Social Science Research 95 (2021) 102522

Contents lists available at ScienceDirect

Social Science Research

journal homepage: http://www.elsevier.com/locate/ssresearch

Reputation effects in peer-to-peer online markets: A meta-analysis*

Ruohuang Jiao, Wojtek Przepiorka *, Vincent Buskens

Utrecht University, Department of Sociology / ICS, Padualaan 14, 3584, CH, Utrecht, Netherlands

ARTICLE INFO

Keywords:
Online market
Trust
Reputation
Reputation system
Reputation effect
Meta-analysis

ABSTRACT

Most online market exchanges are governed by reputation systems, which allow traders to comment on one another's behavior and attributes with ratings and text messages. These ratings then constitute sellers' reputations that serve as signals of their trustworthiness and competence. The large body of research investigating the effect of reputation on selling performance has produced mixed results, and there is a lack of consensus on whether the reputation effect exists and what it means. After showing how the reputation effect can be derived from a game-theoretic model, we use meta-analysis to synthesize evidence from 107 studies investigating the reputation effect in peer-to-peer online markets. Our results corroborate the existence of the reputation effect across different operationalizations of seller reputation and selling performance. Our results also show the extent to which the reputation effect varies. We discuss potential explanations for the variation in reputation effects that cannot be attributed to sampling error and thereby point out promising avenues for future research.

1. Introduction

Reputation as a mechanism to govern market exchanges is undergoing its most successful propagation. Although humans' ability to share information about others' deeds and misdeeds has promoted market exchange throughout history (Greif 1989; Hillmann 2013; Diekmann and Przepiorka 2019) modern information and communication technology (ICT) has reduced the costs of sharing information to a minimum (Rifkin 2014). In recent years, online markets have increased in popularity and fundamentally transformed the ways in which we engage in economic exchange. There are online markets for consumer goods, loans, plumbing work, academic positions, illegal drugs, etc. If one conceives of markets as institutions that facilitate exchange (Coase 1988), it becomes apparent how online markets have made generalized social exchange possible at a large scale (Blau 1964). The number of online platforms for dating, car-sharing, time-sharing, house-swapping, etc. is steadily increasing (Botsman and Rogers 2010).

We would like to thank Andreas Diekmann, Sanne Kellij, Irene Klugkist, Andreas Schneck, Jeroen Weesie and Rafael Wittek for their insightful comments and suggestions. R.J. gratefully acknowledges financial support from the China Scholarship Council [grant no. 201707720047].

* Corresponding author.
E-mail address: w.przepiorka@uu.nl (W. Przepiorka).

https://doi.org/10.1016/j.ssresearch.2020.102522
Received 15 May 2020; Received in revised form 11 November 2020; Accepted 23 December 2020
Available online 2 January 2021

Most of these market exchanges are governed by online reputation systems (Dellarocas 2003), which allow buyers to comment on sellers' behaviors and attributes with ratings and text messages.[1] These ratings then constitute these sellers' reputations, which can be conceived of as signals of their trustworthiness and competences (Li et al., 2019; Przepiorka and Berger 2017). Indeed, research shows that information about seller reputations can be predictive of online exchange fraud and disputes (Gregg and Scott 2006; Macinnes et al., 2005). However, trustworthy and untrustworthy sellers are indistinguishable when they enter a market because they have no records of past behavior. One way for honest market entrants to build their reputation is to lower prices or offer other types of discounts to attract interaction partners and prove their trustworthiness and competence. Building a good reputation is therefore costly. However, honest agents will be compensated for their investment if they remain in the market long enough, whereas dishonest agents will not bother to invest in building a good reputation. Hence, traders can infer potential trading partners' honest intentions from their good reputations (Przepiorka 2013; Shapiro 1983). A similar argument can be made with regard to the quality of commodities and services offered via online markets.

This argument implies that online sellers' reputations and their business success will be correlated, that is, sellers with a better online reputation will realize more sales at higher prices. We henceforth call this the *reputation effect*.[2] However, there is a lack of consensus regarding the existence and meaning of the reputation effect (see, e.g., Lindenberg et al., 2020), especially because in many previous studies that estimated it, the reputation effect appeared to be small (Livingston 2005; Snijders and Matzat 2019; Standifird 2001). Given the persistently increasing popularity of online market platforms that are governed by reputation systems, it is important to understand in what ways the reputation effect can be meaningful.

The aim of this paper is first and foremost to establish the general existence and the variation of the reputation effect reported in the literature. We do this by synthesizing evidence from 107 studies that investigated the reputation effect. Our meta-analysis includes 378 coefficients estimated based on 181 different datasets comprising a total of 14.04 million observations of online market transactions. More precisely, we conduct twelve separate meta-analyses, one for each combination of three types of seller reputation variables (number of positive ratings, number of negative ratings, overall reputation scores) and four types of selling performance variables (probability of sale, selling price, selling quantity, ratio of selling price to reference price). By splitting the data into these twelve subsets, we establish the robustness of the reputation effect across different operationalizations of seller reputation and selling performance. We leave elaborations and tests of explanations for the variation in the size of the reputation effect within these subsets for subsequent studies.

Before we describe how we conduct our meta-analyses and present our results, we recap the game-theoretic underpinnings of the reputation effect in the next section. In the concluding section, we highlight possible reasons for the variation of the reputation effect that cannot be attributed to sampling error and suggest directions for future research. In particular, we list moderating factors that could be used in subsequent sub-group analyses and meta-regressions to identify potential determinants of the size of the reputation effect.

2. Theory

An interaction between a buyer and a seller in an online market is usually conceptualized as a trust game with incomplete information (Güth and Ockenfels, 2003). In the standard trust game (Dasgupta 1988, see the right sub-tree denoted TG in Fig. 1), a first moving agent (the buyer) decides whether to trust the second moving agent (the seller) and send money to buy an item. Upon receipt of the money, the seller decides whether to ship the item the buyer paid for. In the TG, the payoffs are ordered such that the seller would not ship ($T > R$) and, therefore, the buyer does not buy ($P > S$). As a result, the exchange does not take place and both earn payoff P, which is lower than the gains from trade R.

Unlike the standard trust game, the trust game with incomplete information (TGI) accounts for the fact that the buyer is uncertain about the seller's incentives and ability to be trustworthy because the seller holds private information about their preferences and constraints (Raub 2004). In the TGI (Fig. 1), this important aspect of a trust problem is modelled with the buyer being in one of two games, the assurance game (AG) or the TG. The two games merely differ in the order of seller payoffs. In the AG, the seller has an additional benefit b and/or cost c for shipping or not shipping, respectively, such that $R + b > T - c$. This order of payoffs implies that, in the AG, the seller has an incentive to ship if the buyer buys. The uncertainty of the buyer is modelled by a so-called move of nature (N) determining which game the buyer is in. While the seller knows whether they are in the AG or the TG, the buyer only knows the probability α of being in the AG. This is denoted by the dashed line connecting the two decision nodes of the buyer. Given the limited information the buyer has about the seller, how does the buyer decide whether to choose 'buy' or 'not buy' in the TGI?

Knowing α and the TGI payoffs, the buyer calculates the expected payoff from either of their two actions and takes the one that

[1] Although reputation systems and recommender systems are related in that they are employed by online market platforms and fueled by user generated data, they must be kept distinct. A reputation system can be defined as socio-technical structure through which third-party information about an actor's past behavior is collected, transmitted and aggregated (see also Resnick et al., 2000). A recommender system can be defined as a software tool or technique that suggests items that are of potential interest to a particular actor (see, e.g., Ricci et al., 2015; Palopoli et al., 2013, 2016).

[2] Several studies have corroborated that the relation between seller reputation and business success is causal (Przepiorka 2013; Resnick et al., 2006; Snijders and Weesie 2009). Snijders and Weesie have established the causal relation between seller reputation and success by estimating buyers' willingness to pay for seller reputation. Note that we refer to the reputation effect as the premium sellers can expect for their reputation rather than buyers' willingness to pay for reputation. All studies included in our meta-analysis conform with our definition of the term.

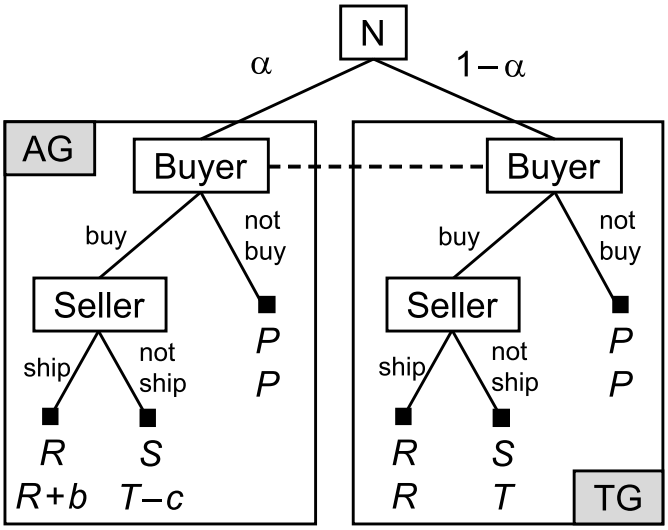

Fig. 1. Trust game with incomplete information (TGI).

maximizes their expected payoffs. If the buyer buys, their expected payoff is $\alpha R + (1-\alpha)S$. If the buyer does not buy, their payoff is P in either case. The buyer chooses 'buy' if doing so gets them a larger expected payoff than choosing 'not buy', that is, if $\alpha R + (1-\alpha)S > P$. Since the payoffs of the TGI are fixed, we solve this equation for α and obtain

$$\alpha = \frac{P-S}{R-S} \tag{1}$$

Hence, if α, the probability of being in the AG, is larger than a certain threshold value, the buyer buys and abstains from buying otherwise. Note that the α-threshold is determined entirely by the buyer's payoffs.

Theoretically, information about a seller's reputation tells buyers something about the likelihood of being in the AG or the TG. Without additional information about a particular seller, α might correspond to buyers' prior beliefs about the trustworthiness of online sellers in general. How does information about seller reputation affect a buyer's belief about a seller's trustworthiness?

