have shown that removing negative seller ratings of buyers in eBay’s classic feedback system, and with it potential retaliation to negative buyer ratings, has resulted in a significant improvement in DSRs especially for sellers that previously were rated poorly; and in no change in sellers’ exit behavior, especially that of the poorly rated ones. In Section 6, 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. We conduct

³⁰As to detail, if buyers’ feedbacks are delayed, then we predict from this simple paradigm a downward jump in the feedback score right after the May 2008 change, resulting from the fact that before that change, sellers exercised moral hazard in transactions rated negatively by the buyers right after the change, whereas in transactions after the change, sellers would strategically anticipate unbiased buyer rating.

an additional robustness check in Appendix C.

7.1 Evidence on the Assumptions Underlying Our Explanation

The key underlying assumption for 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 the actual time and conscientiousness of the seller when describing, and of the effort waged when delivering the good. These measures would go way beyond what eBay itself knows about the transaction. A close-to-ideal measure would be the report of an independent party that observes all aspects of all transactions of a seller in a given period of time. But that measure is also not available. As a matter of fact, such more direct measures also tend not to exist even for brick-and-mortar stores.

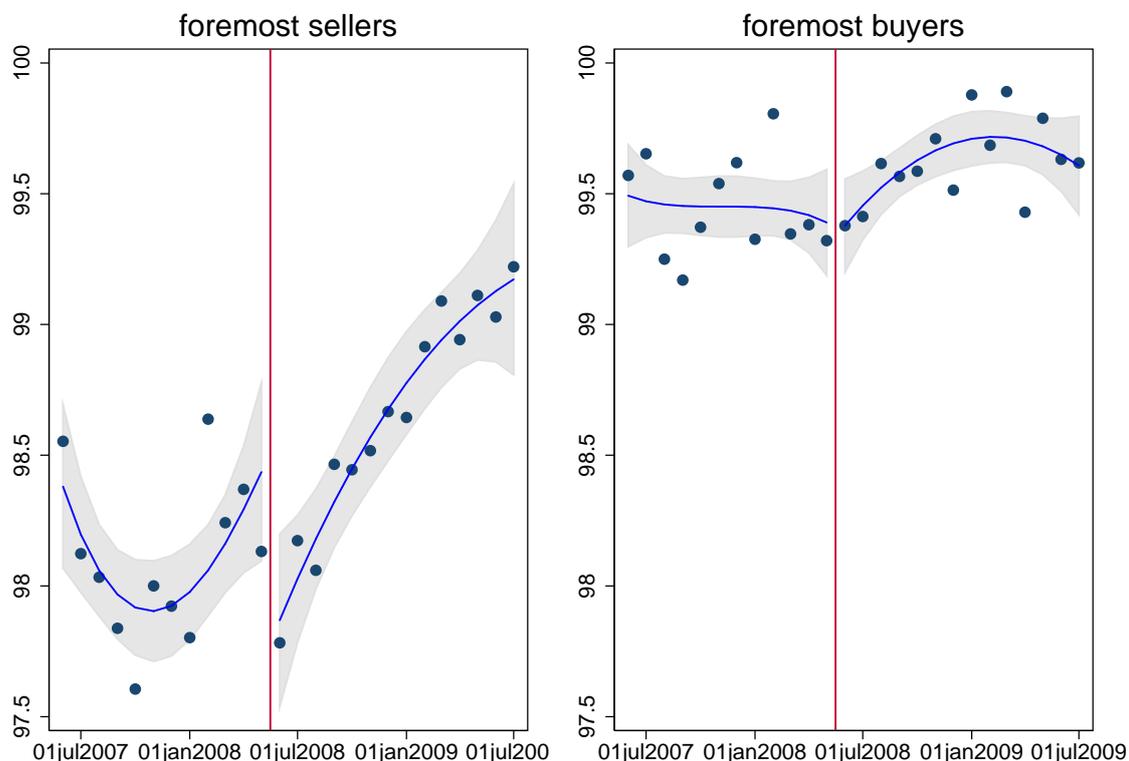
Our measure of seller effort is a reported average of buyers' ratings of seller performance. This report is *not* provided within the classic feedback system whose change we analyze; but in a second system, the DSR system introduced one year before the classic system was changed, in which the buyers'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

A 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 controlled for.³¹ Indirect evidence for this is provided by the fact that the number of DSRs received remains unchanged.³² 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

³¹Formally, a sufficient condition for this to be true is that the propensity that a buyer leaves 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 Mill's ratio term stays constant because the index that changes the probability would remain unchanged.

³²Table 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 6. It shows that the number of ratings only changed for the worst sellers. As we explain below, when we drop this group, then we obtain results that are very similar to the main results reported above. See the explanation in Section 7.2 and Table 7 below.

Figure 5: Effect on classic feedbacks



Notes: The left figure shows the percentage positive feedbacks over time. 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 dots are averages per wave. The solid vertical line depicts the change to the classic feedback mechanism.

ratings stayed the same, as documented in Table 1 and formally tested in Section 7.2. This suggests that the decision whether or not buyers rate was not affected so that changes in DSR ratings for a given seller indeed reflect changes in buyer satisfaction.

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

In Figure 5 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 identi-

Table 4: Effect on classic feedbacks

bandwidth	foremost sellers				foremost buyers			
	50	100	200	300	50	100	200	300
local linear	-.369 (.470)	-.542* (.296)	-.328 (.227)	-.256 (.204)	.052 (.219)	.053 (.177)	-.002 (.122)	.030 (.125)
local quadratic	-.490 (.638)	-.408 (.497)	-.727*** (.296)	-.762*** (.349)	-.085 (.378)	.098 (.230)	-.049 (.187)	-.081 (.158)

Notes: This table shows estimated effects of the feedback change on the percentage positive classic ratings received by users who were either foremost sellers or buyers. These were obtained by performing kernel regressions. We used an Epanetchnikov kernel. The cells contain estimates for local linear and local quadratic regressions and the respective standard errors 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, by 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. Significance at the 10 and 1 percent level is indicated by * and ***, respectively.

fied 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 Footnote 6), and thus exhibited opportunistic behavior right until the May 2008 change. Effective this very date, however, buyers could leave negative classic ratings on this opportunistic seller behavior without the risk of seller retaliation. 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, i.e. before these sellers could react to that change. This, in our view, provides rather convincing evidence that first, buyer ratings quite accurately reflect seller behavior; and second, that buyers have reputational concerns. In view of this, the increasing ratings thereafter suggest that, as time goes on, the sellers do react to unbiased buyer ratings by improving their service.³³

Table 4 contains the corresponding formal tests. The four columns on the left contain results for foremost sellers, and the four columns on the right 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

³³As 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.

