Raising the Bar: Certification Thresholds and Market Outcomes[*]

Xiang Hui[†] Washington University

Maryam Saeedi[‡] Carnegie Mellon University

Giancarlo Spagnolo[§] SITE, Tor Vergata, EIEF & CEPR

Steve Tadelis[¶] UC Berkeley, NBER & CEPR

September 21, 2021

Abstract

Certification of sellers by trusted third parties helps alleviate information asymmetries in markets, yet little is known about the impact of a certification's threshold on market outcomes. Exploiting a policy change on eBay, we study how a more selective certification threshold affects the distribution of quality and incumbent behavior. We develop a stylized model that shows how changes in selectivity change the distribution of quality and prices in markets. Using rich data from hundreds of online categories on eBay.com, we find support for the model's hypotheses. Our results help inform the design of certification selectivity in electronic and other markets.
JEL Codes D47, D82, L15, L86

[*]We thank David Byrne, Sven Feldmann, Kate Ho, Hugo Hopenhayn, Tobias Klein, Greg Lewis, Ryan McDevitt, Brian McManus, Peter Newberry, Rob Porter, David Ronayne, Konrad Stahl, and many seminar and conference participants for excellent suggestions, discussions and comments. We are grateful to eBay for providing access to the data and to several eBay executives for providing valuable input. An earlier version of this paper was circulated under the title "Certification, Reputation and Entry: An Empirical Analysis".

[†]hui@wustl.edu

[‡]msaeedi@andrew.cmu.edu

[§]spagnologianca@gmail.com

[¶]stadelis@berkeley.edu

1 Introduction

Various institutions have emerged to help mitigate frictions caused by asymmetric information, including warranties (Grossman (1981)), reliance on past reputation (Shapiro (1983)), and regulated certification by a trusted intermediary (Leland (1979)). In their extensive survey of quality disclosure and certification, Dranove and Jin (2010) observe that for many important purchases, whether for consumption goods, durable goods, services, healthcare, or schooling, "from cradle to grave, consumers rely on quality disclosure to make important purchases."

A variety of third-party agencies issue quality certifications, including government agencies (e.g., trade licensing), for-profit rating agencies (e.g., credit ratings), independent NGOs (e.g., green certification), producers' associations (e.g., sustainable agriculture), and online platforms (e.g., seller quality), to name a few. Like strong brand reputations, certification by a trusted intermediary is often based on past performance and reduces asymmetric information. Furthermore, both strong brand reputations and trusted certification can become barriers to entry for sellers who do not have a certifiable track record (Klein and Leffler (1981), Grossman and Horn (1988)). It seems, therefore, that changes in certification criteria will impact the perceived quality of sellers both with and without certification and, in turn, the resulting market structure mix of incumbents and entrants. A number of complex questions emerge: How would more stringent certification criteria impact the type of sellers who enter the market and the incentives they face? How would it change the quality distribution of sellers and the prevailing prices in the market? And, under what conditions does the quality change associated with higher standards lead to higher social welfare?

In this paper we take a step towards answering these questions with an eye towards helping inform regulators and market designers on how to set their certification bar. We begin by developing a parsimonious asymmetric information model of a marketplace in which quality is endogenous and certification affects entry, behavior, and market structure. Using the model's testable hypotheses, we exploit a policy change that occurred in 2009 when eBay, one of the largest online marketplaces, replaced the "Powerseller" badge awarded to particularly virtuous sellers with the "eTRS" badge, which had more stringent requirements, hence "raising the bar" and becoming more selective.

The model shows that more stringent certification increases in the average quality of both badged (certified) and unbadged (uncertified) sellers. Sellers who lose their badge are worse than those who remain badged, but are better than those who were previously unbadged. As a consequence, when the certification bar is raised, entry is encouraged at the tails of the quality distribution, while discouraged in its center. That is, potential entrants with the highest quality benefit from the more selective badge and those with the lowest quality benefit from being pooled with better unbadged sellers. Entry becomes less attractive for mid-range quality sellers for whom it is harder to obtain a badge. Hence, our first main testable hypothesis predicts that changing the certification stringency will increase the dispersion of quality in the market. A second, more nuanced result shows that markets that are more impacted by the increased stringency of the badge will display a more dispersed distribution of entrant quality. Finally, the model also shows that only marginal mid-range quality sellers who can increase their quality at a low cost will exert higher effort to profitably obtain the more selective badge, while others will join the ranks of unbadged sellers.

We take these predictions to the data with an identification strategy that exploits the differential impact of the policy change across 400 separate subcategories (markets) on eBay's marketplace. Through the lens of our model, we assume that the composition of seller quality-types drives the differential impact across markets because the policy change itself was identical in all markets. This leads to heterogeneous effects of the policy on the fraction of badged sellers who lose their badge after the policy change. Indeed, we document a significant drop in the share of badged sellers at the policy change date, which is what the policy change was designed to do, and show that there is substantial heterogeneity of this effect across subcategories.

Using a verified measure of quality we find that the distribution of the entrants' quality indeed exhibits "fatter tails" after the policy change, consistent with our theoretical hypothesis. That is, the average quality of entrants increases in the upper deciles and drops in the bottom deciles of the quality distribution. Furthermore, fatter tails are more pronounced in markets that were more affected by the policy change, as predicted by our model.

To test our model's prediction that only marginal mid-range quality sellers who can increase their quality at a low cost will exert higher effort to obtain the more selective badge, we study the evolution of quality provided by four exclusive groups of *incumbent* sellers, depending on whether or not they had a badge before and after the change in policy. Consistent with our model, the only incumbents that show a significant change in behavior are those who lose their badge and, by improving quality provision, manage to regain the new badge within three months.

We then study how prices change for these four groups of incumbent sellers—with and without a badge after the policy change. The results confirm our model's predictions: First, sellers who lose their badge experience a decrease in the relative price that they receive. Second, sellers who remain badged and those who remain unbadged experience higher prices. Third, these changes are more noticeable in markets more affected by the policy change.

To conclude our analysis, we compute a back-of-the-envelope measure of consumer surplus to assess the impact of the policy change. We find that on average, consumer welfare increases by 2.2%. However, our estimates differ across different categories. Using machine learning techniques we shed light on factors that correlate with higher welfare for consumers as a result of the policy change. We find that higher gains in consumer surplus happen in markets with a higher share of consumer complaints per transaction. We interpret this as suggestive evidence that in markets for which consumers have higher preferences for quality, given by their higher tendency to file a complaint, they benefit more from more stringent certification requirements.

An important identifying assumption is that there are no time-varying heterogeneities across subcategories that simultaneously affect changes in the share of badged sellers and in entry. We perform placebo tests and find no impact, consistent with the exclusion restriction of our econometric specification. We also control for observable and time-varying variables for robustness. The results are all qualitatively similar to those in our main specification.

Our results help guide the design of certification mechanisms in electronic markets, where a host of performance measures can be used to set certification requirements and increase buyers' trust in the marketplace. They may also offer useful insights for other markets with high levels of asymmetric information where certification is ubiquitous. These markets include financial markets where credit ratings are used to obtain the "investment-grade" badge, many final and intermediate goods markets where labelling institutions certify various forms of quality, and public procurement markets where regulatory certification can significantly change the competitive environment and reduce the costs of public services.[1] According to our findings, if a platform (or a large procurer, or buyer) is concerned about too much mass in the middle of the quality range, while there are two few high- and low-quality sellers, it should increase the stringency of the certifying badge to stimulate entry at the tails of the quality distribution (and vice versa). Furthermore, our results suggest that raising the certification bar is more likely to increase consumer welfare where more buyers' have a preferences for high quality and in industries where more sellers can adjust (or sellers can more easily adjust) the quality of their product in response to the policy change.

Our paper joins a growing literature that uses rich online data to understand how to alleviate

[1]For example, concerns have been expressed by several prominent U.S. senators, as well as in the EU, that the extensive use of past performance information for selecting federal contractors could hinder the ability of new or small businesses to enter public procurement markets. The debate led the General Accountability Office to study dozens of procurement decisions across multiple government agencies, but the resulting report (GAO-12-102R) was rather inconclusive (see further discussions in Butler et al. (2020)).

asymmetric information in markets. The closest papers to ours are Elfenbein et al. (2015), Klein et al. (2016), and Hui et al. (2018), which also use eBay data to study the effects of different information policies on market structure. Elfenbein et al. (2015) study the value of a certification badge across different markets and show that certification provides more value when the number of certified sellers is low and when markets are more competitive. However, they do not study the impact of certification on the dynamics of entry and changes in market structure. Klein et al. (2016) and Hui et al. (2018) exploit a different policy change on eBay after which sellers could no longer leave negative feedback for buyers, making it easier for buyers to leave negative feedback. Both studies find an improvement in buyers' experience after the policy change. Using scraped data, Klein et al. (2016) take advantage of the evolution of both public and anonymous feedback of Detailed Seller Ratings to show that the improvement in transaction quality is not due to exit from low-quality sellers. Using internal data from eBay, Hui et al. (2018) complement Klein et al. (2016) and investigate changes in the size of incumbents. They show that although low-quality sellers do not exit after the policy change, their size shrinks dramatically, accounting for 49%–77% of the quality improvement. In contrast with these three papers, our paper explicitly studies the impact of certification on the dynamics of entry and the changes in market structure, as well as the quality provided by entrants and incumbents before and after the policy change.