To answer this question, we first have to expand on the reasons for why some sellers are trustworthy and others are less so and thus what b and c in the AG comprise. Of course, many sellers are just honest; they never think about cheating buyers. These sellers might have social preferences either of the type that increase their utility if the buyer's utility is increased ($+b$) or of the type that decreases their utility if the buyer's utility is decreased ($-c$) or both (Becker 1976; Fehr and Schmidt 1999). However, let's not be too quick with invoking social preferences to explain why some sellers are more trustworthy than others.

Most sellers are interested in making money from doing business online. These sellers thus have an interest in staying in the market and expanding their business. Although fraudulent sellers also have an interest in staying in the market, a reputation system would not allow them to stay if they behave untrustworthily. The reason is that once a seller who received a buyer's money and did not send back the item the buyer paid for is rated negatively by the buyer, they will be perceived as dishonest and no other buyer will buy from them in the future. In other words, a onetime failure to reciprocate a buyer's trust will result in the seller having to leave the market and possibly re-enter the market using a new pseudonym (Friedman and Resnick 2001). However, re-entering the market under a new pseudonym implies that the seller has no previous records of past transactions (neither good or bad). Hence, sellers who value the stream of future gains from trading with buyers higher than making a gain from cheating a buyer once and having to start from scratch will enter the market and build up their good reputation through cooperative and honest business conduct (see also Buskens and Raub 2013). But how do these sellers build their reputation given that without one they are indistinguishable from dishonest market entrants?

A good reputation is a reliable signal of trustworthiness because it is costly to produce and separates sellers who can incur the costs to produce it from those who cannot (Przepiorka and Berger 2017). Building a reputation is costly not only because sellers have to behave persistently cooperatively but also because market entrants, i.e. sellers without a reputation, have to accept lower prices for their offers. New sellers have to offer their items at prices that make buyers indifferent between their offer and an offer by a seller with a good reputation.

In terms of our TGI, for buyers to be indifferent between a seller with a good and a seller with an unknown reputation, buyers' payoffs from buying from either of them must be equal. Since a seller with a good reputation is in the AG, a buyer's payoff from buying from that seller is R. Recall that a buyer's expected payoff from buying from a seller with an unknown reputation is $\alpha R + (1-\alpha)S$. Hence, sellers with an unknown reputation must offer a discount d, such that $\alpha R + (1-\alpha)S + d = R$. When solving this equation for d, we obtain

$$d = (1-\alpha)(R-S) \tag{2}$$

In other words, *d* corresponds to a buyer's expected net loss from buying from a seller with no reputation. However, once honest sellers without a reputation have received positive ratings (because of the great buyer experience), they do not need to offer this discount anymore and can increase their prices to a level that compensates them for their initial investment in building a reputation (Przepiorka 2013; Shapiro 1983).

From this argument it follows that sellers' reputations and their selling performance in terms of prices will be positively correlated. Moreover, in case supply exceeds demand and the market does not clear, by the same argument we can expect sellers' reputations and their selling performance in terms of probability of sale to be positively correlated:

H1. The better a seller's reputation, the higher is the price the seller can obtain for their items.

H2. The better a seller's reputation, the higher will be the probability the seller's items will be sold.

Note that H1 and H2 imply that sellers with a better reputation will obtain higher prices compared to a reference price and will sell larger quantities of their items in a given time frame, respectively:

H3. The better a seller's reputation, the higher is the price the seller can obtain for their items compared to a reference price.

H4. The better a seller's reputation, the larger will be the quantities at which the seller's items will be sold.

The previous literature included in our meta-analysis has tested these hypotheses more or less explicitly using different operationalizations of seller reputation. In our meta-analysis, we assess whether these four hypotheses are generally supported, considering each with three different operationalizations of seller reputation: number of positive ratings, number of negative ratings and reputation score (i.e. the number of positive ratings minus the number of negative ratings).

3. Methods

In this section we describe in detail how we conducted our meta-analysis. We first describe our literature search and the criteria for the inclusion of previous empirical studies in our analysis. Next, we describe the model selection process that we employed when studies reported more than one model estimating reputation effects based on the same data. Finally, we describe our approach to making effect sizes comparable so they could be included in our meta-analyses (see Tong and Guo 2019).

3.1. Literature search

We conduct a meta-analysis on the relation between selling performance and seller reputation with results from existing empirical studies. The process of collecting all relevant articles on reputation effects in peer-to-peer online markets starts with two previous

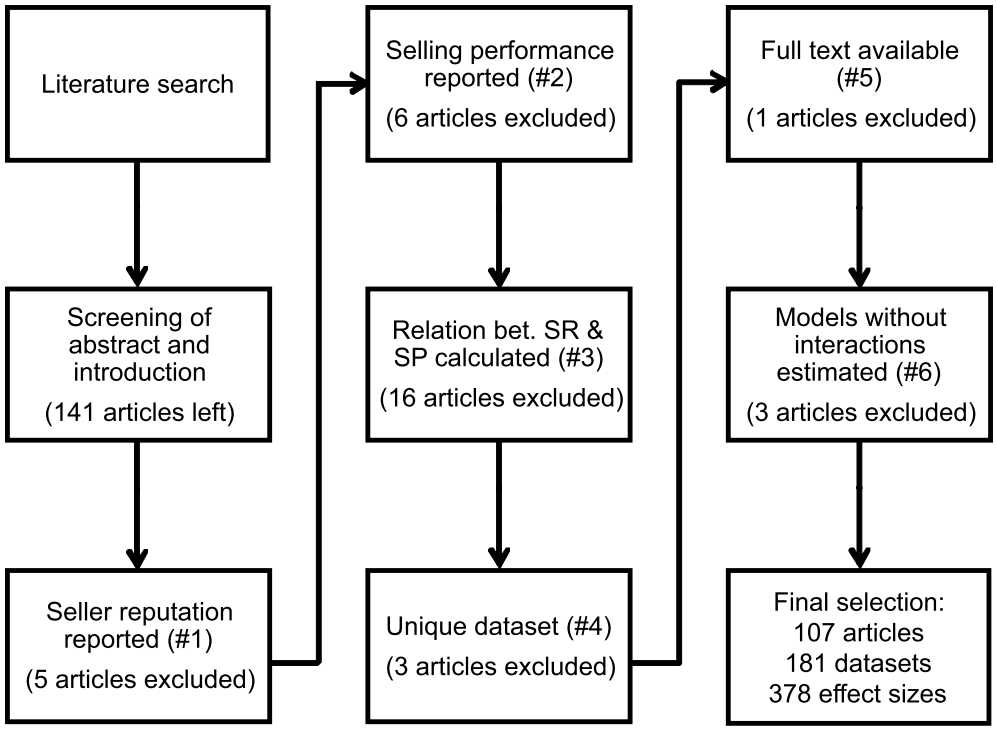

Fig. 2. Steps of literature search, study and model selection.

meta-analytic studies.

In their meta-analysis, Liu et al. (2007) focus on the relationship between seller reputation (number of positive, neutral and negative ratings) and the success of online auctions (number of bids, price premium and selling probability). Their paper integrates findings from 42 articles, and uses combined significance tests for the meta-analysis. That is, their analysis only takes into account whether a regression coefficient has the expected sign and is statistically significant, and tests if overall a significant effect exists for a certain relationship (e.g. between the number of negative ratings and the number of bids). Another meta-analytic study conducted by Schlägel (2011) includes 58 articles researching the effects of seller reputation (number and/or ratio of positive ratings, neutral and negative ratings) on online auction outcomes (selling probability, number of bidders, number of bids and the final price). Similarly to Liu et al. (2007), Schlägel (2011) only considers the direction and sign of each type of relationship to obtain an overall assessment.

These two articles provide a first general overview of existing research on reputation effects in online markets. Our study extends these two studies in the following respects: (1) While Liu et al. (2007) and Schlägel (2011) account for the literature published before 2007 and 2011, respectively, our meta-analysis includes studies published until September 2018. (2) Our analysis also considers effect sizes and not only the signs and statistical significance of reputation effects, so called vote-counting. Vote-counting has been criticised for being ineffective in finding small but exiting effects in research fields with mostly underpowered studies (Combs et al., 2011). (3) We make effect sizes as comparable as possible by using appropriate transformations and specifying their relative importance by accounting for the sample size based on which they are estimated. (4) Liu et al. (2007) and Schlägel (2011) limit their analyses to online auctions whereas we also consider studies that analyzed fixed-price transactions.

The studies included in Liu et al. (2007) and Schlägel (2011) form our initial set of studies to be included in our analyses. We conducted our literature search on Web of Science, Google Scholar, RePEc (Research Papers in Economics) and CNKI (China National Knowledge Infrastructure, in Chinese). The following search strings were used: (online auction OR Internet auction OR eBay OR Taobao) AND (reputation OR rating OR feedback). Next we checked the titles, abstracts, and introduction sections of each study for their relevance for our analyses. Moreover, we checked the online platform, the time period of data collection and the type of products for each dataset to make sure there is no overlap between studies in terms of datasets. The search process resulted in 141 relevant research articles written in English, Chinese or German. The reference list of all these articles is provided in the online supplementary material, Section A. Fig. 2 summarizes the study selection process that we employed. For a general approach to study selection for systematic reviews and meta-analyses, see Moher et al. (2009).

In the next step, we took a closer look at the full body of each article, in particular at hypotheses, descriptions of datasets, and results. At this stage, 31 of the 141 articles were excluded for the following reasons:

(1) No comparable seller reputation information reported: Seller reputation is mostly measured by means of the reputation score (i.e. the number of positive ratings minus the number of negative ratings), the number or percentage of positive, neutral, and/or negative ratings. Studies using other types of reputation measures are excluded. For example, some studies use the duration of sellers' membership in the online market as a measure of these sellers' reputations, and some operationalize seller reputation as a dummy variable indicating whether a seller is a 'top-seller'. These studies are excluded (indicated with #1 in online supplementary material Section B).
(2) No selling performance reported: Selling performance is mostly measured by means of the final price of a fixed-price transaction or auction (i.e. highest bid), the selling probability (i.e. whether or not an item was sold), the selling volume (i.e. amount of items sold within a limited time period), and the price ratio of a sold item indicating the rate of the selling price to the standard or average price of similar items. Studies are excluded if they do not report any of these variables measuring selling performance (indicated with #2 in online supplementary material Section B).
(3) No relationship between selling performance and seller reputation is reported: Our meta-analysis requires that the association between seller reputation and selling performance is calculated and reported including information on coefficients and *t*-values (or *p*-values and standard errors). Studies without empirical analysis of the relationship between selling performance and seller reputation, or studies in which these relationships are reported without explicit mention of test-values, are excluded (indicated with #3 in online supplementary material Section B).
(4) Same dataset used in another study: Some authors use the same dataset in more than one article. In these cases only the most recent analysis or the one with the best fitting model is included (indicated with #4 in online supplementary material Section B).
(5) Full text of the article is not available: One paper included in Schlägel (2011) is not available in online databases, and the author's contact information is not available. It is therefore not included in our analysis (indicated with #5 in online supplementary material Section B).