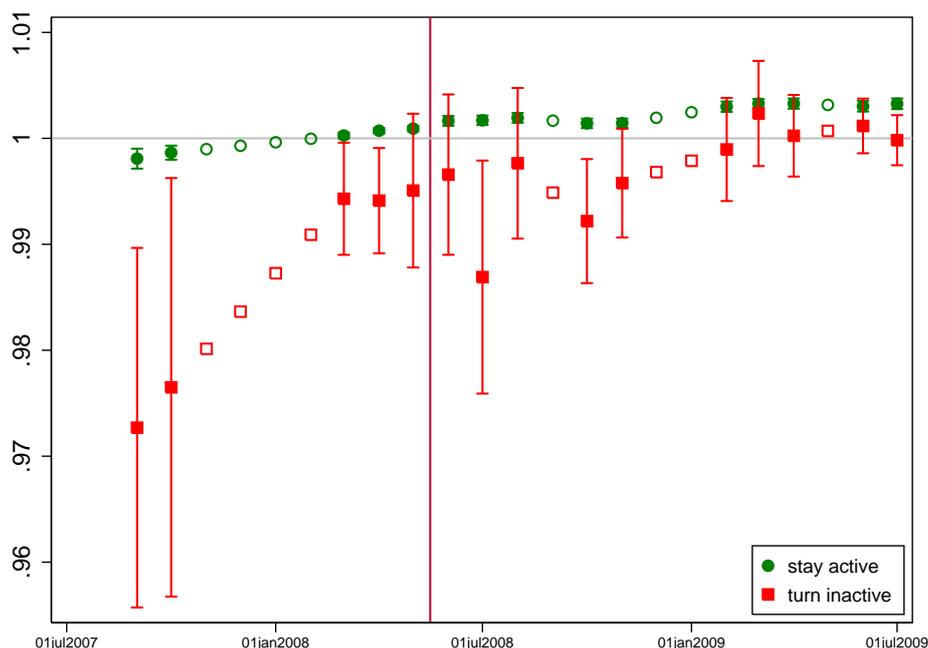
5 suggests that a bandwidth of 200 fits the data well when we use a local quadratic specification. The corresponding estimate for sellers is -.727. It is significant at the one percent level.

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. Towards an answer, 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 $\alpha_i + \left(\sum_{\tau=t-12}^{t-1} w_i^\tau \cdot \varepsilon_{i\tau}\right)$, 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 (x100) 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.³⁴ This is in line with our view that sellers differ by type.

Finally, Cabral and Hortasçsu (2010) argue that in an anonymous market such as the one under discussion, we expect a correlation between exit and ratings because rational sellers change their behavior just before leaving the market and are more likely to leave the market after having received negative feedback because of the lower value to staying in the market. In light of this, we should consider evidence along those lines as an indication supporting our claim that buyers correctly value the transaction via the DSR if we would see buyer ratings on exiting sellers to degenerate. In Figure 6, we compare the continuing and the exiting sellers' DSR scores, relative to those obtained three months earlier. Whereas the DSR scores of the continuing sellers remain essentially unchanged in the time window considered, the exiting sellers' DSR scores go down on average. The error bars seem to suggest that this difference is not significantly different from zero. As before, however, the confidence intervals are point-wise. The ratio is significantly different between exiting and staying sellers 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

³⁴We calculate the ratio in the usual way after performing fixed effects estimation. However, to be precise, this is the ratio between the variance of the fixed effect and the total variance in the reported DSR ratings. 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 to 50 percent of the variance originating from the seller fixed effects.

Figure 6: Behavior prior to inactivity



Notes: In this figure we compare the ratio between the current DSR and the DSR three months before for exiting users (depicted by the squares) to that of the stayers (depicted by the dots). We used linear interpolation in case 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.

error of 0.0001.³⁵ 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 trends downwards before retirement. The standard deviation of the ratio in a given wave, e.g. May 2008, is 0.0077, so the effect is equal to 65 percent of this, which arguably is non-negligible.

7.2 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 neglect possible other causes, such as changes in buyer or seller behavior unrelated to the change of the system addressed by us, other contemporaneous changes in eBay's rules, or changes in the macro-environment. Ideally, one would assess these

³⁵The figure also shows that in the first two months, the difference between those turning inactive and those who don't is much bigger (the following squares are only linear interpolations). This is, however, misleading because, as we have described above, the first two data points are based on much less observations, as DSR scores are only reported if at least 10 DSR ratings were received.

alternative explanations using a ‘control group’ from a market in which 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, however, we go through an extensive list of competing explanations, and conclude that none of those is likely to have caused the increase in DSRs.

First, the results could have been generated simply by grade inflation rather than seller effort. Figure 2 strongly 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, some buyers who wanted to leave a negative rating without retaliation could have done so only by leaving a negative DSR. After the change, they could safely leave a negative classic rating, and therefore abstain from leaving that negative DSR. By this, the DSRs, as reported averages, would just increase because negative DSRs would not be left anymore by those users. One way to test this is to check, for each seller, 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. Towards 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, and sellers ranked poorly a smaller share of the transactions. Table 6 in Appendix B gives evidence to the contrary: the number of DSRs remained unaffected for all but the worst sellers. More importantly, our results would even be 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 are able to also control for seller exit—in some sense a more extreme form of a composition effect—, as discussed in Section 5 above.

An alternative, fourth explanation of our results could be another change in eBay’s allocation mechanism implemented within our observation window. Three months prior to the change whose effects are discussed here, eBay changed the order in which listings were displayed when buyers on eBay searched for an item. Before that, offers were simply ranked by the time remaining until the offer was closed. Under the new ranking scheme, called “Best Match” (BM), eBay introduced a number of factors, by which it determined the sequence of listings. One of these factors was the DSR score. The introduction of BM thus provided an incentive for sellers to improve their performance.³⁶ The ranking scheme was modified several times since. The exact way of ranking listings is a trade secret highly guarded by eBay, as is e.g. Google’s search algorithm. We now assess whether the introduction of BM could have geared our results.

Feiring (2009, 3rd ed, p. 16) reports that within the time window of our analysis, the ranking induced under the BM scheme affected only the very poorest sellers, namely those for whom *Item as Described* and *Communication, Shipping Time, or Shipping and Handling Charges* were ranked only 1 or 2 (out of 5) stars in more than 3 percent, and more than 4 percent of their transactions, respectively. We concentrate our robustness check on these. We will show first, that this is a small group of sellers, and second, that excluding them from our analysis does leave our results essentially unaffected.

As the first order effect of introducing BM, we expect the sellers with relatively poor records to realize fewer transactions, and correspondingly obtain significantly fewer DSRs. So we looked at shifts in the number of DSRs received post March 2008 by percentile of sellers distributed by DSR scores. Table 6 in Appendix B shows that the number of DSRs received after the introduction of BM decreased significantly only for the 10 percent poorest sellers (the effect is -5.76 from a level of 44.44 ratings per month before that, with a standard error of 1.50). Re-doing the regressions that underlie the results in Table 2 and dropping the 10 percent worst sellers yields Table 7, also in in Appendix B. The results are very similar, thus supporting our claim that our analysis is not affected by the introduction of BM.

This leads us to a fifth competing explanation. At the same time at which the BM ranking

³⁶That change was obviously motivated by the increased attractiveness of the fixed price over the auction format to sellers: *A related reason was, we introduced the fixed price format of listings. They could be 30 days, 60 days, and 90 days. And when you have fixed price listings that can be live on the site for 30 days, 60 days, 90 days, “time ending soonest” which was a sort on eBay, no longer made sense for those types of listings. You have a 30-day listing that might only come up to the top of the results 30 days after it was listed. So we had this problem, lots of fixed price inventory, 30 days and 60 days.* (Taken from http://files.meetup.com/1537023/Best_Match_Transcript.doc, last accessed in June 2013.)

scheme was announced, eBay declared that it would provide fee discounts to PowerSellers with favorable DSR ratings (see <http://pages.ebay.com/sell/update08/overview/index.html>, last accessed June 2014). In particular, from July 2008 onwards PowerSellers received a 5 percent final value fee discount if they had received DSRs of 4.6 and above in the last 30 days, and a 15 percent final value fee discount if they had received DSRs of 4.8 or more in the last 30 days. Importantly, these incentives are only provided to a small group of (potential) high-volume PowerSellers. According to Table 1, only 7 percent of the sellers in our sample were PowerSellers on June 1, 2007, and therefore we do not expect this to overturn our main results. To assess this more formally, we excluded all those sellers from the sample who were observed to be a PowerSeller at least once. Then, we re-ran the regression underlying the results in column (1) of Table 2. Based on 40,393 observations for 3,500 sellers we find that the effect of the feedback change is 0.0517, from a baseline level of 4.7079 and with a standard error of 0.0034, which is still highly significant and of very similar magnitude as reported for the whole sample.