A related literature analyzes the effects of changes in eBay's feedback mechanisms on price and quality (e.g., Klein et al. (2016), Hui et al. (2016), and Nosko and Tadelis (2015)). Consistent with these papers, we find that sellers who were badged both before and after the policy change were of higher quality than sellers who were badged before but not after the change. Our paper also broadly relates to the literature that ties reputation, certification, and transparency to sales performance, including empirical studies such as Cabral and Hortacsu (2010), Hui et al. (2016), and Fan et al. (2016).[2] Last, our analyses are related to the empirical literature on adverse selection and moral hazard, e.g., Greenstone et al. (2006), Einav et al. (2013) and Bajari et al. (2014).

The remainder of the paper is organized as follows. Section 2 provides details about the platform and the policy change, while Section 3 presents a stylized theoretical model that illustrates how the policy change affects entry and quality choices. Section 4 describes our data, and Section 5 discusses our empirical strategy. Our results appear in Section 6, Section 7 deals with endogeneity concerns and offers several robustness tests, and Section 8 concludes the paper.

[2] See also Bajari and Hortacsu (2004), Cabral (2012), and Tadelis (2016) for surveys and Avery et al. (1999), Jullien and Park (2014), Stahl and Strausz (2017), and Hopenhayn and Saeedi (2019) for related theoretical studies.

2 Background and Policy Change

eBay is known for its well-studied feedback system in which sellers and buyers can rate one another with positive, negative, or neutral feedback. eBay later introduced "detailed seller ratings," (henceforth, DSR), in which buyers leave sellers anonymous ratings between 1 and 5 stars along four dimensions (item as described, communication, shipping rate, and shipping speed). In 2008, to combat concerns that seller retaliation deters buyers from leaving negative feedback, eBay made the feedback rating asymmetric so that sellers could leave only positive or no feedback for buyers.

In addition to user-generated feedback, eBay started certifying sellers it deemed to be of the highest-quality by awarding them the "Powerseller" badge. To qualify, a seller had to sell at least 100 items or at least $1,000 worth of items every month for three consecutive months.[3] The seller also had to maintain at least 98% positive feedback and 4.6 out of 5.0 DSR. Finally, a seller had to be registered with eBay for at least 90 days. The main benefit of being a Powerseller was receiving discounts on shipping fees of up to 35.6%. Though different levels of Powersellers depended on the number and value of annual sales, all Powersellers enjoyed the same direct benefits from eBay. An indirect benefit of the badge was its salience, suggesting that the seller is of higher quality.

eBay revised its certification requirements and introduced the "eBay Top Rated Seller" (eTRS) badge, which was announced in July 2009 and became effective in September 2009.[4] To qualify as eTRS, a seller must surpass the Powerseller status by *additionally* having at least 100 transactions and at least $3,000 in sales over the previous 12 months, and must have less than 0.5% or two transactions with low DSRs—1 or 2 stars out of 5—and less than 0.5% or two complaints from buyers.[5] The information on dispute rates, only available to eBay, has not been used before. It is also important to note that after eTRS's introduction, sellers can still obtain the Powerseller status but it is no longer displayed as a badge for buyers to observe.

[3] There were six tiers of Powerseller: Bronze, Silver, Gold, Platinum, Titanium and Diamond. To reach a higher tier, sellers must increase the volume of sales in the past three months. For example, to achieve the Silver tier, the minimum quantity and value must exceed 3,000 and $300, respectively. However, the quality requirements did not increase for sellers of higher tiers. Therefore, we focus on having the badge before, rather than on the tier of the badge the sellers had beforehand.

[4] If sellers changed their behavior between the announcement date and the implementation date, this would imply a smaller drop in the share of badged sellers and smaller changes in outcome variables, which likely attenuates our estimation results.

[5] A senior director involved in the change explained that there were two main reasons for the change: First, the Powerseller program rewarded sellers with higher discounts on their final value fees based on their sales volume, paying less attention to their performance, which created an incentive for sellers to sell more, sometimes at the cost of the experience they were delivering. Second, buyers perceived the Powerseller badge to mean eBay endorsed the seller. This skewed purchasing towards Powersellers, who already had a pricing advantage over non-Powersellers due to their discounts, but had little incentive to deliver great service. The eTRS badge introduced more stringent performance requirements to obtain discounts by using maximum thresholds of low DSRs and dispute rates.

Obtaining the eTRS badge is harder than obtaining the Powerseller badge, but also provides greater benefits. Top Rated Sellers receive a 20% discount on their final value fee (a percent of the transaction price) and have their listings positioned higher on eBay's "Best Match" search results page, which is the default sorting order, promoting more sales. Finally, the eTRS badge appears on listings, signaling the seller's superior quality to all potential buyers.

Two other simultaneous changes occurred on eBay with this policy.[6] One introduced easier selling procedures across all categories (e.g., faster processing of unpaid items, removal of negative feedback if a dispute is resolved, and easier management of buyer messages). The second is a change in the search ranking algorithm, mainly that (i) ranking became based on sales per impression instead of sales; (ii) the title's relevance was enhanced; and (iii) eTRS were promoted in the default search ranking algorithm. Changes (i) and (ii) are controlled for with our DiD approach because the causal effect of each of these changes is orthogonal to our sub-category exposure measures. For (iii), we include time-varying market characteristics in the regression and replicate the key results of our paper. For example, we control for the share of badged sellers in a market, because non-badged sellers appear less in the search results page in markets with a higher share of badged sellers due to the change in the search ranking algorithm. If our results are somehow mechanically driven by the share of badged sellers, then the estimates would be reduced after controlling for the share of badged sellers in a market. However, after controlling for these time-varying market characteristics, we do not see any qualitatively different results as presented in Section 7.2.[7]

3 Certification and Entry: A Simple Model

We present a simple model that incorporates both hidden information (adverse selection) and hidden action (moral hazard) in the spirit of Diamond (1989). The model generates comparative statics that offer a series of testable implications, and clarifies the assumptions needed to empirically identify the effect of a more stringent certification on market outcomes.

Supply: Consider a market with a continuum of sellers. Each seller can produce one unit of output with zero marginal costs and fixed costs $k \in [0, \infty)$, independently distributed with the continuous and strictly increasing cumulative distribution function $G(k)$, with $G(0) = 0$ and $G(\infty) = 1$. There are three types of sellers: a measure μ_ℓ of "low-quality" sellers, indexed by ℓ,

[6] https://pages.ebay.com/co/es-co/sell/July2009Update/faq/index.html#2-1 (accessed on 10/30/2018).

[7] Ideally we would want to control for the number of times a listing is shown to buyers in the search results page. However, these data only exist since 2011.

who can only produce low quality L; a measure μ_h of "high-quality" sellers, indexed by h, who can only produce high-quality H; and a measure μ_s of strategic sellers, indexed by s, who can each choose whether to exert extra effort at a cost e and produce high quality H, or whether to shirk at no cost and produce medium quality M, where $H > M > L > 0$. The cost of effort $e \in [0, \infty)$ is independently distributed across all s-type sellers with the continuous and strictly increasing cumulative distribution function $F(e)$, with $F(0) = 0$ and $F(\infty) = 1$. Hence, s-type sellers have two dimensions of cost heterogeneity, (k, e), while ℓ- and h-type sellers only differ across k.[8]

Demand: Each of a continuum of buyers demands one unit of a good and is willing to pay up to the expected quality of the good. To simplify, we assume that the buyers are on the "long side" of the market so that market clearing prices leave buyers with no surplus and the price of each good will be equal to its expected quality.

Information: Buyers cannot observe the quality of any given seller. A marketplace regulator can, however, produce an observable "badge" $B \in \{M, H\}$ that credibly signals if a seller's quality is at least at the threshold B. Given a badge B, let $\underline{v}_B$ denote the expected quality of sellers who are below the badge threshold and let $\bar{v}_B$ denote the expected quality of sellers who are at or above the badge threshold. Then, if a positive measure of sellers of all types are in the market, then $\bar{v}_H = H$ and $M > \underline{v}_H > L$, whereas $H > \bar{v}_M > M$ and $\underline{v}_M = L$.

Equilibrium: Let $\mu_{\theta B}$ denote the measure of type θ sellers that enter a market with badge B. Let π denote the fraction of active s-type sellers who choose to exert effort. An *equilibrium* for threshold $B \in \{M, L\}$ consists of (i) a pair of prices $\underline{p}_B$ and $\bar{p}_B$, (ii) measures of each type of sellers, $\mu_{\theta B}$, and (iii) the proportion of s-type sellers who enter and work, π, such that prices equal expected qualities, which in turn are consistent with Bayes rule given the measures of sellers of each type above and below the threshold, and that all sellers are best responding to prices.

We are interested in the comparative statics of making the badge more restrictive by switching from $B = M$ to $B = H$ so that higher-quality is needed to obtain a badge.

3.1 Lax Badge: $B = M$

Because $B = M$, all s-types qualify to be badged whether they choose to exert effort or not. Since prices depend only on the badge, there are no returns to effort while the cost of effort is positive

[8] We can alternatively assume that strategic types who do not exert effort will have a baseline quality L instead of M. They can then increase their quality to M by paying the cost e, or increase their quality to H by paying a higher cost $e' > e$. Results remain mostly the same except for the prediction that prices increase for unbadged sellers. In this case the price remains the same for unbadged sellers before and after the policy change.

for all s-types. The following observation is straightforward:

Lemma 1. *All s-types choose to shirk when $B = M$.*

The equilibrium when $B = M$ is therefore characterized as follows:

1. **prices:** $\bar{p}_M = \bar{v}_M = \frac{\mu_s M + \mu_h H}{\mu_s + \mu_h}$, and $\underline{p}_M = \underline{v}_M = L$,
2. **entry:** $\mu_{\ell M} = G(L)\mu_\ell$, $\mu_{sM} = G(\bar{v}_M)\mu_s$ and $\mu_{hM} = G(\bar{v}_M)\mu_h$,
3. **behavior:** All s-types who enter choose to shirk.