3.2. Model selection

After removing studies based on the exclusion criteria listed above, 110 research articles are left. During the screening process, we noticed that many authors ran multiple models on the same kind of reputation effect on the same type of selling performance, and these models often produced similar results. To avoid the inclusion of the same kind of effect from the same study in our analysis more than once, only one model for each type of selling performance is selected as the final model for the calculation of effect sizes. The selection is based on the following criteria:

(1) No interaction effects with seller reputation: Some studies include interaction effects with seller reputation to explore how other factors (i.e. moderators) may affect the reputation effect. For example, in the study by Cai et al. (2013), the second model reported in table 6 includes seller reputation and an interaction of seller reputation and a dummy variable indicating if it is before or after a 'buyer protection system' is implemented in the online market. The result shows that there is a significant interaction effect of seller reputation and the introduction of a 'buyer protection system', which indicates that the reputation effect varies between two subsamples. However, no further information regarding the subsamples is reported in the article. Therefore, we cannot use the coefficients of this model in our meta-analysis. Unless there is enough information provided to disentangle reputation effects for each subsample (such as in Jin and Kato 2006), models that interact seller reputation with other variables are excluded. There are three studies reporting only regression models with such interaction effects, so these studies are excluded at this stage (indicated with #6 in online supplementary material Section B).
(2) Interpretation by authors in the result section: Models analyzed and interpreted by the authors as the main result of the research are prioritized. Since these models were selected by the authors as final research findings, these models are regarded as the most informative and suitable for drawing substantial conclusions.
(3) The best fitting model: Adding to the two conditions above, the selection of the final model is made based on the reported goodness of fit. With all other conditions equal, the model with the best model fit is selected.

3.3. Effect sizes

The correlational effect sizes (r), variances of these effect sizes (v_r), and the corresponding sample sizes are necessary to perform a meta-analysis. Depending on the features of statistical modelling used in each study, we use one of two methods to calculate r (and v_r) from reported coefficients.

3.4. Pearson correlation coefficients

Some of the studies that we selected for our meta-analyses report effect sizes from bivariate relationships only. If the Pearson correlation coefficient (ρ) is used to quantify the relation between seller reputation and selling performance, ρ is used as the correlational effect size (eq. (3)). The variance of the effect size is calculated based on r and the corresponding sample size n (eq. (4)) (see Borenstein et al., 2009: Ch. 6).

$$r = \rho \tag{3}$$

$$v_r = \frac{(1 - r^2)^2}{n - 1} \tag{4}$$

3.5. Multiple linear regression coefficients

Of the 378 coefficients included in our meta-analysis, 40 are zero-order correlations. However, 338 coefficients stem from multiple regression models. In these cases, the methods used for meta-analysis of effect sizes from bivariate relationships cannot be applied (see e.g. Lipsey and Wilson 2001: 67–71). The reporting of non-standardized coefficient estimates and the different sets of explanatory and control variables used across different studies make comparability of coefficients difficult (although see Bowman 2012; Peterson and Brown 2005, respectively). However, our primary aim is to establish the general existence and variation of the reputation effect across different operationalizations of seller reputation and selling performance. This means that we do not need to rely on effect sizes from bivariate relationships only. In this case, the literature suggests several ways forward (also see Aloe and Becker 2012; Bowman 2012; Tong and Guo 2019).

First, Borenstein et al. (2009: 314) point out that regression coefficients and their standard errors could be used directly in meta-regression if the aim is to examine in how far study-level characteristics affect effect sizes rather than obtaining an overall regression coefficient. This approach could be particularly fruitful if the research question addressed by means of multiple regression is unambiguous and consensus exists on which explanatory and control variables should be used in the models (e.g., as in the estimation of the determinants of housing prices; see Sirmans et al., 2006). This does not apply in our case.

Second, unstandardized regression coefficients can be standardized and used in meta-analysis if information on both the standard deviation (SD) of the predictor variable X and the SD of the target variable Y are reported (Bowman 2012). However, in our case, the necessary information is mostly lacking and even if studies report the SDs of the predictor and target variables, if log-transformed variables are used for example, it is not possible to calculate their SDs from the SDs of the untransformed variables.

Third, Aloe and Becker (2012) show how the semi-partial correlation between variables X and Y included in a multiple regression model can be computed from the t-value of the coefficient of X, the R^2 of the regression model and the model's degrees of freedom ($df =$

$n - k - 1$, where n is the number of cases and k is the number of model parameters). Aloe and Becker (2012) point out that the use of semi-partial correlations (rather than partial correlations) is preferable as it comes closest to the bivariate correlation coefficients used in standard meta-analysis. However, in our case, this is not the aim and we therefore favor the fourth approach.

Fourth, effect sizes can be calculated as partial correlations (Aloe 2014; Aloe et al., 2017; Rosenthal 1991; Tong and Guo 2019).[3] We use partial correlations in our meta-analyses because they present the relationship of the effect of interest while controlling for the number of predictors included in each regression model. Partial correlations have been used as effect sizes in previous meta-analytic studies because they can be easily calculated from reported significance tests and make effect sizes comparable across different operationalizations of the variables of interest (e.g., Djankov and Murrell 2002; Doucouliagos and Laroche 2003). The partial correlational effect sizes (r) can be calculated based on the corresponding t-value and degrees of freedom (df) of the regression model (eq. (5)). The variance of the effect size is then calculated accordingly (eq. (6)).

$$r = \frac{t}{\sqrt{t^2 + df}} \tag{5}$$

$$v_r = \frac{(1 - r^2)^2}{df} \tag{6}$$

Given that our partial correlations are quite heterogeneous in terms of controls used in the different studies, we also include the bivariate correlations as they form just a special case (with no controls), which is not in principle different from partial correlations with different sets of controls.

It is important to note moreover that the degrees of freedom (df) must be calculated differently, if studies account for clustered data by estimating cluster-robust standard errors. Studies that estimate reputation effects based on data containing, for example, repeated observations on same sellers must take into account that offers posted by the same seller are not independent. This can be done by applying various multilevel techniques, one of which is the calculation of cluster-robust standard errors of regression coefficients (Cameron and Trivedi 2005; Snijders and Bosker 2012). In the large majority of cases, the calculation of cluster-robust standard errors results in these standard errors becoming considerably larger (as compared to calculations that treat every observation as independent) and, consequently, corresponding t-values becoming considerably smaller. In these cases, to calculate effect sizes and variances of these effect sizes correctly, the appropriate degrees of freedom are $df = n_c - k - 1$, where n_c is the number of clusters and not the number of cases (see, e.g., StataCorp 2015: 478).[4]

Not all of the studies that we selected for our meta-analyses report the t-values of coefficient estimates. For studies that report standard errors (SE) of regression coefficients, the t-values are obtained by dividing the regression coefficients by the corresponding SEs. However, in 43 cases, the information on t-values and SEs is missing, and p-values are only reported in terms of 'stars' (i.e. p-value ranges). We apply two strategies to obtain the t-values from p-value ranges in these cases (see Table 1). We either take the median of the p-value range (Strategy 1), or we take the upper bound of the p-value range if the p-value is reported to be smaller than 0.1 (Strategy 2). For insignificant coefficients (i.e. for $p > 0.1$ or $p > 0.05$, depending on a study's cut-off for statistical significance), we assume $p = 0.5$. Our analysis is based on Strategy 1; our results do not change much when Strategy 2 is used instead (see online supplementary material Section C).

The t-values of regression coefficients can be calculated based on the corresponding p-values and df. These t-values are then used to calculate effect sizes and corresponding variances as described above (eqs. (5) and (6)).

This approach for linear models (e.g., OLS) we use as well for non-linear regression models such as logit and probit. In the latter case, however, we use z-values from these models instead of t-values to calculate partial correlational effect sizes and the corresponding effect size variances using equations (7) and (8), respectively. Since z-values do not depend on the number of model parameters, we use the number of cases (n) or the number of clusters (n_c) instead of the degrees of freedom (df) in these calculations.

$$r = \frac{z}{\sqrt{z^2 + n_c}} \tag{7}$$

$$v_r = \frac{(1 - r^2)^2}{n_c} \tag{8}$$

3.6. Fisher z-transformation

Correlational effect sizes are bound to be between −1 and 1. Their sampling distribution is therefore not normal, which makes the

[3] A semi-partial correlation establishes the relation between a predictor variable X and a target variable Y net of the portion of Y explained by other predictors used in the model. A partial correlation establishes the relation between a predictor variable X net of the portion of X explained by other predictors used in the model and a target variable Y net of the portion of Y explained by other predictors used in the model (Aloe and Becker 2012).

[4] In case of other multilevel approaches such as random intercept models, the appropriate degrees of freedom can be calculated from a weighted average of the number of cases (n) and the number of clusters (n_c) where the weight is the proportion of between-cluster variation (see also Aloe et al., 2017; Hedges 2007). None of the coefficient estimates included in our meta-analysis stems from such multi-level regressions.

Table 1
Strategies to determine *p*-values from reported *p*-value ranges.