A second to last and sixth alternative reason why ratings could have increased could be that buyer demand has shifted from auctions to fixed-price offerings, as documented by [Einav, Farronato, Levin, and Sundaresan \(2013\)](#), in particular in their Figure 1. As one can see there, however, the decrease was gradual at least until September 2008, while our Figure 2 shows that DSRs increased already before that, right after the change to the feedback mechanism.³⁷ Moreover, and more importantly, one would not expect that a change in format should have an effect on buyer satisfaction as measured in our paper. The reason is that the DSR score we used is the average over the four DSR scores in the rating categories item description, communication, shipping charges and shipping speed. Arguably, none of them is related to whether or not the item has been offered in an auction. After all, our interpretation as based on our toy model equally applies to auctions.

Seventh and finally, one might argue that the time period around the change was one of great macroeconomic turmoil, and that this may have had important effects that we attribute to the change of the system. To begin, Figure 2 shows that before the feedback change there was no time trend in the DSR ratings, even though there was macroeconomic turmoil at that time. This already indicates that DSR ratings would not have increased had the classic feedback system not been changed. Towards further evidence against the influence of macroeconomic turmoil we

³⁷See their Section 7 for an explanation why the jump in September 2008 is mechanical.

collected data on average prices of a selection of camera models from the website Pixelpeeper. Figure 15 in Appendix B shows respective time series of the log of the average price on eBay against time. It already suggests that time effects did not play a role. To analyze this more formally, we have regressed log average prices on a full set of time and camera model indicators, as well as on indicators for the total number of models in the market and the number of models of the same brand. Figure 16 shows the fitted time trend and confirms that indeed, there has not been any structural break in the time effects. We complement this evidence with the number of searches for the camera models over time. Figure 17 does not show any structural breaks either. Taken together, we take this as evidence that even though the time period was one of great macroeconomic turmoil, this does not seem to have affected transactions on eBay in drastic ways. Therefore, macroeconomic developments are likely not an alternative explanation for our findings either.

8 Conclusion

In anonymous markets, buyers (and sellers) often rely on reports of each others' performance. In this paper, we use changes in the mechanism by which buyers can report on seller 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 non-anonymous 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 dismissed with 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 the change in focus, 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 non-anonymous, could be opportunistically biased. This gives rise to the research design we exploit: we use the unbiased Detailed Seller Rating 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 toy model that focuses on the effects of this natural experiment. We 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, i.e. 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 this way should results 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.

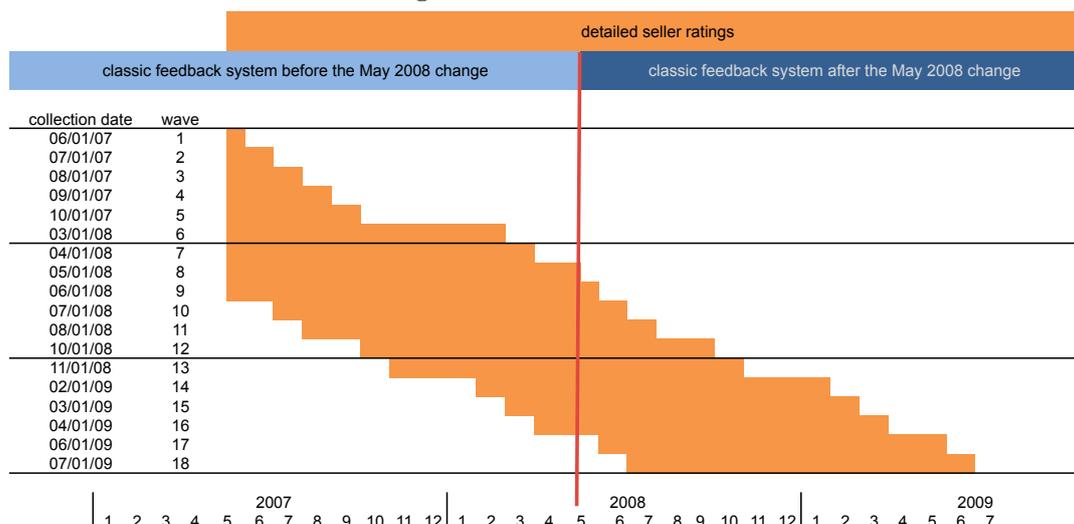
References

- ABBRING, J., P. CHIAPPORI, AND J. PINQUET (2003): "Moral hazard and dynamic insurance data," *Journal of the European Economic Association*, 1(4), 767–820.
- ABBRING, J., J. HECKMAN, P. CHIAPPORI, AND J. PINQUET (2003): "Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?," *Journal of the European Economic Association*, 1(2-3), 512–521.
- AKERLOF, G. A. . (1970): "The market for lemons: Quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, 84, 488–500.
- ANDERSON, M., AND J. MAGRUDER (2012): "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database," *Economic Journal*, 122(563), 957–989.
- ARON-DINE, A., L. EINAV, A. FINKELSTEIN, AND M. CULLEN (2012): "Moral hazard in health

- insurance: How important is forward looking behavior?," NBER Working Paper No. 17802, National Bureau of Economic Research, Cambridge, MA, U.S.A.
- AVERY, C., P. RESNICK, AND R. ZECKHAUSER (1999): "The market for evaluations," *American Economic Review*, 89(3), 564–584.
- BAJARI, P., H. HONG, AND A. KHWAJA (2006): "Moral hazard, adverse selection and health expenditures: A semiparametric analysis," NBER Working Paper 12445, National Bureau of Economic Research, Cambridge, MA, U.S.A.
- BAJARI, P., AND A. HORTASÇSU (2004): "Economic insights from internet auctions," *Journal of Economic Literature*, 42(6), 457–489.
- BOLTON, G., B. GREINER, AND A. OCKENFELS (2013): "Engineering trust: Reciprocity in the production of reputation information," *Management Science*, 59(2), 265–285.
- CABRAL, L. (2012): "Reputation on the internet," in *The Oxford Handbook of the Digital Economy*, ed. by M. Peitz, and J. Waldfogel, chap. 13, pp. 343–354. Oxford University Press, Oxford, U.K.
- CABRAL, L. M., AND A. HORTASÇSU (2010): "The dynamics of seller reputation: Theory and evidence from eBay," *Journal of Industrial Economics*, 58(1), 54–78.
- CHIAPPORI, P. (2000): "Econometric models of insurance under asymmetric information," in *Handbook of Insurance*, ed. by G. Dionne, Huebner International Series on Risk, Insurance, and Economic Security, pp. 365–393. Kluwer, Boston, MA, U.S.A.
- CHIAPPORI, P., AND B. SALANIÉ (2000): "Testing for asymmetric information in insurance markets," *Journal of Political Economy*, 108(1), 56–78.
- DELLAROCAS, C., AND C. WOOD (2008): "The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias," *Management Science*, 54(3), 460–476.
- DRANOVE, D., AND G. Z. JIN (2010): "Quality disclosure and certification: Theory and practice," *Journal of Economic Literature*, 48(4), 935–963.
- EINAV, L., C. FARRONATO, J. LEVIN, AND N. SUNDARESAN (2013): "Sales mechanisms in online markets: What happened to internet auctions?," Working Paper, Stanford University, Stanford, CA, U.S.A.
- EINAV, L., A. FINKELSTEIN, S. RYAN, P. SCHRIMPF, AND M. CULLEN (2011): "Selection on moral hazard in health insurance," NBER Working Paper No. 16969, National Bureau of Economic Research, Cambridge, MA, U.S.A.
- FANG, H., M. KEANE, AND D. SILVERMAN (2006): "Sources of advantageous selection: Evidence from the Medigap insurance market," NBER Working Paper No. 12289, National Bureau of Economic Research, Cambridge, MA, U.S.A.
- FEIRING, D. (2009): "Best match made simple," available at <http://www.auctioninsights.info/bestmatchbook>.
- FINKELSTEIN, A., AND J. POTERBA (2004): "Adverse selection in insurance markets: Policyholder evidence from the UK annuity market," *Journal of Political Economy*, 112(1), 183–208.
- GREENSTONE, M., P. OYER, AND A. VISSING-JORGENSEN (2006): "Mandated disclosure, stock returns, and the 1964 Securities Acts amendments," *Quarterly Journal of Economics*, 121(2), 399–460.