That is, $\bar{p}_M$ is equal to the expected quality given the weights of the s- and h-types in the population because $G(\cdot)$ is i.i.d. across all types, both s- and h-types receive the same price, and have the same zero-profit condition. The measure of sellers who enter are determined by those who can cover their fixed costs given the two equilibrium prices.

3.2 Stringent Badge: $B = H$

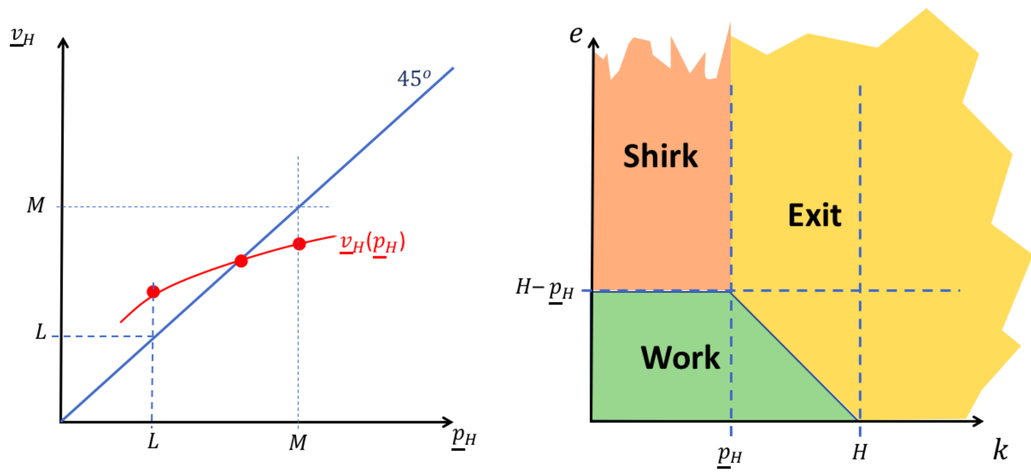
When $B = H$, s-type sellers will be badged if and only if they choose to exert effort.

Lemma 2. $1 > \pi > 0$ *in any equilibrium with $B = H$.*

Proof. Because $G(\cdot)$ is continuous and increasing on $[0, \infty)$, and $G(0) = 0$, a positive measure of all types will enter the market. This implies that in any equilibrium with $B = H$, $\bar{v}_H = H$ and $M > \underline{v}_H > L$, resulting in $\bar{p}_H > \underline{p}_H$. In turn, because both $F(\cdot)$ and $G(\cdot)$ are continuous and increasing on $[0, \infty)$, and $G(0) = F(0) = 0$, it follows that a positive measure of s-types will prefer to enter, exert effort and be badged over not being badged. Finally, because the support of $F(\cdot)$ is unbounded, and because $\bar{p}_H - \underline{p}_H$ is bounded, then not all s-types who enter will exert effort. □

To characterize the equilibrium when $B = H$, it is illustrative to graphically describe the structure of any equilibrium as shown in Figure 1. The right panel shows the two-dimensional cost-space of s-type sellers who have both entry costs k and effort costs e. Because $\bar{p}_H = H$, any s-type with $k > H$ cannot earn positive profits and will exit. Similarly, any s-type with $k < \underline{p}_H$ can enter and earn $\underline{p}_H - k > 0$ by shirking. For these entrants, the benefit from working outweighs the cost of working if and only if $H - \underline{p}_H > e$. Finally, for those with fixed costs $H > k > \underline{p}_H$, if $k + e < H$ then they prefer to enter and work over exit (shirking yields negative profits), while if $k + e > H$ then they prefer to exit. This observation helps characterize equilibrium as follows:

Figure 1: Equilibrium when $B = H$



Proposition 1. *When $B = H$ there exists an equilibrium with $\bar{p}_H = H$ and $M > \underline{p}_H > L$.*

Proof. Market prices determine entry and each s-type's choice to work. By construction, $\bar{p}_H = H$. Consider the lowest possible unbadged price, $\underline{p}_H = L$. Because $L > 0$, a proportion $G(L)$ of ℓ- and s-types with fixed costs $k < L$ will enter, of which a proportion $\pi = F(H - L)$ of s-types will work and obtain a badge, and the remainder will shirk and produce quality M. But because a positive measure $G(L)(1 - F(H - L))$ of s-types enter and are unbadged, it follows that $\underline{v}_H > \underline{p}_H = L$, so this cannot be an equilibrium. Define $\underline{v}_H(\underline{p}_H)$ as the unbadged quality that would be obtained following an unbadged price $\underline{p}_H$ and in which all sellers act optimally. We can explicitly write the function $\underline{v}_H(\underline{p}_H)$ for any $M > \underline{p}_H > L$ as follows:

$$\underline{v}_H(\underline{p}_H) = \frac{\mu_\ell G(\underline{p}_H)L + \mu_s G(\underline{p}_H)(1 - F(H - \underline{p}_H))M}{\mu_\ell G(\underline{p}_H) + \mu_s G(\underline{p}_H)(1 - F(H - \underline{p}_H))}$$

As established above, $\underline{v}_H(L) > L$, and $\underline{v}_H(M) < M$ because both shirking s-types and ℓ-types will enter and be unbadged. Because both $G(\cdot)$ and $F(\cdot)$ are continuous, the function $\underline{v}_H(\underline{p}_H)$ is continuous, and must cross the 45-degrees line at least once. Hence, an equilibrium exists. □

The left panel of Figure 1 illustrates the logic of Proposition 1. The upshot from the description of equilibria above is that any equilibrium $B = H$ will satisfy the following:[9]

1. **prices:** $\bar{p}_H = \bar{v}_H = H$, and $\underline{p}_H = \underline{v}_H \in (L, M)$,

[9] The double integral represents the s-types who enter with $\underline{p}_H < k < H$ and for whom $e + k < H$ so they prefer to enter and work over exiting or entering and shirking.

2. **entry:** $\mu_{\ell H} = G(\underline{v}_H)\mu_\ell$, $\mu_{sH} = G(\underline{v}_H)\mu_s + \int_{\underline{p}_H}^{H}\int_{H-\underline{p}_H}^{H-k} dG(x)dF(y)dxdy$, and $\mu_{hH} = G(H)\mu_H$,

3. **behavior:** Some s-types who enter choose to work and some to shirk. The measure of s-types who shirk is $G(\underline{v}_H)(1 - F(H - \underline{p}_H))\mu_s$.

Note that there may potentially be more than one equilibrium, and conditions on $G(\cdot)$ and $F(\cdot)$ can be described to guarantee uniqueness, yet this is not a concern given our interest in comparing any equilibrium with $B = H$ to the unique equilibrium with $B = M$.

3.3 Comparative Statics

The shapes of $G(\cdot)$ and $F(\cdot)$ will determine whether a more stringent badge will cause aggregate entry–and entrant quality–to either increase or decrease. However, the following five corollaries follow immediately from comparing prices across the two equilibria identified above and lead to testable empirical predictions.

Corollary 1. $\underline{p}_H < \bar{p}_M$.

Hence, s-types who lose their badge are hurt by a lower price, implying that those with high enough entry and effort costs will not enter after the change.

Corollary 2. $\underline{p}_H > \underline{p}_M$ *and* $\bar{p}_H > \bar{p}_M$.

$\bar{p}_H > \bar{p}_M$ follows from the definition of a more stringent badge, and $\underline{p}_H > \underline{p}_M$ because unbadged sellers now include both qualities L and M, which leads to a higher average quality than just L.

Corollary 3. *Entry increases for ℓ and h-types and decreases for s-types.*

This Corollary follows directly from Corollaries 1 and 2, which together imply that the distribution of entrants will have "fatter tails" after the more stringent badge is implemented.

Corollary 4. *s-types who retain their badge will increase quality and produce H instead of M.*

This follows because all badged s-types shirk when $B = M$ while they must work when $B = H$.

Corollary 5. *Let market A have measure μ_s^A and let market B have measure $\mu_s^B > \mu_s^A$, fixing the other measures of ℓ- and h-types across the markets. If both markets experience a change of badge from lax to stringent, then more entry of h-types will occur in market B.*

This result follows from the fact that, fixing the measure of h-types, an increase in s-types means a lower price $\bar{p}_M$ in market B. This in turn implies that when the badge becomes stringent, and $\bar{p}_H = H$ in both markets, then the badged-price increases more in market B, and hence there will be more entry of both h-types, as well as s-types who choose to work.

This last corollary is critical in generating the main comparative static that guides our empirical analysis. Naturally, when there are more s-types in a market, then more sellers will necessarily lose their badge after an increase in stringency. Hence, if a policy change occurs, then one can *infer* that market B had a higher measure of s-types than market A if a larger fraction of sellers lose their badge in market B. Hence, if a policy change is implemented simultaneously in many markets, then corollary 4 implies that in those markets that lost a higher fraction of badged sellers, the impact on the tails of the distribution of entry will be larger, an insight we take to our data.

It is worth noting that instead of only three discrete quality levels and two badge levels we explored a more elaborate model with a continuum of baseline quality-types where each type can increase quality by exerting effort, and the badge can be set at any level of quality in the interior of the type-range. The results are similar though the analysis is more involved and distracts from the key economic forces at play. Namely, by increasing the selectivity of the badge, the distribution of types above and below the badge changes in ways that increase average quality for both groups, and this in turn impacts incentives to work as well as incentives to enter. The heterogeneity across markets for Corollary 4 would result from heterogenous distributions of types in a similar manner. The model we use is, in our view, the most parsimonious and easy to follow.