Reported *p*-value range ('stars')	Number of cases	Used *p*-value is median (Strategy 1)	Used *p*-value is upper bound (Strategy 2)
$0 < p < 0.001$	2	0.0005	0.001
$0 < p < 0.01$	16	0.005	0.01
$0.01 < p < 0.05$	8	0.03	0.05
$0.05 < p < 0.1$	7	0.075	0.1
$p > 0.05$	3	0.525	0.5
$p > 0.1$	7	0.55	0.5

Notes: Two-tailed tests are assumed if not otherwise specified.

calculation of confidence intervals and comparisons of correlational effect sizes difficult. Fisher's *r-to-z* transformed correlation conversion is used, so that after the transformation, the sampling distribution of r becomes normally distributed (Fisher 1921). The transformed correlation coefficient (z) and the variance of z (v_z) can be calculated by means of equations (9) and (10), respectively. These transformed correlation coefficients are then used as units of analysis in our meta-analyses. The results of the meta-analyses are transformed back to correlation coefficients (r) for interpretation and presentation. However, there is an ongoing debate about whether the Fisher z-transformation should be applied to partial correlations as well (Aloe and Becker 2012; Suurmond et al., 2017). Our results hardly change if we perform the analyses described below without first transforming the data.

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{9}$$

$$v_z = \frac{1}{n-3} \tag{10}$$

4. Data

Our dataset consists of 378 effect sizes, estimated with 181 different datasets, reported in 107 studies. The dataset was created by hand-coding the relevant information contained in the 107 studies. In a first step, the hand coding was performed and double-checked by one of the authors. In a second step, the other two authors independently checked a total of 125 (about 33%) of the data rows pertaining to the 378 coefficients. In a third step, the entire dataset was checked once more and updated based on insights gained in the second step. Upon publication of this article, we will make our data available via a public repository for reproduction purposes.

Figs. 3 and 4 provide overviews of the dependent and independent variables in the included studies. Descriptive statistics of our dataset are presented in Table 2, and, where available, mean prices of products at dataset level are reported in Table 3.

Fig. 3 shows the number of studies that use one or several of the four dependent variables in their estimation of reputation effects. For example, out of the 107 studies included in our meta-analyses, 42 studies estimate reputation effects using only the final price as the dependent variables, and 15 studies use both the final price as well as the selling probability as dependent variables. Correspondingly, Fig. 4 shows the number of studies that use one or several of the three operationalizations of seller reputation in their estimation of reputation effects. For example, there are 34 studies, that use only the reputation score and 25 studies that use both the number of positive and the number of negative ratings but not the reputation score.

5. Results

Meta-analyses are performed separately for each of the twelve types of reputation effects, i.e. each combination of type of seller reputation and selling performance. As mentioned above, for methodological reasons, the effect size r is z-transformed for meta-analysis; thereafter, the overall effect is converted back to facilitate interpretation. The meta-analyses are performed with the 'metafor' package in R (Viechtbauer 2010). The overall reputation effects are estimated using random-effects models. We also assess the heterogeneity of the results with Q-statistics and I^2 as well as publication bias by reporting the Egger's regression test in Table 4.

5.1. Overall effect sizes

Table 4 lists the overall effect sizes (*ES*) and corresponding 95% confidence intervals (CI) that resulted from the twelve meta-analyses performed for each combination of outcome and reputation variables. For each overall effect size, the table also shows the results of heterogeneity measures (Q and I^2) and the assessment of publication bias (Egger's test). We discuss overall effect sizes first.

The results reported in Table 4 corroborate the general existence of the reputation effect. All overall effect sizes point in the expected direction. However, three of the twelve overall effect sizes are statistically insignificant. With final price as the outcome variable, the reputation score and the number of positive ratings have a significantly positive overall effect ($ES = 0.05$, $p = 0.025$ and $ES = 0.11$, $p < 0.001$, respectively), and the number of negative ratings has a significantly negative overall effect ($ES = -0.10$, $p < 0.001$). Results are similar if the selling price relative to a reference value (i.e. price ratio) is used as the outcome variable. The reputation score and the number of positive ratings have a significantly positive effect ($ES = 0.08$, $p = 0.039$ and $ES = 0.28$, $p < 0.001$, respectively), and the number of negative ratings shows a negative but statistically insignificant overall effect ($ES = -0.06$, $p = 0.160$).

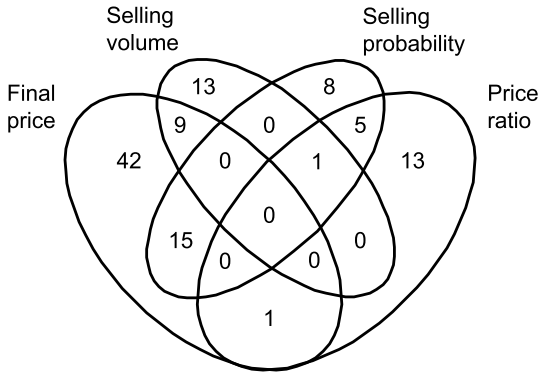

Fig. 3. Summary of dependent variables in included studies.

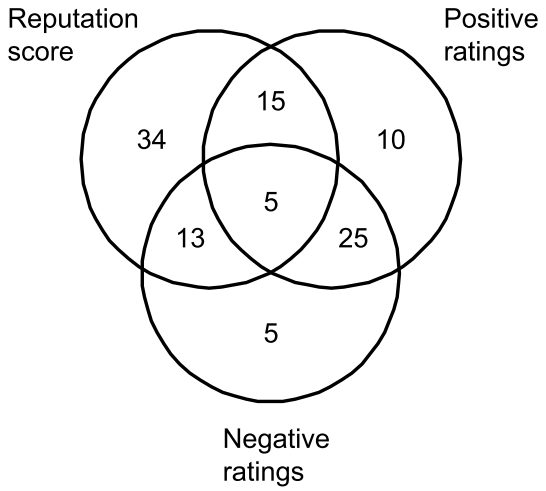

Fig. 4. Summary of independent variables in included studies.

With selling probability as the outcome variable, reputation scores and the number of positive ratings exhibit statistically significant increases in the probability of sale ($ES = 0.04$, $p = 0.016$ and $ES = 0.07$, $p < 0.001$, respectively), whereas the number of negative ratings has a significantly negative effect on selling probability ($ES = -0.05$, $p < 0.001$). Finally, with selling volume as the outcome variable, the reputation score and the number of positive ratings exhibit positive effects on the number of sold items but only the overall effect of the number of positive ratings is statistically significant ($ES = 0.08$, $p = 0.085$ and $ES = 0.14$, $p < 0.001$, respectively). The effect of the number of negative ratings points in the expected direction but is statistically insignificant ($ES = -0.06$, $p = 0.331$).

Overall, these results corroborate that a good seller reputation has a positive effect on selling performance. Especially the number of positive ratings exhibits a consistent, significantly positive effect on all types of outcome variables. The reputation score also has positive but generally smaller effects on selling performance than the number of positive ratings. Results regarding the number of negative ratings are mixed although all overall effects are negative, as expected.

Given that correlational effect sizes can range between −1 and 1, the overall effect sizes reported in Table 4 appear relatively small. This by no means should be interpreted as 'weak' or small average reputation effects. In order to interpret reputation effects substantially, a single study should be considered. For example, the coefficient of the log-transformed number of positive ratings in a model of log-transformed price (in EUR) reported by Przepiorka (2013) has a partial correlational effect size of 0.08. However, the raw coefficient is 0.078 and can be interpreted as follows: a tenfold increase in the number of positive ratings (e.g., from 40 to 400) corresponds with a final price increase of $100 \times [\exp(0.078 \times \log 10) - 1] = 20\%$. Based on the average selling price of items analyzed in this study (about EUR 15), the increase in a seller's positive reputation corresponds to a price increase of EUR 2.95. The overall effect sizes reported in Table 4 can be interpreted as corroborations of the existence of reputation effects across different studies and operationalizations of seller reputation and seller performance. We will have a closer look at effect size heterogeneity next.

5.2. *Effect size heterogeneity*

Statistical heterogeneity refers to the variability in effect sizes that cannot be attributed to sampling variability only. For each type of reputation effect reported in Table 4, we assess the extent of statistical heterogeneity using Cochrane's homogeneity test and the I^2 statistic.

Cochran's Q is a measure of weighted squared deviations around the overall effect size (see Borenstein et al., 2009:109–113). All

Table 2
Descriptive information on included studies and datasets.

Study level	*N*	*n* Min	*n* Max
Publication status			
Journal (English)	78	14	3,981,429
Journal (Chinese)	12	137	182,853
Working paper	1	180	473,152
Conference/workshop	4	22	1379
Book chapter	4	117	2133
Dissertation/thesis	6	129	16,032
Unpublished	2	169	807
Study type			
Observational	105	14	3,981,429
Experimental	1	137	982
Mixed [a]	1	100	1124
Dataset level	***N***	***n* Min**	***n* Max**
Country			
US	113	14	1,311,452
China	48	22	182,853
Europe [b]	18	192	3,981,429
Mixed [c]	2	418	55,094
Platform			
eBay	120	14	339,517
Taobao	38	22	1,311,452
Yahoo	7	91	551
Eachnet	4	1053	182,853
Priceminister	4	1,759,572	3,981,429
Silk Road	3	119	16,243
Other/mixed [d]	5	204	15,033
Year of data collection [e]			
1998	2	407	460
1999	11	94	1822
2000	40	14	861
2001	12	100	9981
2002	13	82	182,853
2003	10	117	2133
2004	10	126	339,517
2005	8	91	1665
2006	10	107	89,982
2007	12	38	14,689
2008	15	95	3,981,429
2009	8	22	4226
2010	4	445	1,311,452
2011	6	205	1251
2012	4	196	16,032
2013	2	119	1379
2014	2	3433	16,243
2015	10	137	982
2016	2	237	15,033
Transaction type			
Auction	132	14	339,517
Fixed-price	44	22	3,981,429
Unknown/mixed [f]	5	38	14,689

[a] Jin and Kato (2006) reported results from an observational and experimental study.
[b] The category 'Europe' includes France, Germany, Finland, Poland, Switzerland.
[c] Lei (2011) collected a dataset with eBay sellers from 42 countries, and Snijders and Zijdeman (2004) combined data from two Dutch and two US websites.
[d] Other platforms are Alegro, Huuto, Bizerate and Ricardo. The dataset of Snijders and Zijdeman (2004) was collected from eBay.nl ($n = 111$), Ricardo. nl ($n = 125$), eBay.com ($n = 103$), and ePier.com ($n = 79$).
[e] If year of data collection was not specified, the study's publication year was used.
[f] Chen et al. (2018), Przepiorka (2013) and Zhu et al. (2009) reported datasets combining auctions and fixed-price offers.

Table 3
Descriptive information on included studies' datasets.