- HECKMAN, J. J. (1978): “Dummy endogenous variables in a simultaneous equation system,” *Econometrica*, 49, 931–959.
- HENZE, B., F. SCHUETT, AND J. P. SLUIJS (forthcoming): “Transparency in markets for experience goods: Experimental evidence,” *Economic Inquiry*, Tilburg University, Tilburg, The Netherlands.
- JIN, G., AND A. KATO (2008): “Price, quality, and reputation: Evidence from an online field experiment,” *The RAND Journal of Economics*, 37(4), 983–1005.
- JIN, G. Z., AND P. LESLIE (2003): “The effect of information on product quality: Evidence from restaurant hygiene grade cards,” *Quarterly Journal of Economics*, 118(2), 409–451.
- KLEIN, T., C. LAMBERTZ, G. SPAGNOLO, AND K. STAHL (2006): “Last minute feedback,” C.E.P.R. Discussion Paper No. 5693, Centre for Economic Policy Research, London, U.K.
- KLEIN, T. J., C. LAMBERTZ, G. SPAGNOLO, AND K. O. STAHL (2009): “The actual structure of eBay’s Feedback mechanism and early evidence on the effect of recent changes,” *International Journal of Electronic Business*, 7(3), 301–320.
- LAFFONT, J., AND D. MARTIMORT (2002): *The theory of incentives*. Princeton University Press.
- LUCKING-REILEY, D., D. BRYAN, N. PRASAD, AND D. REEVES (2007): “Pennies from eBay: The determinants of price in online auctions,” *Journal of Industrial Economics*, 55(2), 223–233.
- MAYZLIN, D., Y. DOVER, AND J. CHEVALIER (forthcoming): “Promotional reviews: An empirical investigation of online review manipulation,” *American Economic Review*.
- MCCRARY, J. (2008): “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142(2), 698–714.
- MELNIK, M., AND J. ALM (2002): “Does a seller’s reputation matter? Evidence from eBay auctions,” *Journal of Industrial Economics*, 50(3), 337–349.
- NOSKO, C., AND S. TADELIS (2014): “Quality externalities and the limits of reputation in two-sided markets,” Mimeographed, University of Chicago, Chicago, IL, U.S.A.
- 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*, ed. by M. Baye, vol. 11, pp. 127–157. Elsevier Science, Amsterdam, The Netherlands.
- RESNICK, P., R. ZECKHAUSER, J. SWANSON, AND K. LOCKWOOD (2006): “The value of reputation on eBay: A controlled experiment,” *Experimental Economics*, 9(2), 79–101.

Figure 7: Data collection



Notes: The shaded area depicts the time interval DSR scores are covering. The change to the classic feedback system was implemented during the month of May 2008.

A Details on Data Collection and Missing Values

As described in Section 4, we drew sellers from five product categories. These are given in Table 5. We then followed these sellers over time, doing our best to keep track of changes of the user names. There are months in which we were not able to collect data on the first day of the month for technical reasons, and therefore have missing values in our data set, as visualized in Figure 2, for instance. It is important to keep in mind, however, that DSR scores are moving averages, and thus contain information on DSR ratings received in the previous 12 months. This is also illustrated in Figure 7. It shows the points in time at which we have collected data, and the corresponding 12 month periods the DSR scores are calculated for. The data collected in the 12 months after one month with missing data will therefore be useful to infer the flow of missing DSR ratings. We do so indirectly when performing regressions, as in Section 5. In order to calculate the corresponding weights, we assume that the ratings that were received between two points in time were spread evenly over the months.

We could directly re-calculate some information for classic ratings. This is because in addition to numbers on how much positive, negative, and neutral feedback a user has received over the last month, eBay also reports these numbers for the last six months, and for the last twelve months. Say we have collected data for some user on the first of January. We then have

numbers for December, but also for July to December, and for January to December. Even without additional observations, we can already infer from those the numbers for January to June, January to November, and July to November.

Combining numbers for overlapping periods from different data collections, we were able to calculate numbers on classical positive, negative, and neutral feedback received for months when we could not collect data. The lower bars in Figure 8 show in which months there was no collection. We could not recover DSR numbers for these months, as eBay does not report DSR statistics in a similar fashion. But for classical feedback, we were able to calculate 12-month-figures for all but one of these months, as reflected in the upper bars of that same figure.³⁸

We use these re-constructed data to analyze how the percentage of positive feedback evolves over time. For this, we do not use the percentage share prominently reported by eBay, but calculate it from the raw numbers for positive, negative, and neutral feedback. This does not only provide us with numbers for months when we could not collect data. It is also necessary to get comparable numbers, whereas eBay repeatedly made changes to how they calculate the percentage share they report; or to what period it refers. In our analyses, it is consistently calculated as positive divided by the total number (positive plus negative plus neutral) of ratings received in the same period.

Figure 8 also shows how many DSRs were received in the preceding year and over time. It increases until May 2008 because DSRs were only introduced in May 2007. Since then, the number of DSRs received in the previous 12 months is stable over time. Moving forward one more month means that the number of DSR received thirteen months in the past does now not enter anymore and is replaced by the number of DSRs received in the previous month. The observation that the number of DSRs received in the previous 12 months does not change over time therefore means that the number of DSRs received in every single month is stable over time. The figure also shows that on average (across users), more feedbacks are received than DSRs. The reason for this is that DSRs can only be received when acting as a seller and when a classic rating is left (but not *vice versa*). Moreover, for this figure we have counted DSRs as zero when they were not displayed and DSRs are only displayed if at least 10 DSRs were left in the previous 12 months.

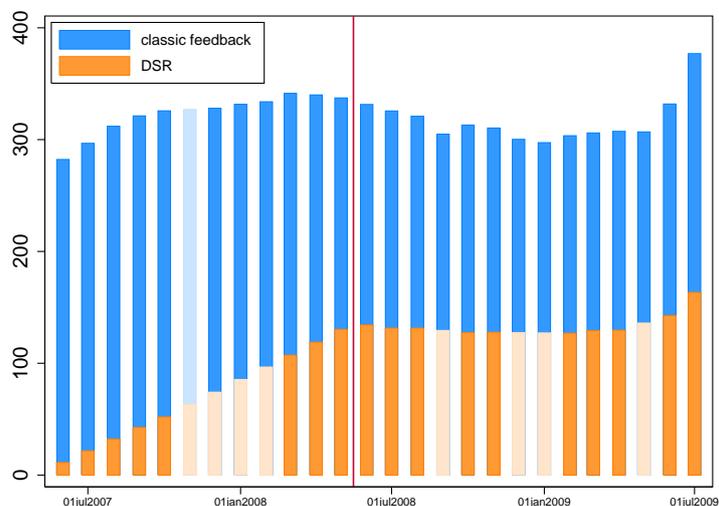
³⁸We were able to re-construct the number of classic ratings received in each of the months in the figure, but not for all months prior to the sample period. Therefore, there is one month of missing data in the figure.