We now summarize the main empirical predictions of our model that we take to the data: (i) The quality distribution of entrant sellers exhibits fatter tails; (ii) across markets, in a market with more s-types, there will be a larger impact on the tails of the quality distribution of entrants;[10] (iii) Incumbents who would lose their badge but instead retain it must increase their quality; and (iv) Perceived quality and prices increase for both badged and unbadged sellers.

4 Data

We use proprietary data from eBay that include detailed characteristics on product attributes, listing features, buyer history, and seller feedback. Our data cover the period from October 2008 to September 2010, and include all listing and transaction data in the year before and the year

[10] The mirror image of (i) and (ii) is that the quality distribution of those who exit has thinner tails. We show some evidence of this in the online Appendix.

after the policy change. An important feature of our data is information on product subcategories cataloged by eBay. There are about 400 subcategories (which we also call markets), such as Lamps and Lighting, Beads and Jewelry Making, Video Game Memorabilia, Digital Cameras, and others. A subcategory is the finest level of eBay's catalog that includes all listings on the site.

Though it is hard to observe a firm's entry date before it has made a sale or reached a certain size, in our detailed data we observe a seller's first listing in different subcategories on eBay. We treat this date as a seller's entry date into the subcategory. Additionally, we observe the number of incumbents in any month in each subcategory. This allows us to compute a normalized number of entrants across subcategories, which we call the entrant ratio.

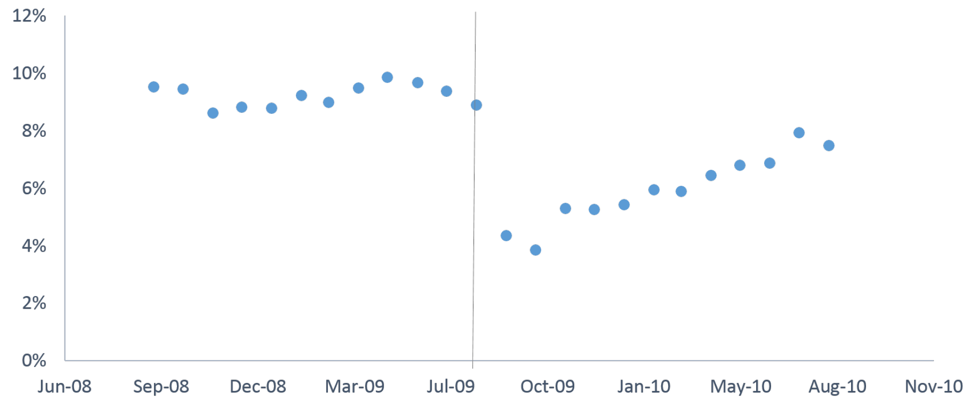
Finally, the use of internal data allows us to construct a quality measure that is not observed publicly. Every seller has a reputation score and percent-positive (PP) on eBay, the latter being the number of positive ratings divided by the total number of ratings. Nosko and Tadelis (2015) demonstrate the extreme skewness of PP (the mean is 99.3% and the median is 100%), a finding consistent with those of others who documented biases in reviews (Zervas et al., 2015; Luca, 2011; Fradkin et al., 2017). Nosko and Tadelis (2015) construct a measure they call "effective percentage positive" (EPP), which is the number of positive feedback transactions divided by the number of total transactions, and show that EPP contains much more information on a seller's quality than conventional feedback and reputation scores. We follow their approach to compute each seller's EPP and use it as a measure of quality. We construct a seller's EPP using the number of transactions and the number of positive feedback ratings in the first year of entry, conditional on the entrant's survival in the second year (selling at least one item in both the first and second years after entry). The conditioning is intended to eliminate the survival effect from the quality effect.

We also consider alternative measures of quality in place of EPP: PP, low DSRs, and the number of claims filed against sellers. As we report in the online appendix, all the signs are consistent with our main specification, though some of the regressions are not significant at the 95% confidence level. Also, variations of EPP with different time intervals and without conditioning on sellers' survival yield similar results, as reported in the online appendix.

5 Empirical Strategy

The policy change described in Section 2 offers a quasi-experiment, and Figure 2 demonstrates that this change caused a significant decrease in the share and number of badged sellers. The average

Figure 2: Share of Badged Sellers



Notes: Average monthly share of badged sellers on eBay. The vertical line indicates the policy change, which made it harder for a seller to obtain a badge. All the averages are statistically different from each other at the 1% level.

share of badged sellers dropped from around 10% in the year before the change to about 4% right after the change. Given that badged sellers on eBay account for roughly half of all sales in the marketplace, the drop in the share of badged sellers creates a big change in the share of badged listings in the search results page. Sellers could lose their badge if they do not meet the new quality or sales requirements, and among sellers who lost their badge, 98% did so because they did not meet the quality requirements; therefore, we concentrate on change in quality requirements.

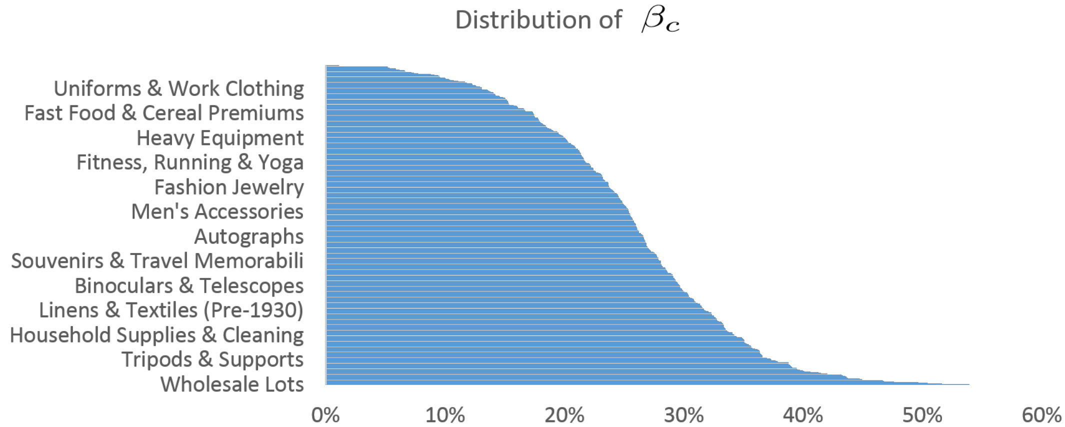
We take advantage of the fact that a "one-size-fits-all" policy change was implemented across heterogeneous markets, each having its own distribution of sellers, as modeled in Section 3. Our goal is to create treatment and control groups using variations in policy exposure across different markets on eBay. Consider two such markets; after the policy change, one market loses a larger fraction of its badged sellers than the other. Through the lens of the model and Corollary 4, variation in the intensity of the number of sellers who lose their badge is an indication of the number of s-type sellers. It follows that a market with a larger drop in the share of badged sellers should exhibit a larger change in outcome variables. We assume that this variation is exogenous to other aspects of a market aside from the distribution of types and test this assumption with different measures of policy exposure in the online appendix as well as by using a placebo test.[11]

To measure the policy exposure across markets, we first use the new criteria of a badge to simulate the percentage drop in the share of badged sellers. In particular, we apply the new certification requirements on badged sellers in the month before the policy change and compute the drop in the number of badged sellers divided by the total number of badged sellers.[12]

[11] A similar approach is used in Mian and Sufi (2012).

[12] We establish the robustness of our results by using other measures of policy exposure and report the results

Figure 3: Policy Exposure in Different Subcategories



Notes: Policy exposure is the percentage drop in badged sellers caused by the policy change in different subcategories on eBay. There are about 400 subcategories; the labels on the left axis are some examples.

The horizontal bars in Figure 3 are the ex-ante simulated percentage drop in the share of badged sellers across markets, which measures policy exposure. The decrease in the share of badged sellers varies dramatically across markets, from less than 10% to as much as 50%.

Our main identification strategy exploits the variability in policy exposure in different markets induced by the policy change using a continuous difference-in-difference (DiD) approach. In particular, we estimate the impact of the policy change by comparing the changes in the number and quality of entrants in markets that are *more* affected by the policy change to those in markets that are *less* affected over the same time period. This DiD approach is continuous in the sense that the treatments (i.e., impacts of the change on the share of badged sellers across markets) take continuous values between 0 and 1. The DiD specification is given as

$$Y_{ct} = \gamma \widehat{\beta}_c Policy_t + \mu_c + \xi_t + \epsilon_{ct}, \tag{1}$$

where Y_{ct}'s are the outcome variables of interest in subcategory c in month t (e.g., quality, or entry); $\widehat{\beta}_c$ is the simulated impact of the policy change on the share of badged sellers from our first stage, shown in Figure 3; $Policy_t$ is a dummy variable that equals 1 after the policy change; μ_c are

in the online appendix. In particular, we (1) apply an immediate change in the share of badged sellers using data from the week before and the week after the policy change; (2) estimate the change using an event study in the one, three, and six months before the policy change; (3) use the drop in the number of badged sellers instead of shares; and (4) use the percentiles of measures of policy exposure across subcategories. Our preferred measure is based on the simulation approach because it is an ex-ante measure of the policy exposure. In particular, in the event-study approaches, the change is estimated based on the share of badged sellers after the policy change, which itself may endogenously depend on changes in entry due to the policy change.

Table 1: Policy Change Impact on Rate and Quality of Entrants

Panel A. Entrant Ratio			
	(1)	(2)	(3)
	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	0.124***	0.066***	0.010
	(0.021)	(0.016)	(0.032)
R^2	0.911	0.888	0.685
Panel B. EPP Conditional on Survival in the Second Year			
	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	0.064***	0.039***	0.043***
	(0.019)	(0.014)	(0.016)
R^2	0.771	0.728	0.699

Notes: The regressions are performed at the subcategory-month level.
*** indicates significance at p $\leq$ 0.01; ** p $\leq$ 0.05; * p $\leq$ 0.10.

subcategory fixed effects; ξ_t are month fixed effects; and ϵ_{ct} are error terms. We cluster standard errors at the subcategory level in the estimation.