Category/Product	Mean price (in USD)	# Datasets
Toy and hobby		
Boad game	131.9	1
Game ticket	n/a	1
Computer/video games	29.11	5
Coin	62.28	25
Stamp	33.07	2
Artwork	47.74	1
Music instrument	1636.04	2
Baseball card	75.45	4
Sport equipment	509.96	1
Toy	263.21	2
CD/DVD	9.77	13
Book	180.48	4
Computer and electronics		
PDA	252.75	4
Scanner and printer	260.5	2
Memory disk	76.06	7
Computer software	434.36	5
(Video) camera	681.27	12
DVD player	286.1	2
MP3 player	197.12	13
Mobile phone	453.74	16
Game station	n/a	1
Computer/laptop	n/a	2
Other electronic device	167.13	7
Other		
Car	7601	3
Prepaid/gift card	31.4	6
Kitchen supply	56.2	8
Illegal drug	99.8	3
Clothing	19.35	2
Food/drink	33.98	7
Watch	791.61	2
Cosmetics	18.98	3
Mixed/unavailable	256.96	15

Notes: The mean prices are calculated at the dataset level.

Table 4
Overall effect sizes.

				Heterogeneity test		
Relation	*ES*	95% CI	*N*	Q	I^2 (%)	Egger's test
Final price						
Reput. score	0.05*	[0.01, 0.09]	66	1118.06***	99.76	−0.91
Pos. ratings	0.11***	[0.06, 0.15]	53	555.07***	99.15	3.12**
Neg. ratings	−0.10***	[−0.13, −0.07]	44	288.01***	92.04	−2.42*
Price ratio						
Reput. score	0.08*	[0.00, 0.15]	16	410.68***	99.17	0.25
Pos. ratings	0.28***	[0.18, 0.37]	35	485.31***	99.62	5.10***
Neg. ratings	−0.06	[−0.15, 0.03]	25	183.76***	87.14	1.94
Selling probability						
Reput. score	0.04*	[0.01, 0.07]	26	965.71***	96.59	−2.25*
Pos. ratings	0.07***	[0.04, 0.10]	19	1146.62***	97.20	0.05
Neg. ratings	−0.05***	[−0.07, −0.03]	16	39.86***	62.36	−1.73
Selling volume						
Reput. score	0.08	[−0.01, 0.16]	31	5514.45***	99.81	−2.12*
Pos. ratings	0.14***	[0.09, 0.20]	31	892.68***	99.41	4.29***
Neg. ratings	−0.06	[−0.17, 0.06]	16	103.85***	92.58	−3.15**

Notes: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

but one Q-values are statistically significant leading us to reject the null hypothesis of homogeneity in effect sizes in these cases. I^2 describes the percentage of between-study variability to total variability (i.e. within and between study variability in effect sizes). It indicates how (in)consistent findings are across studies; it is not a measure of the variation of the true effect (see Borenstein et al., 2009:117–119). Eleven out of the twelve I^2 values presented in Table 4 are above 75%, suggesting that a large part of effect size heterogeneity results from differences in true effect sizes rather than sampling variability.

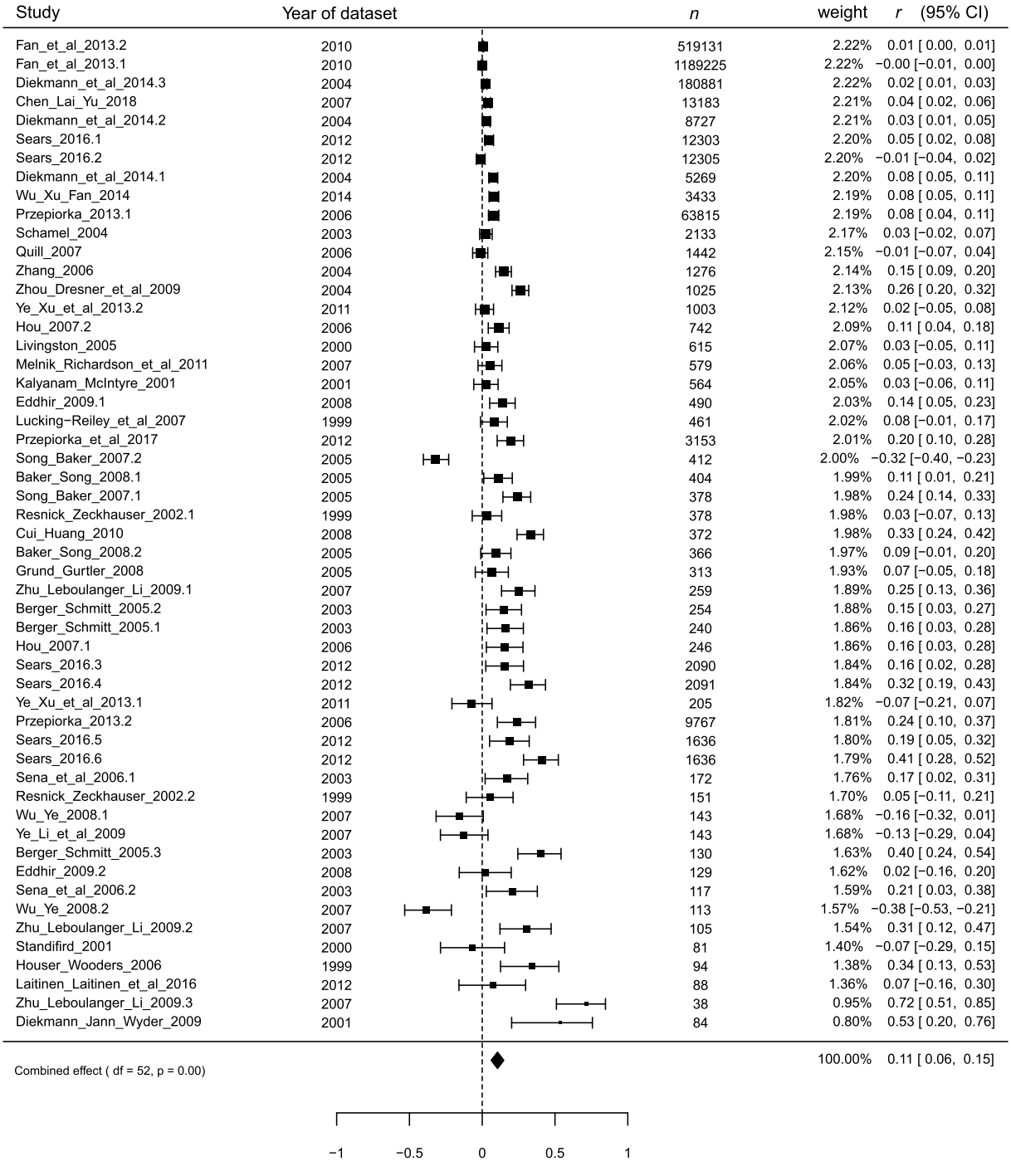

Fig. 5. Forest plot of effect sizes of positive ratings on final prices.

These results suggest that statistical heterogeneity is very high. It can be attributed to differences in study designs (e.g., type and number of explanatory and control variables in multiple regression models), market platforms (eBay, Taobao, etc.), samples of product items (price levels, unaccounted heterogeneity) etc. (also see Tables 2 and 3). To be better able to assess the extent of effect size heterogeneity, we will next have a look at four forest plots.

The four forest plots shown in Fig. 5 through 8 list the study abbreviations, years of data collection, sample sizes, meta-analytic weights, and effect sizes together with their 95% confidence intervals. The effect sizes are also plotted with their 95% confidence

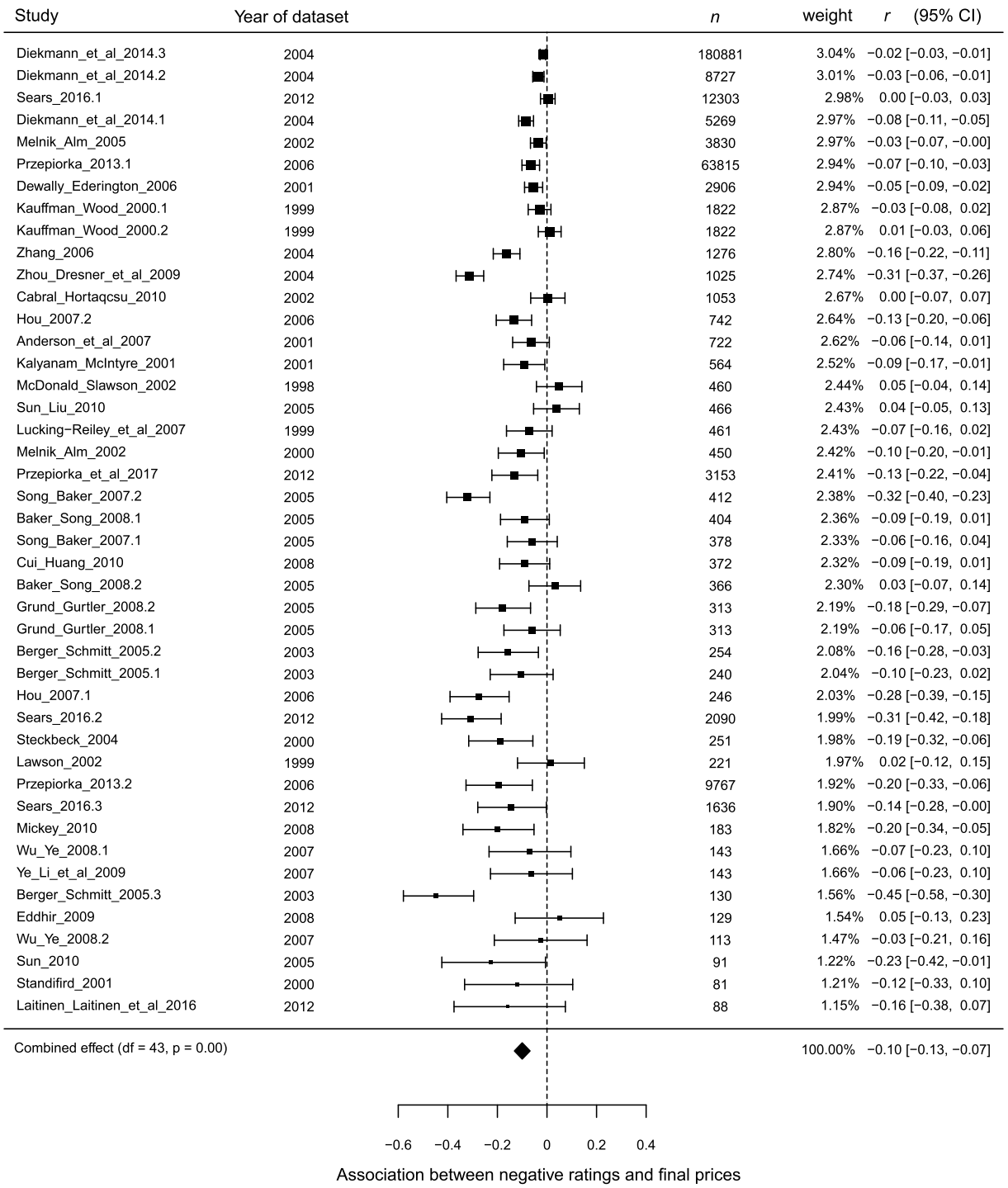

Fig. 6. Forest plot of effect sizes of negative ratings on final prices.