Table 5: Product categories

Home → All Categories → Computers & Networking → Laptops, Notebooks
Home → All Categories → Consumer Electronics → Apple iPod, MP3 Players
Home → All Categories → Toys & Hobbies → Model RR, Trains
Home → All Categories → Collectibles → Trading Cards
Home → All Categories → Home & Garden → Food & Wine

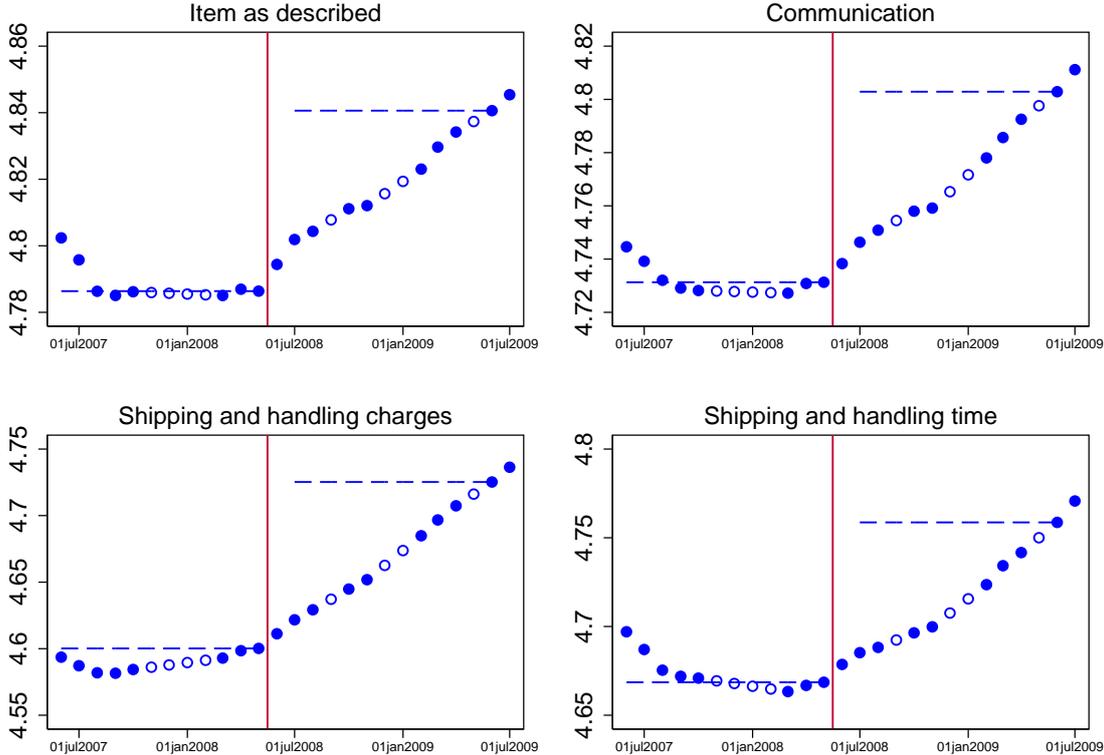
Notes: As of February 2008.

Figure 8: Number of classic feedback and DSRs received in the previous 12 months



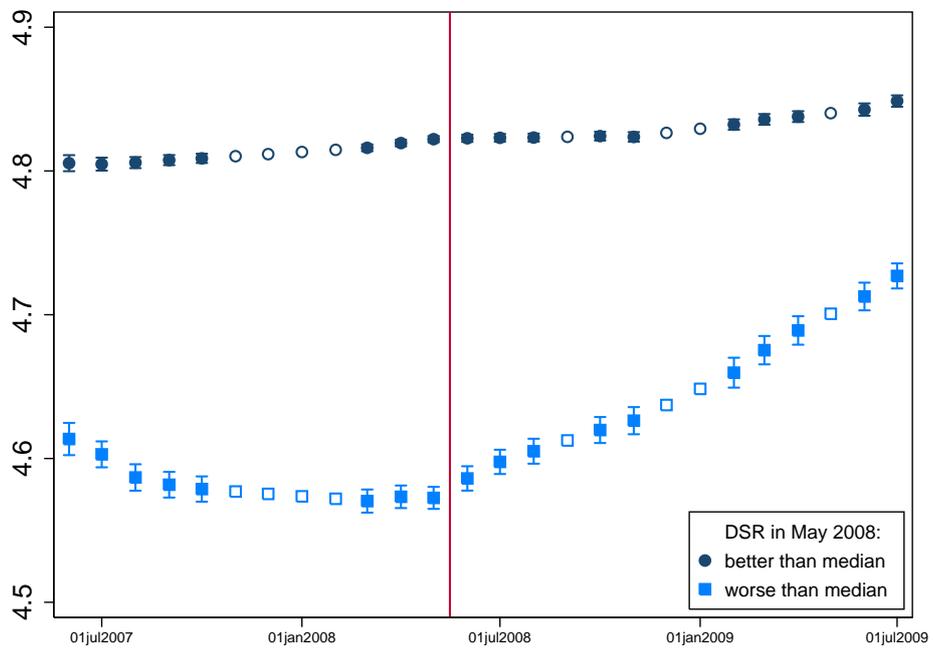
B Additional Tables and Figures

Figure 9: Average DSR score by category



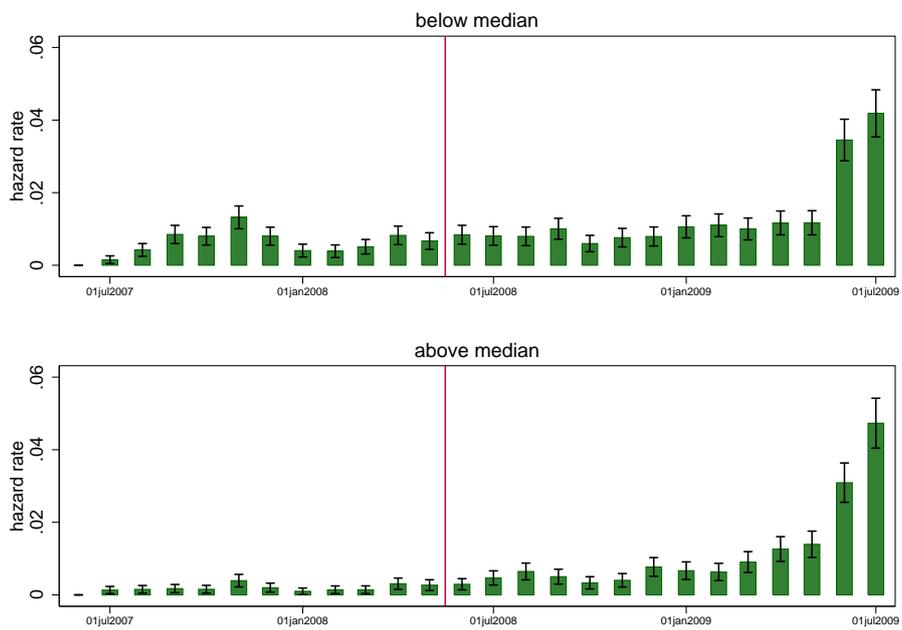
Notes: Figure shows how the average of the four DSR rating categories changed over time.