Our coefficient of interest is γ, which indicates the percentage change in the outcome variable as a result of variations in the share of badged sellers due to the policy change. Specifically, a statistically significant positive $\widehat{\gamma}$ means that a larger *decrease* in the share of badged sellers *increases* the outcome variable. Possible endogeneity issues are addressed in Section 7.

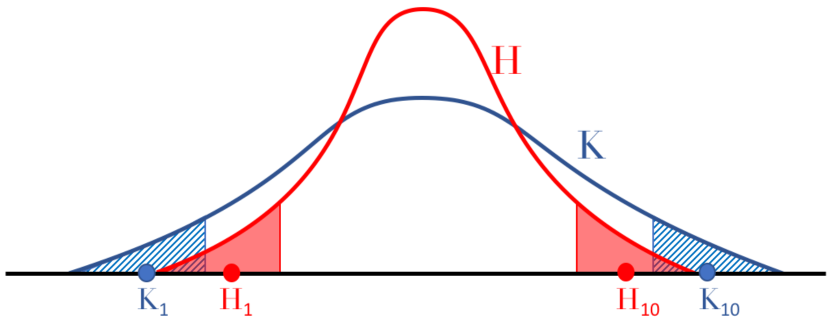
The DiD approach controls for time-invariant differences in the variables of interest across subcategories; for example, the entrant ratio in the Clothing market is higher than that in the Antiques market. The approach also controls for differences in the entrant ratio over time, for example, changes in the overall popularity of selling on eBay over time. As in most DiD approaches, our key identification assumption for a causal interpretation of $\widehat{\gamma}$ is that time-varying unobserved errors do not systematically correlate with $\widehat{\beta_c}$ and Y_{ct} simultaneously. We provide robustness tests of this identification assumption in Section 7.

6 Results

We first present some descriptive statistics of the effects of the policy change on the average rates of entry and quality provided by the entrants, followed by empirical tests of our model's predictions.

Our model does not predict whether the number of entrants, or their average quality, increases or decreases (these depend on the distributions of types and entry costs). Table 1 reports $\widehat{\gamma}$ from regression (1) for the entry rate and quality of entrants. Recall that a positive γ means that the increase in the outcome variable is larger in more impacted markets, i.e., a larger drop in the

Figure 4: Example of a Comparison of Two Distributions



share of badged sellers. In Panel A of Table 1, column 1 shows that the entrant ratio is higher in markets that are more affected by the policy change, using data from three months before and after the policy change (June 20–September 19 and September 20–December 19, 2008). A 10% larger decrease in the share of badged sellers leads to 1.2% more entrants. The estimate in column 2 is smaller when we use data from six months before and after the policy change. In column 3, we study the impact seven to twelve months *after* the policy change (relative to the six months before the policy change), where the estimate is even smaller and is not statistically significant.[13]

The positive coefficients in panel B in Table 1 show that there is an increase in the average quality (EPP) of entrants in the more affected subcategories after the policy change. Column 3 shows that the increase in EPP remains significant from the seventh to the twelfth month after the policy change, suggesting that its impact on entrants' quality is persistent.

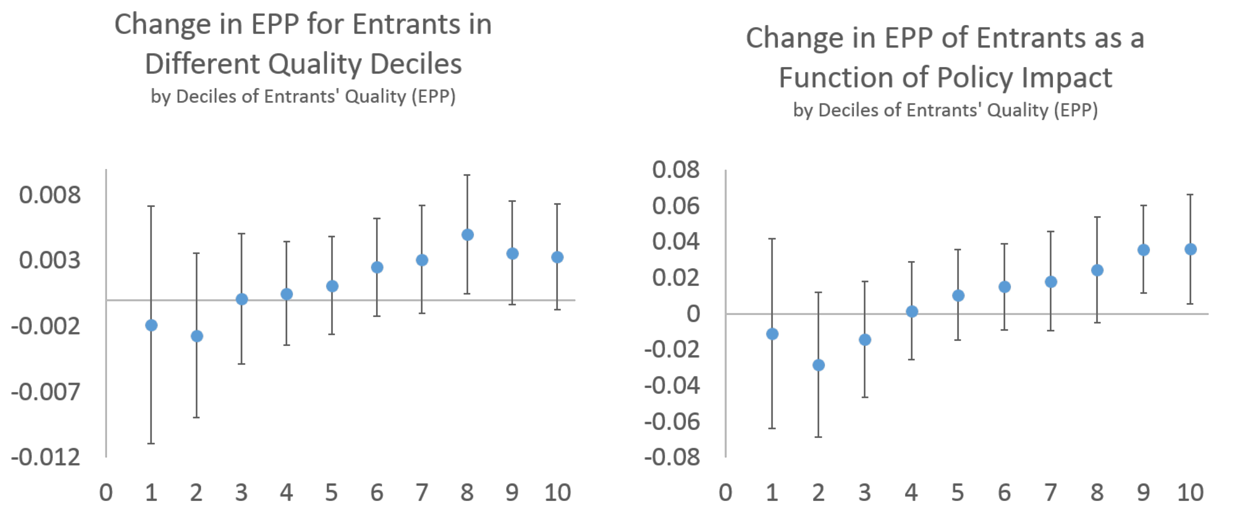
6.1 Quality Distribution of Entrants: Fatter Tails

Predictions (i) and (ii) at the end of Section 3 relate to fatter tails, which correspond to *within*-market fatter tails and *across*-market fatter tails respectively. The intuition for testing these is shown in Figure 4. Consider two distributions of entrants' EPP scores in the first year after these sellers' entry, H and K, the latter having fatter tails. Begin by partitioning entrants into deciles based on their EPP scores. Denote the average quality of the top decile of H by H_{10} and of K by K_{10}, and, similarly, denote the average quality of the bottom decile of H by H_1 and of K by K_1. Since K has fatter tails, it follows that $H_{10} < K_{10}$ and $H_1 > K_1$. These differences will be smaller for less extreme deciles and will all but disappear for the middle deciles.

To test for within-market changes in the distribution we rely on an event-study approach to

[13]We do not include longer time periods because eBay introduced eBay Buyer Protection in September 2010.

Figure 5: Change in EPP for Entrants in Different Quality Deciles



Notes: The left figure shows average within-subcategory changes in EPP. The right figure shows across-subcategory changes in EPP as a function of policy exposure. Bars indicate 95% confidence intervals.

estimate the effect of the policy change on EPP for each market, while to test for across-market distributional changes, we perform our DiD specification for different deciles, as we explained in more detail next. For both specifications, a positive coefficient for the top deciles indicates that the average entrant quality is higher after the policy change, and that the average entrant quality is higher after the policy change in markets with higher policy exposure, respectively. Similarly, a negative estimate for the bottom deciles will confirm the hypotheses for the bottom tail.

Figure 5 plots point estimates with 95% confidence intervals from running regression (1) on entrants' first-year EPP for different EPP deciles, with 10 being the highest decile and 1 being the lowest. For consistency, we condition the EPP calculation on an entrant's survival in the second year. Entrants are counted every two months and we restrict attention to markets with at least 10 entrants in each decile. Hence, for each market we have three observations (six-month equivalent) before the policy change and three observations after. Additionally, we only consider markets that have entry in all of the six two-month periods, leaving 228 of the 400 eBay subcategories.[14]

The left panel shows that the distribution of entrant quality exhibits fatter tails *within* each market. For each quality decile of a market, we estimate how the policy change impacts the EPP of entrants in an event-study manner (i.e., regressing EPP on a constant, policy dummy, and linear bi-monthly trend). For each quality decile, we plot the points calculated by averaging these estimates

[14]Performing the analysis on all subcategories preserves the monotonically increasing estimates as a function of quality deciles that we find, but the results are not significant. This is likely due to the noise induced by having too few entrants in the deciles of some markets.

across markets. The confidence intervals are constructed based on the standard errors of these estimates. In the right panel, we test whether the distribution of entrant quality exhibits fatter tails *across* markets. For each quality decile of a market, we perform the DiD estimation across markets, and the plotted points are the estimated γ in specification 1 with their 95% confidence intervals. In both figures, the top-two decile point estimates are positive. They are statistically significant at the 90% level in the left panel, and at the 95% level in the right panel, as predicted by the theory. The other estimates exhibit an overall increasing relationship that is consistent with our model's fatter-tail predictions, but no two are statistically different.[15] This in turn implies that sellers in the middle of the quality distribution enter less frequently.[16,17]

6.2 Incumbent Behavior: Some Higher Effort

Empirical prediction (iii) stated that some incumbent sellers who retain their badge after the policy change must have exerted some investment or effort to increase their quality. We define incumbents as sellers who listed at least one item both before and after the policy change. Incumbents' EPPs are computed using transactions in a *given month* to capture potential changes in behavior and quality of service from month to month.

To directly test our third empirical prediction, we divide incumbents into four mutually exclusive and collectively exhaustive groups based on their certification status before and after the policy change. The BB group consists of sellers who were badged both before and after the policy change; BN consists of sellers who were badged before the policy change but lost their badge after; We similarly label groups NB and NN.[18] We consider a seller to be badged before the policy change if she was badged for at least five out of six months before the change.[19] The seller's badge status afterwards depends on whether she meets the new policy requirements by the end of the day before the policy change. In other words, a seller's badge status before the policy change is based on the actual measure, and her status after the change is based on simulation, to be consistent with the construction of our policy exposure measure. The largest group is the NN group, with over 50% of sellers, while the smallest is the NB group, at 4%.