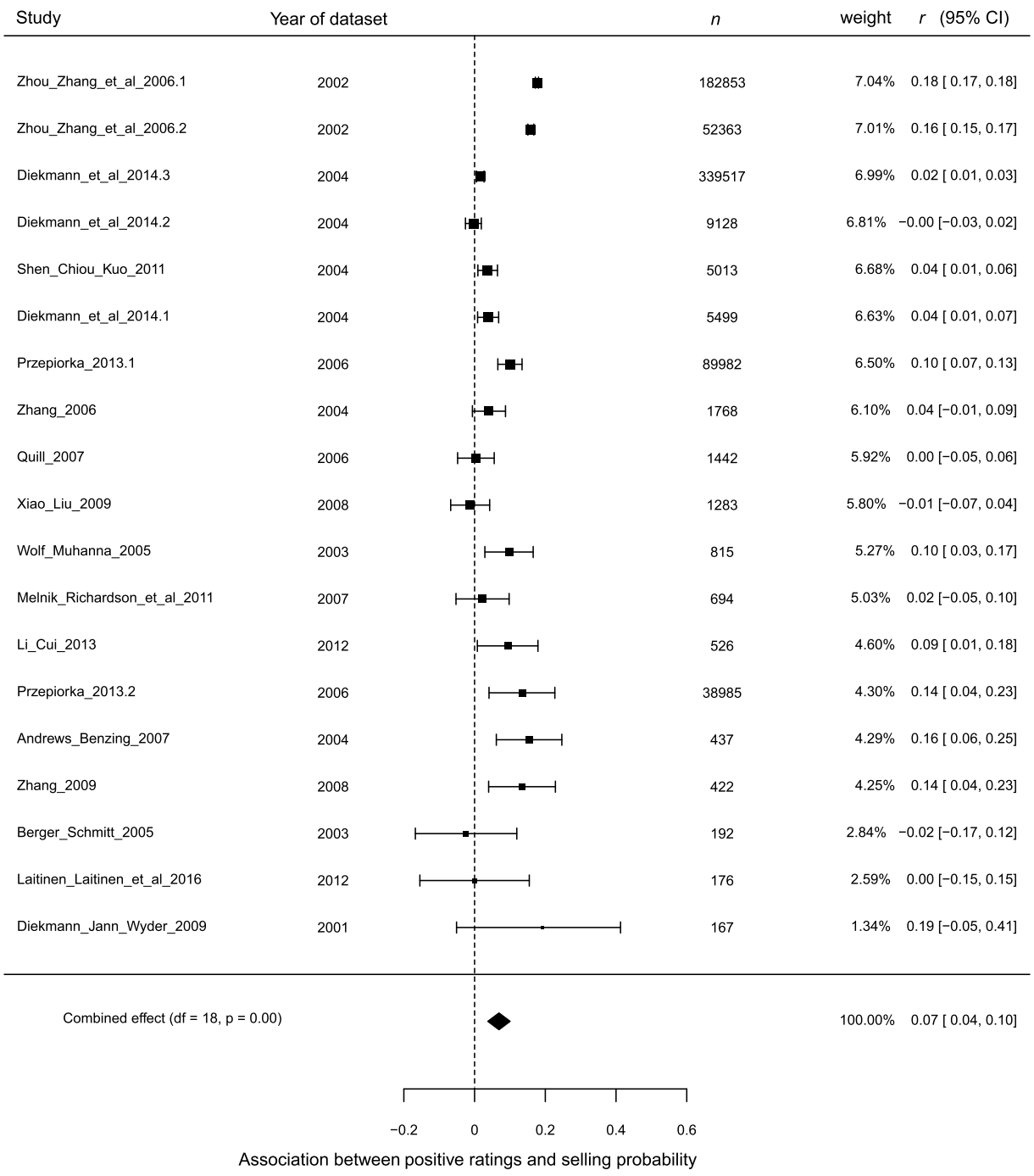

Fig. 7. Forest plot of effect sizes of positive ratings on selling probabilities.

intervals. The diamond at the bottom of each figure indicates the combined result of all individual effect sizes (i.e. the overall effect size that is also reported in Table 4). The four forest plots are for effect sizes of positive ratings on final price, negative ratings on final price, positive ratings on selling probability and negative ratings on selling probability.

Fig. 5 shows the forest plot of the effect sizes of the number of positive ratings on final price. Positive ratings have a significantly positive overall effect on final prices ($ES = 0.11$, $p < 0.001$); also, it presents a significant heterogeneity across the included studies ($Q(df = 52) = 555.07$, $p < 0.001$). It is noticeable that the study of Song and Baker (2007) reports a negative correlation with a relatively high weight. Tracing back to the original paper, it indicates a negative Pearson correlation ($\rho = -0.32$) between the number of positive feedback ratings and prices of MP3 players on eBay; but with the same dataset, there is a same negative Pearson correlation ($\rho =$

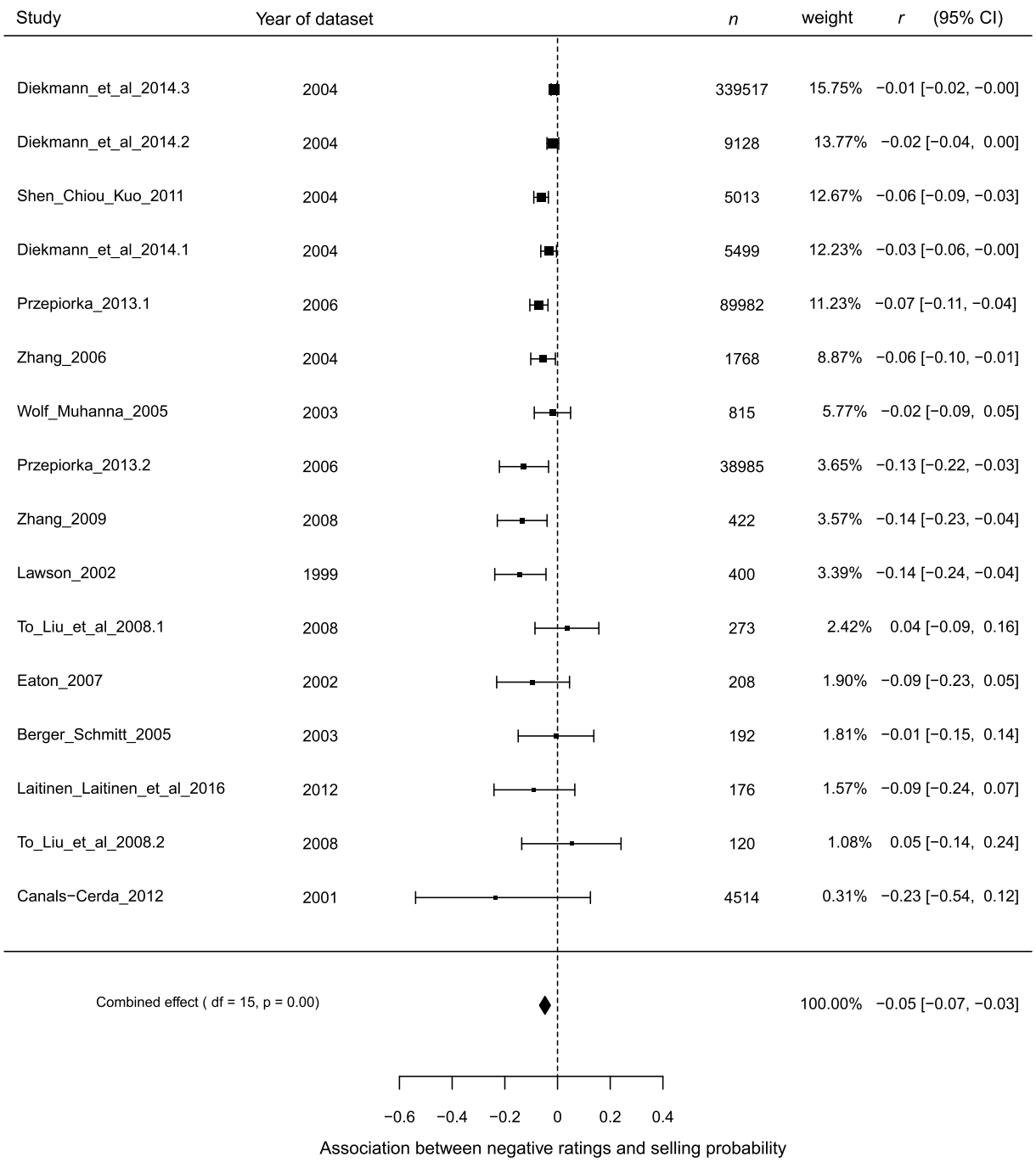

Fig. 8. Forest plot of effect sizes of negative ratings on selling probabilities.