Figure 10: Evolution for two different groups



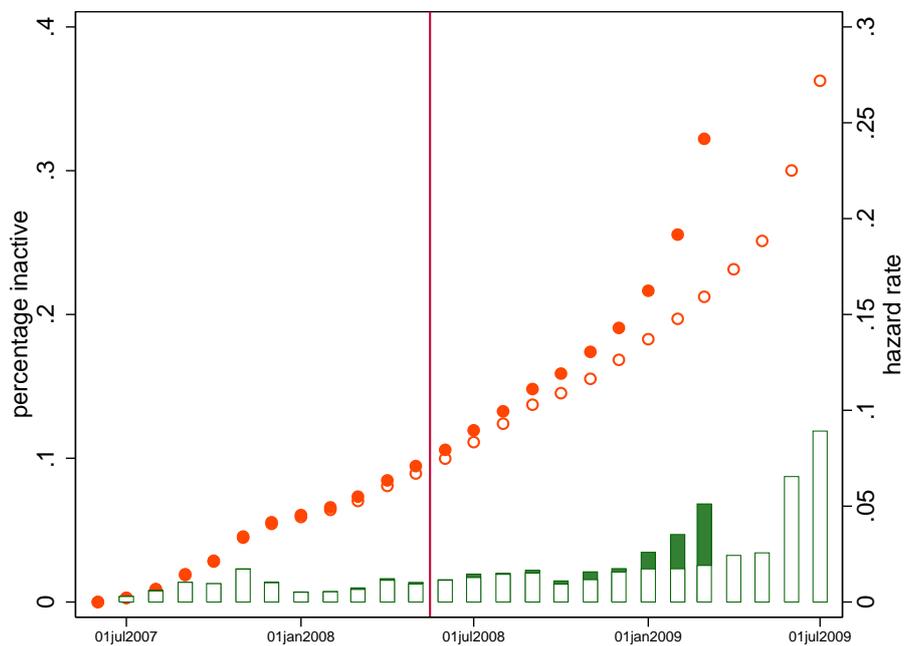
Notes: Figure shows how average DSR score changed over time, with sellers split into those with DSR score above the median of 4.75 prior to the May 2008 change, and those with a score below that. See notes to Figure 2.

Figure 11: Exit for two different groups



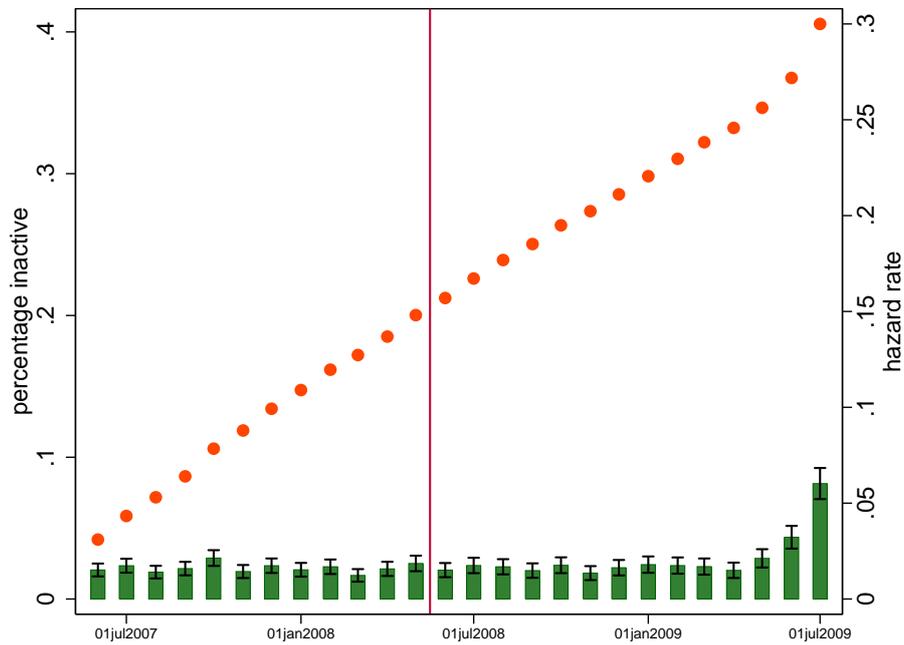
Notes: See notes to Figure 3. Sellers are split into those who had a DSR score above the median of 4.75 prior to the May 2008 change, and those who had a score below that. We used an extrapolated value for May 2008 if the DSR score was only available at a later point in time.

Figure 12: Truncation bias



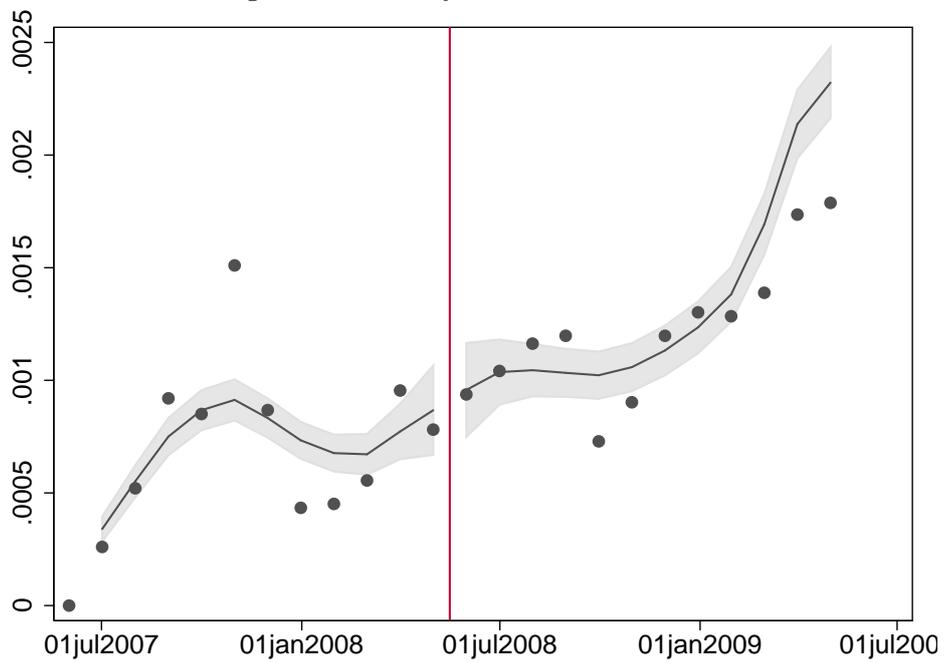
Notes: Figure shows that truncation bias, as discussed in Footnote 9, arises in the last three months. Non-filled dots and bars in this figure correspond to the filled ones in Figure 3. Filled dots are for the case in which we drop the last four waves of data and define inactivity as not observing any 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 3 are not affected until April 2009.

Figure 13: Simulated data



Notes: Figure shows the result of a Monte Carlo simulation based on the data used for Figure 3. We reconstruct the classic feedbacks given for 9 periods prior to the start of our data collection (See Appendix A 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 to exit is 1.5 percent. This generates an increase in the hazard rate in the last three months that solely arises because we misclassify as inactive users that are not active in every month.

Figure 14: Density of the time of exit



Notes: 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 one month time interval and correspond to bins in a histogram. See [McCrary \(2008\)](#) for details.