[15] Note that the estimates from the event-study approach are an order of magnitude smaller than the ones from the DiD approach, probably because the DiD approach can better control for common time trends across markets.

[16] Repeating the analyses by dividing entrants into three bins and five bins yields qualitatively similar results.

[17] We also explore the complement to entry, which is changes in the quality distribution of sellers who exit. The online Appendix shows that the quality distribution of sellers who exit exhibit thinner tails.

[18] The existence of a small group of sellers who are badged only *after* the policy change is due to sellers not being badged instantaneously upon meeting the requirements, but instead being certified once every month.

[19] Using thresholds for each group of three and four months out of six yields qualitatively similar results.

Table 2: Policy Change Impact on Different Incumbent Groups

	+/- 3 Months	+/- 6 Months	Month 7 to 12
Panel A. BB Incumbents			
	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	0.067	0.048	0.107***
	(0.047)	(0.039)	(0.041)
R^2	0.661	0.534	0.509
Panel B. BN Incumbents			
	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	-0.018	0.043**	0.086***
	(0.028)	(0.020)	(0.023)
R^2	0.820	0.779	0.753
Panel C. NB Incumbents			
	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	-0.064	0.014	-0.001
	(0.059)	(0.041)	(0.044)
R^2	0.494	0.473	0.474
Panel D. NN Incumbents			
	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	-0.012	0.007	0.051
	(0.038)	(0.028)	(0.031)
R^2	0.692	0.648	0.624
Panel E. BN Incumbents Who Regain Badge in the 3 Months Following the Policy Change			
	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	0.084**	0.121***	0.134***
	(0.041)	(0.032)	(0.035)
R^2	0.705	0.610	0.590
Panel F. BN Incumbents Who Remain Unbadged in the 3 Months Following the Policy Change			
	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	-0.044	0.005	0.051**
	(0.029)	(0.022)	(0.024)
R^2	0.783	0.740	0.720

Notes: The regressions are performed at the subcategory-month level.
*** indicates significance at p ≤ 0.01; ** p ≤ 0.05; * p ≤ 0.10.

We perform the DiD analyses on the four groups of incumbents in Table 2. Panels A–D show there is no statistically significant change in incumbents' quality when considering the sample period from three months before to three months after the policy change.[20] Using a period of six months before and after the policy change, we find that the only group that experiences a significant increase in EPP in the markets more affected by the policy change is group BN. This

[20]In the NN group, we look only at incumbents who have sold at least six items in the six months before the policy change to eliminate occasional sellers from our analysis.

result is consistent with our model's prediction: some s-type sellers who lose their badge due to the new policy will increase their quality to meet the new badge requirements.

To analyze this further, we distinguish between BN incumbents based on whether they regain their badge within the three months after the policy change.[21] We see in Panel E that a BN incumbent who regains her badge in the near future increases her quality in the three and six months after the policy change. On the other hand, a BN incumbent who remains unbadged in the near future does not increase her quality in neither the three months nor the six months after the policy change. This is an even finer test that is consistent with our model's prediction, showing that some of the quality improvement is due to more effort exerted by some incumbent sellers. The fact that the change in incumbents' behavior is attributed only to a small number of BN incumbents once again suggests that a significant fraction of the increase in quality by entrants at the tails of the quality distribution is likely due to selection rather than to behavioral changes.

6.3 Prices: Increases and Badge Premiums

Prediction (iv) states that prices increase for both badged and unbadged sellers. A challenge in comparing prices on eBay is that products vary wildly because sellers sell many different items that can be new or used, with a potentially high variation in the quality of items with the same title.

To establish an apples-to-apples comparison of prices, we follow the literature that studies price changes on eBay (e.g., Elfenbein et al. (2012), Einav et al. (2015), and Hui et al. (2016)), by taking advantage of product IDs in our data to construct an average price for each product that was listed as a new, fixed-price item that was sold. We use product IDs—which are eBay's finest-grain catalogue that is only defined for homogeneous products, thereby excluding heterogeneous products—to construct a data set at the Product ID–month level. For each individual item sold we define its relative price as the item's price divided by the average price of the product.

Columns 1 and 2 of Table 3 show the changes in the relative prices for different groups of sellers using transactions from one and three months before the policy change to one and three months after the change, where NN is the excluded group. The positive coefficient on *Policy*, which is significant for the $+/-3$ month window, shows that overall relative prices increase for unbadged NN sellers. Sellers who lose their badge (BN) experience a slight decrease in prices, while badged sellers (both BB and NB) experience a larger increase in prices than unbadged sellers.

[21] In Panels E and F, all badge statuses are based on the seller's actual status.

Table 3: Changes in Relative Prices: Event Study

	(1)	(2)
	+/−1 Month	+/−3 Months
Policy	0.005	0.027***
	(0.004)	(0.008)
BB*Policy	0.017***	0.027***
	(0.003)	(0.002)
BN*Policy	-0.009***	-0.033***
	(0.002)	(0.002)
NB*Policy	-0.005	0.059***
	(0.011)	(0.008)
Week FE	✓	✓
R^2	0.006	0.004

Notes: The regressions are performed at the transaction level. B or N indicates that the seller is badged or, respectively, not badged. The first (respectively, second) letter refers to the seller's status before (respectively, after) the policy change.

***significance at p $\leq$ 0.01; ** p $\leq$ 0.05; * p $\leq$ 0.1.

6.4 Welfare Impacts: A Back-of-the-Envelope Assessment

A comprehensive welfare analysis requires structurally estimating a variant of our model with additional assumptions, which may seem ad hoc. Instead, we offer a simple back-of-the-envelope analysis to estimate the effect of raising the bar on consumer surplus, which equals the difference between willingness to pay (WTP) for a product and the price paid for the product. We directly observe the sales price but not the WTP, and therefore need a proxy for it.

We therefore focus on items sold by auctions on eBay, which accounted for about half of the listings on eBay at the time. We exploit the fact that eBay's auctions resemble a second-price auction, in which it is known that a bidder's dominant strategy is to bid their valuation. We therefore use the winning bidder's bid as a proxy for their willingness-to-pay, and the difference between the winner's bid and the price paid as an approximation for the winner's consumer surplus.[22] Note that only the final price is shown on eBay's website, yet we observe all the bids using eBay's administrative data. To calculate changes in welfare, we estimate the following equation:

$$log(CS_{ijt}) = \gamma Policy_t + \eta_j + \epsilon_{ijt}, \tag{2}$$

where CS_{ijt} is winning bidder i's bid for product j at time t, minus the price that i pays; $Policy_t$

[22] This measures welfare as perceived by buyers at the time, as bids may not reflect the added benefits from higher quality of service due to higher quality sellers (faster shipping, better handling, etc.).

Table 4: Welfare Analyses

	OLS		LASSO	
	(1)	(2)	(3)	(4)
	log(CS)	log(CS)	γ_c	γ_c
	Full Sample	No Outliers	$\lambda = 5$	$\lambda = 1$
Post	0.022**	0.034***		
	(0.008)	(0.008)		
% Claim			20.371	254.934**
			(83.163)	(112.268)
$\widehat{\beta}_c$				0.176
				(0.672)
% New				-0.347
				(0.583)
% Low DSR2				-44.901
				(40.029)
% Low DSR3				-9.890
				-28.213
R^2	0.226	0.241	0.001	0.187

Notes: The regressions in columns (1) and (2) are performed at the transaction level according to equation (2) using data from one month before and one month after the policy change. The regressions in columns (3) and (4) are performed at the market (or subcategory) level. In column (2), we use only transactions where the winning bidder's bid is no more than twice the final price. In columns (3) and (4), γ_c is the estimated welfare change for each subcategory using equation (2); and λ is the penalty coefficient in LASSO regressions.

is a dummy variable that equals 1 for transactions that took place after the policy change; η_j are Product ID fixed effects; and ϵ_{ijt} are idiosyncratic errors.[23] The parameter of interest is γ, which estimates the inter-temporal change in consumer surplus. Controlling for Product ID fixed effects is essential for this analysis so that we can compare consumer surplus of identical items. Additionally, we use transactions in the month before and the month after the policy change for the estimation to mitigate the concern that item value may change over time.

The results are reported in Table 4. In column (1), we estimate equation (2) using the full sample of auctions of items with a Product ID. The average consumer surplus increases by 2.2%, which indicates that *on average* the policy change is beneficial for consumers. In column (2), we repeat the analysis on the subset of auctions where the winning bidder's bid is no more than twice the final price to remove outlier transactions, and find that consumer welfare increases by 3.4%.

Next, we estimate equation (2) for each subcategory separately. Interestingly, we find that the impact of the policy change exhibits a wide dispersion across different subcategories. Recall from our model that the underlying distribution of seller types determines how the policy change

[23]As mentioned earlier, product IDs are eBay's finest partition of items. (e.g., a white, 64GB, AT&T compatible iPhone 6 has its own product ID, which differs from that of an iPhone of a different color, memory, or carrier.

will impact entry and exit, and consumer preferences determine how welfare changes as a result of raising the bar. Because we cannot predict what characteristics in each subcategory would correlate with welfare changes, a LASSO regression for variable selection performed at the subcategory level. The dependent variable is γ_c, and the right hand side variables include different subcategory characteristics before the change was implemented: the first-stage estimate of the change in the share of badged sellers $\widehat{\beta}_c$; share of badged sellers before the policy change; average price in the category; market share of badged and established sellers; and share of transactions with a bad buyer experience such as buyer claims, low DSRs, and negative feedback ratings.