−0.32) between the number of negative feedback ratings and prices, suggesting positive and negative feedback ratings have the same effect on selling prices, which is an unexpected result.[5]

The forest plot of the association between the number of negative ratings on final price is shown in Fig. 6. The combined result of all individual effect sizes is negative and statistically significant ($ES = -0.10, p < 0.001$); it also presents a significant heterogeneity across

[5] Since it seemed plausible to assume that this was a typo, we re-estimated the overall effect size excluding the negative coefficient reported in Song and Baker (2007). While the overall effect size remains the same, the 95% CI becomes slightly smaller ($ES = 0.11$, $p < 0.001$, 95% CI [0.07, 0.15]).

the included studies (Q (df = 43) = 288.01, $p < 0.001$). However, some studies report non-negative effects. For instance, the dissertation of Sears (2016) reports a null effect of the log-transformed number of negative ratings on selling prices of marijuana sold via the cryptomarket Silk Road. Although small and insignificant, this effect size exhibits one of the highest weights (2.98%) in the meta-analysis.

Fig. 7 presents the forest plot of reputation effects of positive ratings on selling probability. The combined result of all individual effect sizes is positive and statistically significant ($ES = 0.07$, $p < 0.001$); the heterogeneity test on the included effects sizes is significant (Q (df = 18) = 1146.62, $p < 0.001$), indicating that there is a great variation in effect sizes across studies. One of the coefficients pointing in the opposite direction is reported by Xiao and Liu (2009); it suggests that the percentage of positive ratings has a negative but insignificant effect on the selling probability of Nokia mobile phones in the Chinese online market Taobao.

Fig. 8 is the forest plot of the effect sizes for selling probability and the number of negative ratings. The figure shows that negative ratings have a small, yet significantly negative overall effect on selling probability ($ES = -0.05$, $p < 0.001$); and the heterogeneity is significant among the included effect sizes (Q (df = 15) = 39.86, $p < 0.001$). Of the 16 included effect sizes, two exhibit insignificantly positive values (To et al., 2008) on Yahoo.

5.3. Publication bias

Lastly, the Egger's regression test (Egger et al., 1997) for funnel plot asymmetry on each set of meta-analyses is reported in Table 4. Seven out of 12 meta-analyses exhibit asymmetric funnel plots (i.e. statistically significant Egger's test statistics). This is a first indication of publication bias. However, only five of these seven tests have a sign in line with the sign of the hypothesized reputation effect. For example, Egger's test in case of the effect of the number of positive ratings on final price has the same sign as the overall effect size estimate. Here, funnel plot asymmetry is likely due to publication bias. In case of the effect of the reputation score on selling probability, the Egger's test has a negative sign whereas the overall effect size estimate has, as expected, a positive sign. This may imply that smaller studies that show a large standard error and are more prone to publication bias exhibit a lower reputation effect.

Simulation studies have shown that the Egger's test is not sensitive to effect size heterogeneity stemming from sampling error. However, if effect size heterogeneity is due to differences in samples and study designs, which is the case in our meta-analyses, Egger's test may be biased (see, e.g., Schneck 2017; Sterne et al., 2011).

6. Discussion and conclusions

The increasing popularity of peer-to-peer online markets brings attention to the role of reputation systems, which collect and present information on the trustworthiness and competence of traders based on their past online market exchanges (Dellarocas 2003; Diekmann et al., 2014). From a game-theoretic perspective, information about seller reputation helps to promote buyer trust as it decreases untrustworthy behaviors of sellers. Sellers have to behave cooperatively to build and maintain and good reputation, and since they also have to offer discounts when entering the market, their reputations and business success in terms of prices and sales will be correlated (Przepiorka 2013; Shapiro 1983).

This relation between seller reputation and success, which also has been shown to be causal (Przepiorka 2013; Snijders and Weesie 2009), is called the reputation effect. In the last 20 years, a large body of literature has accumulated that seeks to find evidence for the reputation effect in real-world online markets. However, past studies present inconsistent results and there is a lack of consensus on what the reputation effect means and how substantial it might be (Lindenberg et al., 2020; Snijders and Matzat 2019).

In this paper we integrated evidence from 107 existing studies, including 378 coefficients estimated based on 181 different datasets comprising a total of 14.04 million observations of online market transactions. We conducted twelve separate meta-analyses, one for each combination of three seller reputation variables and four seller performance variables commonly used in the literature. This approach allowed us to establish the robustness of the reputation effect across different operationalizations of seller reputation and selling performance.

To our knowledge, our study incorporates the largest number of studies among the existing systematic reviews on the subject of reputation effects and is the first to consider effect sizes rather than only signs and statistical significances of reputation effects (also see Liu et al., 2007; Schlägel 2011). We were also able to interpret papers in languages other than English. There are thirteen papers written in Chinese and one paper in German. Moreover, we exhibited great effort to incorporate any possible study or research outcome. For instance, 11% (43 out of 378) of the coefficients we used were not accompanied with information about standard errors, t-scores or p-values, which are needed to calculate effect sizes and make them comparable. Instead only p-value ranges indicated by stars were reported in these cases. We proposed different strategies (as reported in Table 1) to determine estimated p-values for subsequent calculations of effect sizes. Finally, we used a relatively new approach that relies on the calculation of partial correlation coefficients to make effect sizes comparable (Aloe 2014; Aloe et al., 2017; Rosenthal 1991; Tong and Guo 2019).

Our results show that seller reputations affect seller performance in the expected directions: overall, positive ratings have positive effects on all types of selling performance and negative ratings have negative effects (although two of the four negative overall effects are statistically insignificant). Although the overall effect sizes (as reported in Table 4) appear to be small, they should not be interpreted as 'weak' or substantially insignificant. Tracing back effect sizes to the original studies reveals that what sellers in online markets obtain for a good reputation can be substantial. However, the reputation effects included in our meta-analyses exhibit a high degree of heterogeneity that cannot be attributed to sampling error only. This is not entirely surprising given the differences in market platforms, item data, and modelling approaches used across studies (see Tables 2 and 3). Although we already grouped coefficients that were estimated using the same type of seller reputation and performance variables, we will try to identify the different sources of

variation of reputation effects reported in previous literature by means of subgroup analysis and meta-regression (see, e.g., Tong and Guo 2019) in a subsequent paper.

There are three categories of moderator variables that suggest themselves: (1) contextual factors (e.g., market platform, geo-cultural region of market participants, time), (2) overall characteristics of the traded products (e.g., price category, usage status, complexity), and (3) methodological factors (e.g., number and type of explanatory variables and controls, specification of functional form of seller reputation, statistical model construction). For example, more expensive, more complex and used rather than new products face buyers with higher risks and uncertainty. It can be therefore expected that these product characteristics will have a positive, moderating effect on the reputation effect. However, it remains to be shown in how far meta-analysis that uses coefficient estimates from multiple regression models can shed light on substantial (rather than methodological) reasons for the variation in reputation effects. We conclude this paper with pointing out an often neglected, substantial reason for the excess variation in reputation effects.

Note that in our game theoretic model above, we made the implicit assumption that, after every transaction, a seller is rated truthfully with certainty. Relaxing this assumption does not only pay justice to the fact that a substantial part of transactions are not rated or not rated truthfully (Dellarocas and Wood 2008; Diekmann et al., 2014), but it also unveils a substantial reason for why the size of the reputation effect may vary across markets and within markets over time. The intuition behind this theoretical argument goes as follows (for a formal derivation see Przepiorka 2013): The lower the rate of truthful ratings is, the longer it will take to identify untrustworthy sellers. The longer it takes to identify untrustworthy sellers, the higher is the incentive for these sellers to enter the market. The more untrustworthy sellers enter the market, the higher will be the probability of encountering an untrustworthy seller ($1 - \alpha$). By equation (2), the higher is $1 - \alpha$, the higher will be the price discount d honest sellers will have to make when entering the market.

This argument results in a seemingly counterintuitive conjecture: The more effective a reputation system is in identifying dishonest sellers in an online market, the smaller will be the reputation effect. This is an important point to make because a small reputation effect is often used as first evidence for positive evaluation bias and the malfunctioning of a reputation system (e.g., Tadelis 2016). In other words, a reputation system may be effective not because sellers earn large premiums for their good reputations, but because the mere presence of the reputation system attracts a majority of trustworthy and reliable sellers (Diekmann et al., 2014). Bockstedt and Goh (2011) found evidence that in an online market concentrated with experienced and reputable sellers, the reputation scores indeed are less relevant for seller differentiation. If, as a result, buyers' a priori levels of trust are high, these buyers will be less inclined to pay for reputation and the reputation effect will thus be smaller.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssresearch.2020.102522.