Table 6: Effect of Best Match and feedback change by decile of DSR rating

	decile of DSR score on March 1, 2008									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
number sellers	459	443	450	478	423	451	554	347	451	450
number observations	5,115	6,091	6,177	6,888	6,234	6,700	8,180	5,122	6,554	6,058
<i>effects on the number of DSRs received</i>										
average number DSR before March 2008	44.44*** (3.48)	69.00*** (3.16)	54.17*** (1.50)	41.82*** (3.93)	47.64*** (1.41)	31.97*** (1.15)	34.76*** (1.00)	28.71*** (0.94)	31.53*** (0.85)	18.29*** (0.71)
effect of Best Match	-5.76*** (1.50)	-0.09 (1.33)	-0.72 (0.67)	-2.02* (1.22)	-0.24 (0.48)	0.10 (0.43)	0.26 (0.45)	0.02 (0.38)	0.10 (0.36)	0.18 (0.30)
effect of feedback change	3.09** (1.26)	-2.13 (1.39)	0.14 (0.74)	0.50 (0.70)	0.15 (0.46)	-0.17 (0.36)	-0.56 (0.38)	-0.11 (0.35)	-0.20 (0.33)	-0.15 (0.24)
fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.04	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
<i>effects on the DSR ratings received</i>										
average DSR received before March 2008	4.32*** (0.01)	4.55*** (0.01)	4.63*** (0.00)	4.68*** (0.00)	4.72*** (0.00)	4.75*** (0.00)	4.78*** (0.00)	4.81*** (0.00)	4.84*** (0.00)	4.89*** (0.00)
effect of Best Match	-0.03 (0.05)	-0.03 (0.04)	0.00 (0.02)	0.00 (0.02)	0.04*** (0.01)	0.02 (0.01)	0.00 (0.01)	0.00 (0.01)	-0.01 (0.01)	-0.01 (0.01)
effect of feedback change	0.20*** (0.05)	0.14*** (0.04)	0.09*** (0.02)	0.07*** (0.02)	0.03** (0.01)	0.04*** (0.01)	0.05*** (0.01)	0.03** (0.01)	0.02* (0.01)	-0.01 (0.01)
fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.08	0.08	0.13	0.11	0.19	0.15	0.11	0.06	0.01	0.02

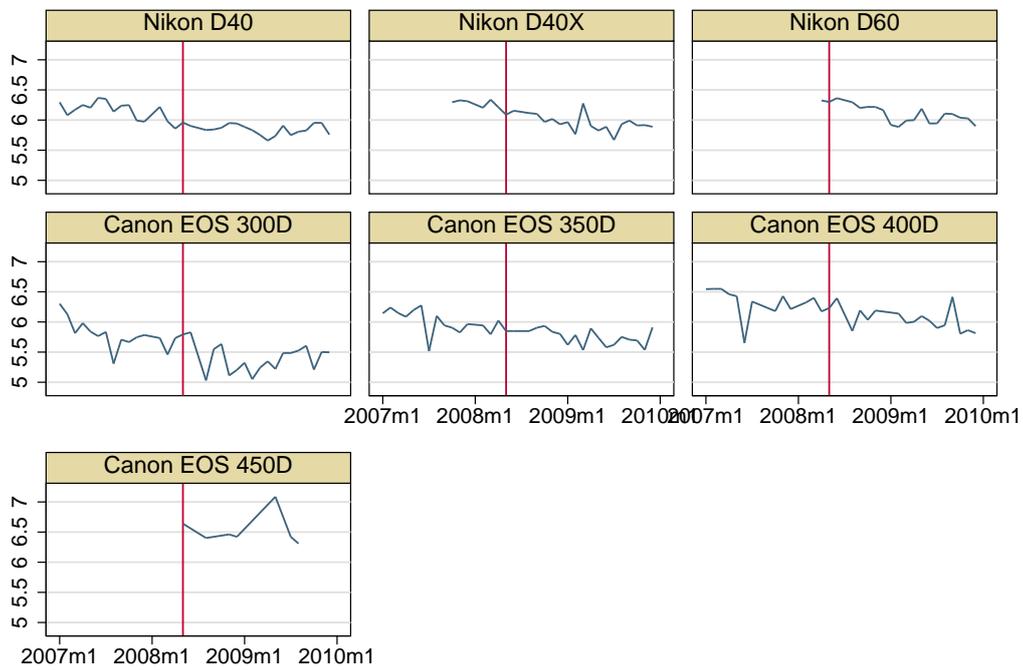
Notes: Upper part of this table shows results of regressions of the number of DSR received per month on the fraction of months in which BM and the new feedback system were in place, respectively. Lower part shows regressions of the average DSR score, averaged over the four categories, on a constant term and the fraction of feedbacks received after the introduction of BM and the May 2008 change to the feedback mechanism, respectively. We assumed that 75 percent of the DSR in March, 2008 were received after the introduction of BM and 50 percent of the DSR received in May 2008 were received after the change to the classic feedback mechanism. One observation is a seller-wave combination. There are 2,337 sellers. Throughout, we control for fixed effects. The R^2 is the within- R^2 . Standard errors are cluster-robust at the seller level and significance at the 10, 5 and 1 percent level is indicated by *, ** and ***, respectively.

Table 7: Effect of the May 2008 change without 10 percent worst sellers

	(1) full sample	(2) small window	(3) time trend	(4) DSR < 4.75	(5) DSR ≥ 4.75
average DSR before change	4.7400*** (0.0006)	4.7426*** (0.0004)	4.7545*** (0.0030)	4.6465*** (0.0010)	4.8129*** (0.0006)
effect of feedback change	0.0535*** (0.0022)	0.0382*** (0.0038)		0.0844*** (0.0041)	0.0306*** (0.0021)
effect of feedback change until September 2008			0.0016 (0.0074)		
effect of feedback change after September 2008			0.0682*** (0.0174)		
linear time trend before change			0.0016*** (0.0004)		
linear time trend after change			-0.0013 (0.0018)		
fixed effects	yes	yes	yes	yes	yes
R^2	0.0703	0.0165	0.0748	0.1067	0.0437
number sellers	4,047	4,047	4,047	1,794	2,253
number observations	58,004	26,358	58,004	25,390	32,614

Notes: See notes to Table 2. Here, we additionally exclude the 10 percent worst sellers, as measured by their DSR score on March 1, 2008.

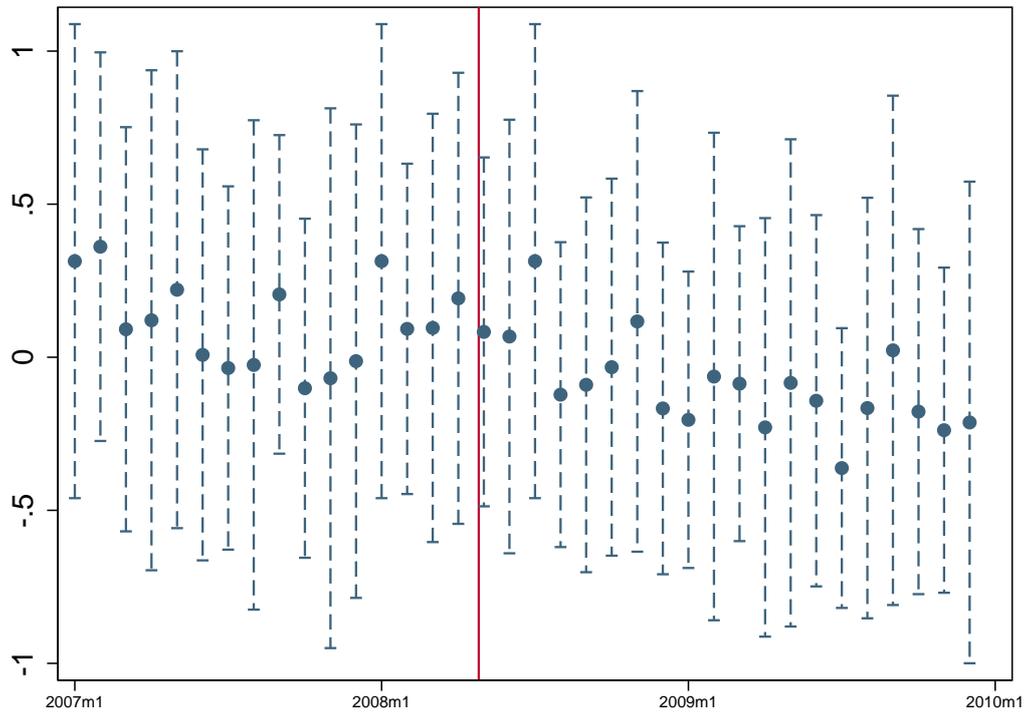
Figure 15: Average auction prices for selected camera models