The LASSO results are reported in Table 4, columns (3) and (4). When the penalization parameter λ is 5, the model selects the most predictive variable of the estimated change in welfare, which is the share of transactions with a buyer claim. If we interpret % Claim as the preference for quality in a market, a positive (but not statistically significant) correlation between this measure and γ_c is consistent with the idea that welfare increases more in subcategories with a stronger preference for quality. With a smaller penalty of $\lambda = 1$, the model selects several more variables: $\widehat{\beta}_c$; the share of transactions in which the product is new; and the share of transactions with low DSRs on communication and shipping speed. What's more, % Claim becomes a statistically significant predictor of changes in welfare across subcategories.[24] Additionally, we see that $\widehat{\beta}_c$ is positively (although not statistically significant) correlated with the estimated welfare change, consistent with the idea that markets with a larger share of s-type sellers are affected more by raising the bar, and hence is correlated with a larger welfare change.

These results sheds some light on how practitioners can use certification thresholds. If consumer welfare is key and consumers value quality a lot, then offering them more precise and stringent signals will be beneficial. Additionally, the effect of information policies is larger in markets where sellers can respond by changing their quality rather than those that are only subject to selection.

7 Endogeneity and Robustness

A critical identification assumption is that there are no time-varying heterogeneities across subcategories that simultaneously affect changes in the share of badged sellers and in the entry variables. Because we cannot directly test the exclusion restriction, we run placebo tests as well as a robustness check for the identification. We also perform several analyses to confirm that our results are

[24]% Low DSR2 and % Low DSR3 are not statistically significant because they are highly correlated with % claim.

robust to different specifications, including applying different time windows used in the definition of EPP and repeating our DiD analyses using an event-study approach instead of a simulation, for brevity, they are reported in the online appendix.

7.1 Placebo Tests on the Exclusion Restriction

Consider the following thought experiment. Suppose there exist serially correlated subcategory-specific confounders that drive our results and assume that they persist over time. This implies that if we repeat our analyses using data from the year before the policy change *as if* there was a hypothetical policy change, we would obtain similar results.

To test this, we perform the following placebo test. We simulate the change in badge requirements in the year before the actual policy change to obtain another set of $\widehat{\beta}$. Then we repeat our main specification using data around this hypothetical policy change date. We perform this analysis to test the sharp predictions of our model. Because the policy change in the previous year is purely hypothetical, we do *not* expect to see results similar to those in our main analyses.

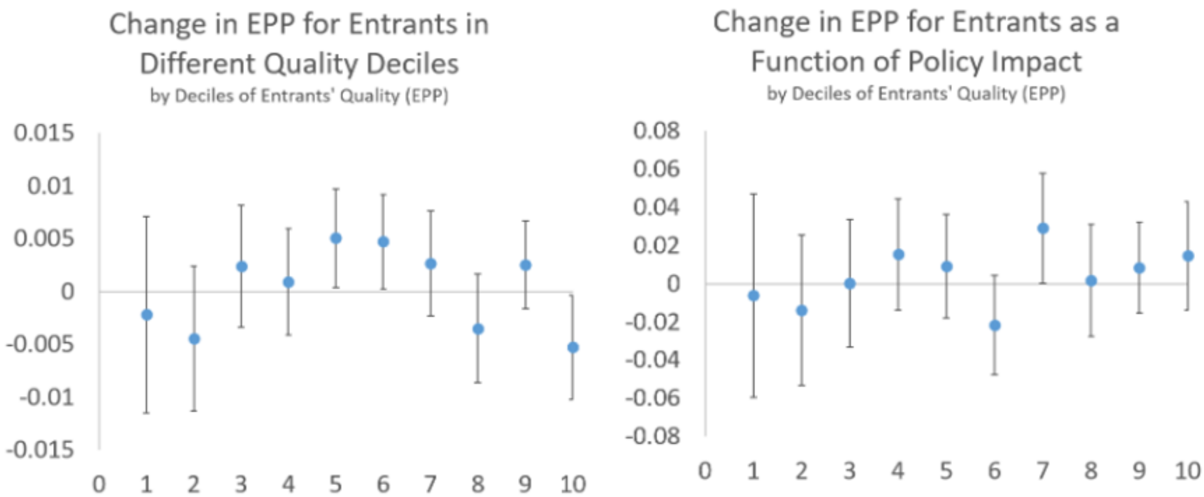
We first replicate our model's prediction on the fattening and thinning of the tails. In Figure 6, we replicate the results in Figure 5 using the placebo date. In both the event study specification and the continuous DiD specification, we do not observe the monotonic relation in the estimates with respect to the deciles of entrants' quality.

Next, we replicate the results on incumbents using the placebo design. Table 5 reports the estimated γ's for EPP. If we focus on short periods (*i.e.,* 3 and 6 months before and after) the hypothetical policy change, we don't see significant changes in incumbents' quality measured by EPP. Importantly, we don't see quality improvement from sellers who "lose" their badge by simulation and "regain" it in 3 months, because they did not actually lose the badge.

Lastly, we replicate the results on price changes for different groups of incumbents using the placebo design, as reported in Table 6. While we see a price increase for the NB group and a price decrease for the BN group, which is intuitive, we don't see price increases for the BB and NN group. The latter results suggest that the price increases for the BB and NN group that we identified in the paper is consistent with the pooling mechanism in our theory.

In principle, there could still exist time-varying but not serially-correlated confounders that can contaminate our causal interpretation, which we talk about in the following section. However, the fact that the estimates in the placebo analyses do not present the same patterns as the ones we see from the main analyses is reassuring.

Figure 6: Placebo: Change in EPP for Entrants in Different Quality Deciles



Notes: The left figure shows the average within-subcategory change in EPP after the hypothetical policy change. The right figure shows the across-subcategory change in EPP as a function of the hypothetical policy exposure. Bars indicate 95% confidence intervals.

Notes: The left figure shows the average within-subcategory change in EPP after the hypothetical policy change. The right figure shows the across-subcategory change in EPP as a function of the hypothetical policy exposure. Bars indicate 95% confidence intervals.

7.2 Controlling for Time-Varying Market Characteristics

Despite controlling for market fixed effects in our DiD specification, a threat to identification arises if there are time-varying market characteristics that simultaneously correlate with the estimated policy exposure and entry. Our placebo tests would detect these time-varying factors only when they are serially correlated. One way to mitigate concerns over time-varying, non-serially correlated market characteristics is to rerun our second-stage regressions while controlling for many time-varying variables that may impact entry and entrant quality and at the same time be correlated with our measure of policy exposure. We can then test whether the estimates are robust to the inclusion of these time-varying market characteristics.

In particular, in Table 7, "DSR1"–"DSR4" are Detailed Seller Ratings for item as described, communication, shipping speed, and shipping charge, on a five-point scale; "% Badged" is the number of transactions by badged sellers divided by the total number of transactions; "Price" is the average sales price of items in a market; "Quantity" is the total number of items that are sold in a market; "# Seller" is the total number of sellers in a market; "# Buyer" is the total number of buyers in a market; "% Claim" is the number of disputes filed by buyers divided by the total number of transactions; "% BBE" is the number of bad buyer experiences (non-positive feedback,

Table 5: Placebo: Hypothetical Policy Change Impact on Different Incumbent Groups

	+/- 3 Months	+/- 6 Months	Month 7 to 12
Panel A. BB Incumbents			
Estimate	-0.060	0.023	0.037
	(0.088)	(0.062)	(0.026)
R^2	0.620	0.518	0.511
Panel B. BN Incumbents			
Estimate	0.075	0.032	0.094**
	(0.069)	(0.048)	(0.044)
R^2	0.750	0.719	0.746
Panel C. NB Incumbents			
Estimate	-0.059	-0.024	-0.006
	(0.090)	(0.063)	(0.060)
R^2	0.470	0.432	0.452
Panel D. NN Incumbents			
Estimate	0.053	0.038	0.003
	(0.063)	(0.045)	(0.043)
R^2	0.850	0.836	0.847
Panel E. BN Incumbents Who Regain Badge in the 3 Months Following the Policy Change			
Estimate	0.026	0.040	0.021
	(0.042)	(0.064)	(0.036)
R^2	0.529	0.466	0.471
Panel F. BN Incumbents Who Remain Unbadged in the 3 Months Following the Policy Change			
Estimate	-0.051	0.080	0.038
	(0.109)	(0.077)	(0.077)
R^2	0.604	0.503	0.533

Notes: The placebo analyses use the $\widehat{\beta}$ estimated from September in the year before the policy change, and we re-perform the second-stage regression using data around that month. The regressions are performed at the subcategory-month level. Badge status is simulated by applying the new policy requirements to incumbent sellers, defined as sellers who list at least one item both before and after the policy change.

*** indicates significance at p $\leq$ 0.01; ** p $\leq$ 0.05; * p $\leq$ 0.10.

one- or two-star DSRs, buyer dispute) divided by the total number of transactions. The coefficient estimates after including these covariates are very similar to those of our baseline model and, in fact, the effect of the policy change is even stronger.