References

Aloe, A.M., 2014. An empirical investigation of partial effect sizes in meta-analysis of correlational data. J. Gen. Psychol. 141 (1), 47–64.
Aloe, A.M., Becker, B.J., 2012. An effect size for regression predictors in meta-analysis. J. Educ. Behav. Stat. 37 (2), 278–297.
Aloe, A.M., Becker, B.J., Duvendack, M., Valentine, J.C., Shemilt, I., Waddington, H., 2017. Quasi-experimental study designs series—paper 9: collecting data from quasi-experimental studies. J. Clin. Epidemiol. 89, 77–83.
Becker, G.S., 1976. Altruism, egoism, and genetic fitness: Economics and sociobiology. J. Econ. Lit. 14 (3), 817–826.
Blau, P.M., 1964. Exchange and Power in Social Life. Wiley, New York.
Bockstedt, J., Goh, K.H., 2011. Seller strategies for differentiation in highly competitive online auction markets. J. Manag. Inf. Syst. 28 (3), 235–268.
Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. Introduction to Meta-Analysis. John Wiley & Sons, West Sussex.
Botsman, R., Rogers, R., 2010. What's Mine Is Yours: the Rise of Collaborative Consumption. Collins, London.
Bowman, N.A., 2012. Effect sizes and statistical methods for meta-analysis in higher education. Res. High. Educ. 53 (3), 375–382.
Buskens, V., Raub, W., 2013. Rational choice research on social dilemmas: embeddedness effects on trust. In: Wittek, R., Snijders, T.A.B., Nee, V. (Eds.), The Handbook of Rational Choice Social Research. Stanford University Press, Stanford, pp. 113–150.
Cai, H., Jin, G.Z., Liu, C., Zhou, L.A., 2013. More Trusting, Less Trust? an Investigation of Early E-Commerce in China, w18961. National Bureau of Economic Research, pp. 1–26.
Cameron, C.A., Trivedi, P.K., 2005. Microeconometrics: Theory and Applications. Cambridge University Press, Cambridge.
Chen, K., Lai, H., Yu, Y., 2018. The seller's listing strategy in online auctions: evidence from eBay. Int. J. Ind. Organ. 5 (6), 107–144.
Coase, R.H., 1988. The firm, the market and the law. In: Coase, R.H. (Ed.), The Firm, the Market and the Law. University of Chicago Press, Chicago, pp. 1–31.
Combs, J.G., Ketchen, D.J., Crook, T.R., Roth, P.L., 2011. Assessing cumulative evidence within 'macro' research: why meta-analysis should be preferred over vote counting. J. Manag. Stud. 48 (1), 178–197.
Dasgupta, P., 1988. Trust as a commodity. In: Gambetta, D. (Ed.), Trust: Making and Breaking Cooperative Relations. Basil Blackwell, Oxford, pp. 49–72.
Dellarocas, C., 2003. The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manag. Sci. 49 (10), 1407–1424.
Dellarocas, C., Wood, C.A., 2008. The sound of silence in online feedback: estimating trading risks in the presence of reporting bias. Manag. Sci. 54 (3), 460–476.
Diekmann, A., Jann, B., Przepiorka, W., Wehrli, S., 2014. Reputation formation and the evolution of cooperation in anonymous online markets. Am. Socio. Rev. 79 (1), 65–85.
Diekmann, A., Przepiorka, W., 2019. Trust and reputation in markets. In: Giardini, F., Wittek, R. (Eds.), The Oxford Handbook of Gossip and Reputation. Oxford University Press, New York, pp. 383–400.
Djankov, S., Murrell, P., 2002. Enterprise restructuring in transition: a quantitative survey. J. Econ. Lit. 40 (3), 739–792.
Doucouliagos, C., Laroche, P., 2003. What do unions do to productivity? A meta-analysis. Ind. Relat. 42 (4), 650–691.
Egger, M., Smith, G.D., Phillips, A.N., 1997. Meta-analysis: principles and procedures. Br. Med. J. 315 (7121), 1533–1537.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114 (3), 817–868.
Fisher, R.A., 1921. Some remarks on the methods formulated in a recent article on "The quantitative analysis of plant growth". Ann. Appl. Biol. 7 (4), 367–372.

Friedman, E.J., Resnick, P., 2001. The social cost of cheap pseudonyms. J. Econ. Manag. Strat. 10 (2), 173–199.
Gregg, D.G., Scott, J.E., 2006. The role of reputation systems in reducing on-line auction fraud. J. Electr. Comm. 10 (3), 95–120.
Greif, A., 1989. Reputation and coalitions in medieval trade: evidence on the Maghribi traders. J. Econ. Hist. 49 (4), 857–882.
Güth, W., Ockenfels, A., 2003. The Coevolution of trust and institutions in anonymous and non-anonymous communities. In: Holler, M.J., Kliemt, H., Schmidtchen, D., Streit, M. (Eds.), Jahrbuch Für Neue Politische Ökonomie Vol 20. Tübingen: Mohr Siebeck, pp. 157–174.
Hedges, L.V., 2007. Effect sizes in cluster-randomized designs. J. Educ. Behav. Stat. 32 (4), 341–370.
Hillmann, H., 2013. Economic institutions and the state: insights from economic history. Annu. Rev. Sociol. 39, 251–273.
Jin, G., Kato, A., 2006. Price, quality, and reputation: evidence from an online field experiment. Rand J. Econ. 37 (4), 983–1005.
Lei, Q., 2011. Financial value of reputation: evidence from the eBay auctions of gmail invitations. J. Ind. Econ. 59 (3), 422–456.
Li, J., Tang, J., Jiang, L., Yen, D.C., Liu, X., 2019. Economic success of physicians in the online consultation market: a signaling theory perspective. Int. J. Electron. Commer. 23 (2), 244–271.
Lindenberg, S., Wittek, R., Giardini, F., 2020. Reputation effects, embeddedness, and granovetter's error. In: Buskens, V., Corten, R., Snijders, C. (Eds.), Advances in the Sociology of Trust and Cooperation: Theory, Experiments, and Field Studies. De Gruyter Oldenbourg, Berlin, pp. 113–140.
Lipsey, M.W., Wilson, D.B., 2001. Practical Meta-Analysis. Sage Publications, Inc, Thousand Oaks.
Liu, Y.W., Chen, H.P., Wie, G.J., Xu, J.L., 2007. Huicui fenxi: xinyong pingjia neng cujin wangshangpaimai ma [does reputation system impact online auction results: a meta-analysis]. Xinxi Xitong Xuebao 1 (1), 16–33.
Livingston, J.A., 2005. How valuable is a good reputation? A sample selection model of internet auctions. Rev. Econ. Stat. 87 (3), 453–465.
Macinnes, I., Li, Y., Yurcik, W., 2005. Reputation and dispute in eBay transactions. Int. J. Electron. Commer. 10 (1), 27–54.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., the PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. 151 (4), 264–269.
Palopoli, L., Rosaci, D., Sarné, G.M.L., 2013. A multi-tiered recommender system Architecture for supporting E-commerce. In: Fortino, G., Badica, C., Malgeri, M., Unland, R. (Eds.), Intelligent Distributed Computing VI. Studies in Computational Intelligence, vol. 446. Springer, Berlin, pp. 71–81.
Palopoli, L., Rosaci, D., Sarné, G.M.L., 2016. A distributed and multi-tiered software architecture for assessing e-Commerce recommendations. Concurrency Comput. Pract. Ex. 28 (18), 4507–4531.
Peterson, R.A., Brown, S.P., 2005. On the use of beta coefficients in meta-analysis. J. Appl. Psychol. 90 (1), 175–181.
Przepiorka, W., 2013. Buyers pay for and sellers invest in a good reputation: more evidence from eBay. J. Soc. Econ. 42, 31–42.
Przepiorka, W., Berger, J., 2017. Signaling theory evolving: signals and signs of trustworthiness in social exchange. In: Jann, B., Przepiorka, W. (Eds.), Social Dilemmas, Institutions and the Evolution of Cooperation. De Gruyter Oldenbourg, Berlin, pp. 373–392.
Raub, W., 2004. Hostage Posting as a mechanism of trust: binding, compensation, and signaling. Ration. Soc. 16 (3), 319–365.
Resnick, P., Zeckhauser, R., Friedman, E., Kuwabara, K., 2000. Reputation systems. Commun. ACM 43 (12), 45–48.
Resnick, P., Zeckhauser, R., Swanson, J., Lockwood, K., 2006. The value of reputation on eBay: a controlled experiment. Exp. Econ. 9 (2), 79–101.
Ricci, F., Rokach, L., Shapira, B. (Eds.), 2015. Recommender Systems Handbook. Springer, New York.
Rifkin, J., 2014. The Zero Marginal Cost Society: the Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism. Palgrave MacMillan, Basingstoke.
Rosenthal, R., 1991. Meta-Analytic Procedures for Social Research. SAGE Publications Inc, Newbury Park.
Schlägel, C., 2011. Country-Specific Effects of Reputation: A Cross-Country Comparison of Online Auction Markets. Springer Fachmedien Wiesbaden, Weisbaden.
Schneck, A., 2017. Examining publication bias—a simulation-based evaluation of statistical tests on publication bias. PeerJ 5, e4115.
Sears, J.M., 2016. A Reputation for the Good Stuff: User Feedback Signaling and the Deep Web Market Silk Road (Doctoral Dissertation). Montana State University-Bozeman, USA.
Shapiro, C., 1983. Premiums for high quality products as returns to reputations. Q. J. Econ. 98 (4), 659–679.
Sirmans, G.S., Macdonald, L., Macpherson, D.A., Zietz, E.N., 2006. The value of housing characteristics: a meta-analysis. J. R. Estate Finance Econ. 33, 215–240.
Snijders, C., Matzat, U., 2019. Online reputation systems. In: Giardini, F., Wittek, R. (Eds.), The Oxford Handbook of Gossip and Reputation. Oxford University Press, New York, p. 479.
Snijders, C., Weesie, J., 2009. Online programming markets. In: Cook, K.S., Snijders, C., Buskens, V. (Eds.), eTrust: Forming Relationships in the Online World. Russell Sage, New York, pp. 166–185.
Snijders, C., Zijdeman, R., 2004. Reputation and internet auctions: eBay and beyond. Analyse & Kritik 26 (1), 158–184.
Snijders, T.A.B., Bosker, R.J., 2012. Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling. Sage, Los Angeles.
Song, J., Baker, J., 2007. An integrated model exploring sellers' strategies in eBay auctions. Electron. Commer. Res. 7 (2), 165–187.
Standifird, S.S., 2001. Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings. J. Manag. 27 (3), 279–295.
StataCorp, L.P., 2015. Stata Treatment-Effects Reference Manual. A Stata Press Publication, College Station, TX.
Sterne, J.A.C., Sutton, A.J., Ioannidis, J.P.A., Terrin, N., Jones, D.R., Lau, J., Higgins, J.P.T., 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 342, d4002.
Suurmond, R., van Rhee, H., Hak, T., 2017. Introduction, comparison, and validation of Meta-Essentials: a free and simple tool for meta-analysis. Res. Synth. Methods 8 (4), 537–553.
Tadelis, S., 2016. Reputation and feedback systems in online platform markets. Ann. Rev. Econ. 8, 321–340.
To, P.L., Liao, C., Liu, Y.P., Chen, C.Y., 2008. Online auction effectiveness: optimal selling strategies for online auction market. Pacific Asia Conf. Inf. Syst. 2008 Proceed. 123.
Tong, G., Guo, G., 2019. Meta-analysis in sociological research: power and heterogeneity. Socio. Methods Res. 11, 1–39.
Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Software 36 (3), 1–48.
Xiao, J.J., Liu, L., 2009. Xiaofeizhe baozhang jihua de youxiaoxing yanjiu-Ji yu C2C wangshangjiaoyi de shizhengfenxi [Research on the effectiveness of buyers' protection policy- an empirical study of C2C online markets]. Caimao Jingji 11, 112–119.
Zhu, Y., Li, Y., Leboulanqer, M., 2009. National and cultural differences in the C2C electronic marketplace: an investigation into transactional behaviors of Chinese, American, and French consumers on eBay. Tsinghua Sci. Technol. 14 (3), 383–389.