Graphs by model

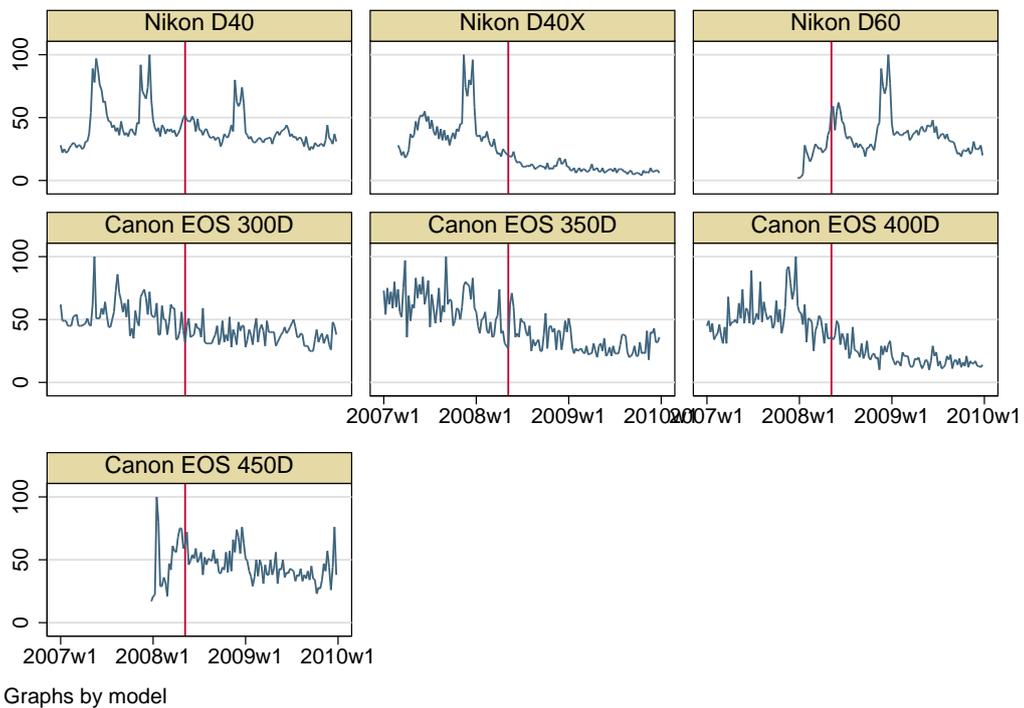
Notes: Figure shows the log of average auction prices for entry-level digital single-lens reflex (DSLR) camera models from Nikon and Canon in North America. Collected from <http://www.pixel-peeper.com/>. Vertical line indicates the change to the feedback system in May 2008.

Figure 16: Time trend in camera prices



Notes: Figure constructed from the data shown in Figure 15 and shows the time fixed effects controlled for camera dummies, time-on-the-market dummies for each model, number of models of the same brand on the market, and total number of models on the market. Bars indicate pointwise 95 percent confidence intervals, based on the reported cluster-robust standard errors, clustered at the level of the camera models. Vertical line indicates the change to the feedback system in May 2008.

Figure 17: Google searches for camera models



Notes: Figure shows time profiles of the number of Google searches for different camera models in North America, available on a weekly basis, downloaded from the Google Trends website. Search terms used are "nikon d40," "nikon d40x," "nikon d60," "canon 300d," "canon 350d," "canon 400d," and "canon 450d." Number of searches rescaled by Google Trends so that highest number is 100 (percent) for each model. Vertical line indicates the change to the feedback system in May 2008.

C Delay between Transaction and Rating

We have performed one additional, more technical robustness check. It is related to the fact that we don't know the exact date of when eBay enacted the change whose effects we are reporting here. To see that this may be a problem in principle, suppose for now that both sellers and buyers would know the exact date of the change (around May 15, 2008), and the seller changed his behavior in a transaction right after the change. That transaction was probably completed by May 25 and this was also the time at which the buyer left a DSR for him. Conversely, if a transaction took place before May 15, 2008, then the seller was not able to react to the feedback change inasmuch unanticipated. Nevertheless, a feedback could have been left for that transaction in the second half of May 2008. In our analysis, we have assumed that half of the DSRs received in May 2008 corresponded to transactions conducted after the feedback change. Disregarding this reporting delay, we attribute the ratings after the change all to transactions thereafter, and with this tend to underestimate the effect of the feedback change. We do not expect this to have big effects, however, because the delay is likely to be small relative to the length of our observation period.

To assess whether this is indeed the case, we check the robustness of our results to changes in the delay we implicitly assume. We don't have a record on delays between announcement, transaction, and feedback. Yet Figure 2 in [Klein, Lambertz, Spagnolo, and Stahl \(2006\)](#) shows the distribution of the time between the end of the auction and the moment at which the first feedback was left. The vast majority of feedbacks is positive and for those about 60 percent are left after 2 weeks, and almost 90 percent after 4 weeks.³⁹ Based on this, we re-did the analysis of Section 7.2 relating to BM and the feedback change, assuming that out of all DSRs received in March 2008, 75 percent of the transactions took place after the introduction of BM. Moreover, we assumed that out of all DSRs received in May 2008, 25 percent of the transactions took place after the change to the feedback system. The results were very similar. We also re-did the analysis underlying Table 2, assuming that 25 percent of the transactions took place after the change to the feedback system. Table 8 shows the results. They are very similar.

³⁹For negative feedbacks, the distribution is shifted to the right. [Klein, Lambertz, Spagnolo, and Stahl \(2006\)](#) argue that this may be due to strategic considerations: both parties had an incentive to wait with their first rating if it was negative, because then it was less likely to be retaliated. After May 2008, these strategic considerations were not important anymore because sellers could not retaliate negative feedbacks anymore.

Table 8: Effects with time delay

	(1) full sample	(2) small window	(3) time trend	(4) DSR < 4.75	(5) DSR ≥ 4.75
average DSR before change	4.7067*** (0.0006)	4.7034*** (0.0004)	4.7150*** (0.0035)	4.5919*** (0.0010)	4.8143*** (0.0006)
effect of feedback change	0.0589*** (0.0024)	0.0435*** (0.0052)		0.0921*** (0.0044)	0.0318*** (0.0021)
effect of feedback change until September 2008			0.0183** (0.0081)		
effect of feedback change after September 2008			0.0652*** (0.0180)		
linear time trend before change			0.0009** (0.0004)		
linear time trend after change			0.0001 (0.0019)		
fixed effects	yes	yes	yes	yes	yes
R^2	0.0583	0.0125	0.0606	0.0820	0.0459
number sellers	5,224	4,919	5,224	2,337	2,337
number observations	67,376	30,488	67,376	31,260	33,508

Notes: See notes to Table 2. The difference between the two tables is that here, we assume that only 25 percent of the feedbacks received in May 2008 correspond to transactions that took place after the change.