Figure 7 plots changes in entrants' quality for different quality deciles after controlling for time-

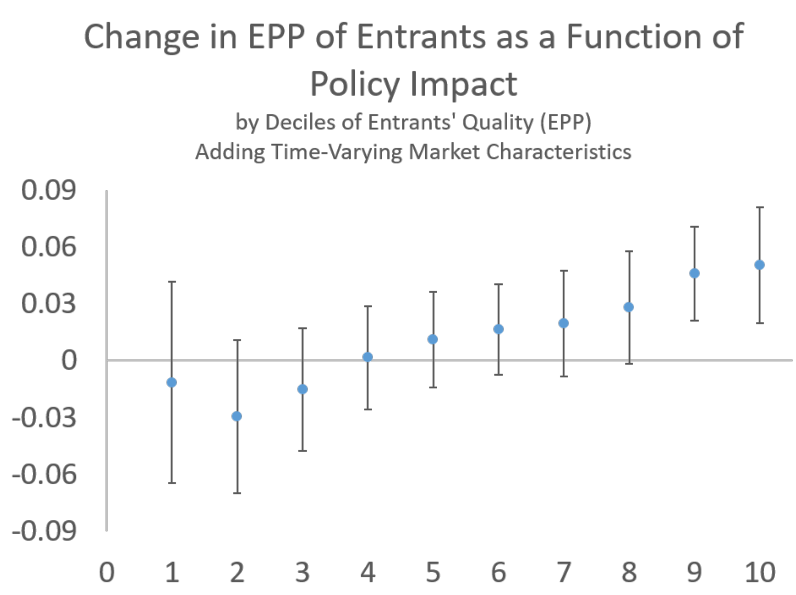
Table 6: Placebo: Changes in Relative Prices: Event Study

	(1)	(2)
	+/−1 Month	+/−3 Months
Policy	-0.006***	-0.015
	(0.001)	(0.029)
BB*Policy	0.008	0.010
	(0.007)	(0.014)
BN*Policy	-0.011***	-0.052***
	(0.003)	(0.006)
NB*Policy	0.015	0.057***
	(0.018)	(0.013)
Week FE	✓	✓
R^2	0.006	0.004

Notes: The placebo analyses use the $\widehat{\beta}$ estimated from September in the year before the policy change, and we re-perform the second-stage regression using data around that month. The regressions are performed at the transaction level. B or N indicates that the seller is badged or, respectively, not badged. The first (respectively, second) letter refers to the seller's status before (respectively, after) the hypothetical policy change.

***significance at p $\leq$ 0.01; ** p $\leq$ 0.05; * p $\leq$ 0.1.

Figure 7: Changes in EPP for Entrants in Different Quality Deciles, Controlling for Time-Varying Market Characteristics



Notes: The figure corresponds to the right graph in Figure 5. It shows the across-subcategory change in EPP as a function of policy exposure using the DiD specification. Bars indicate 95% confidence intervals.

Table 7: Adding More Controls in the DiD Estimation

	Entrant Ratio			EPP		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 3 Months	+/- 6 Months	Month 7 to 12	+/- 3 Months	+/- 6 Months	Month 7 to 12
Estimate	0.135***	0.089***	-0.110***	0.070***	0.035**	0.085***
	(0.022)	(0.017)	(0.034)	(0.020)	(0.015)	(0.026)
DSR1	0.059	0.019	0.054	0.145	-0.104***	-0.285***
	(0.067)	(0.042)	(0.057)	(0.090)	(0.040)	(0.057)
DSR2	-0.193***	0.033	-0.003	-0.141	0.155***	0.097
	(0.068)	(0.048)	(0.087)	(0.087)	(0.038)	(0.072)
DSR3	0.233***	-0.107***	0.054	0.147***	-0.054	0.048
	(0.052)	(0.030)	(0.073)	(0.054)	(0.036)	(0.057)
DSR4	0.110*	0.099***	-0.078	-0.202***	0.063*	-0.055
	(0.057)	(0.035)	(0.060)	(0.057)	(0.035)	(0.050)
% Badged	-0.036**	0.000	-0.262***	-0.020	-0.003	-0.030
	(0.015)	(0.010)	(0.021)	(0.013)	(0.009)	(0.019)
Price	5E-06	7E-06***	-8E-07	3E-04***	4E-04***	-3E-04***
	(5E-06)	(2E-06)	(1E-06)	(1E-04)	(6E-05)	(6E-05)
Quantity	-2E-07	-3E-08	-5E-07***	2E-08	3E-08	-4E-08
	(1E-07)	(4E-08)	(1E-07)	(1E-07)	(4E-08)	(5E-08)
# Seller	4E-06***	3E-06***	-1E-06	-1E-06*	-4E-07	-2E-07
	(8E-07)	(3E-07)	(1E-06)	(7E-07)	(3E-07)	(5E-07)
# Buyer	4E-07	9E-08	9E-07***	4E-08	-5E-08	1E-07
	(3E-07)	(8E-08)	(3E-07)	(2E-07)	(8E-08)	(1E-07)
EPP	-0.010	0.027	0.286***	0.112**	0.066**	0.170***
	(0.060)	(0.034)	(0.068)	(0.055)	(0.031)	(0.047)
% Claim	0.490	0.791***	-0.023	-1.666***	-0.338	1.715***
	(0.462)	(0.256)	(0.452)	(0.445)	(0.249)	(0.281)
% BBE	0.394	-0.587	0.252	1.647***	0.402	-1.613***
	(0.388)	(0.249)	(0.429)	(0.433)	(0.253)	(0.263)
R^2	0.918	0.896	0.697	0.778	0.725	0.778

Notes: The regressions are performed at the subcategory-month level. Here DSR1-4 are Detailed Seller Ratings for item as described, communication, shipping speed, and shipping charge, on a five-point scale; % Badged is the number of transactions by badged sellers divided by the total number of transactions; Price is the average sales price of items in a market; Quantity is the total number of items that are sold in a market; # Seller is the total number of sellers in a market; # Buyer is the total number of buyers in a market; % Claim is the number of disputes filed by buyers divided by the total number of transactions; and % BBE is the number of bad buyer experiences (non-positive feedback, one- or two-star DSRs, buyer dispute) divided by the total number of transactions.

varying market characteristics. It corresponds to the right plot in Figure 5 and it show that the monotonic increasing pattern mostly hold even after controlling for the above-mentioned variables.

8 Conclusion

We develop a parsimonious model of certification in markets with asymmetric information to explore how certification choices impact market outcomes and, in particular, the distribution of quality. We take the model's predictions to data from eBay and exploit the heterogeneous impact of a policy change across different markets. This allows us to identify our model's rich predictions regarding how a policy change will impact the distribution of seller-quality and of prices across markets.

The predictions of our theoretical model are borne out in the data. First, the distribution of quality provided by entrants has fatter tails after the policy change. Second, most incumbents do not change the quality of their service except for a small group of incumbents who regain their badge by increasing their quality. Finally, restricting attention to well-defined products, we find that aside from the products of sellers who lose their badge, relative prices increase.

As we mentioned in subsection 6.2, the fact that only a small fraction of sellers regain their badge suggests that selection is the main force behind our results. The ingredients of our model suggest that potential entrants with lower entry costs will respond more to the change, which has another implication. Namely, new sellers on eBay will have more hurdles to overcome in entering a sub-category compared to existing sellers who are laterally entering the same sub-category, as the former need to learn more about how eBay operates and how to sell effectively. We repeat our analyses for these two types of entrants separately in the Appendix and show that the results are consistent with new sellers having higher entry costs. This too suggest that a significant part of the observed changes in the quality provided by entrants is linked to selection in entry and exit.

Overall, our findings indicate that providing third-party certification not affects the rate of entry in a market, but also the quality of entrants, and hence, how markets evolve over time. As online marketplaces have become more widespread, from products and services, to labor markets and more, our results offer guidance for electronic marketplace designers who wish to use certification mechanisms. We find that raising (respectively, lowering) the bar of the certifying badge will broaden (respectively, contract) the quality distribution of all sellers in the marketplace. The optimal certification bar is dependent on market characteristics and consumer preferences, and as our back-of-the-envelope analysis shows, a well-chosen change in the certification threshold can increase consumer surplus, making the marketplace a more attractive place for consumers.

We view our contribution as a first step in providing insights on matters of quality certification that apply more broadly to other markets with asymmetric information where thresholds-based certification badges (or labels) are commonplace. If anything, the proliferation of online trade and the ability of individuals and firms to access both larger amounts of information and more possible trading parties, should make the use of badges even more widespread.

References

Avery, Christopher, Paul Resnick, and Richard Zeckhauser, "The market for evaluations," *American Economic Review*, 1999, *89* (3), 564–584.

Bajari, Patrick and Ali Hortacsu, "Economic insights from internet auctions," *Journal of Economic Literature*, 2004, *42* (2), 457–486.

_ , Christina Dalton, Han Hong, and Ahmed Khwaja, "Moral hazard, adverse selection, and health expenditures: A semiparametric analysis," *The RAND Journal of Economics*, 2014, *45* (4), 747–763.

Butler, Jeffrey V, Enrica Carbone, Pierluigi Conzo, and Giancarlo Spagnolo, "Past performance and entry in procurement: An experimental investigation," *Journal of Economic Behavior & Organization*, 2020, *173*, 179–195.

Cabral, Luis, "Reputation on the Internet," *The Oxford handbook of the digital economy*, 2012, pp. 343–354.

_ and Ali Hortacsu, "The dynamics of seller reputation: Evidence from ebay," *The Journal of Industrial Economics*, 2010, *58* (1), 54–78.

Diamond, Douglas W., "Reputation Acquisition in Debt Markets," *Journal of Political Economy*, 1989, *97* (4), 828–862.

Dranove, David and Ginger Zhe Jin, "Quality disclosure and certification: Theory and practice," *Journal of Economic Literature*, 2010, *48* (4), 935–963.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen, "Selection on moral hazard in health insurance," *American Economic Review*, 2013, *103* (1), 178–219.

_ , Theresa Kuchler, Jonathan Levin, and Neel Sundaresan, "Assessing sale strategies in online markets using matched listings," *American Economic Journal: Microeconomics*, 2015, *7* (2), 215–47.

Elfenbein, Daniel W, Raymond Fisman, and Brian McManus, "Charity as a Substitute for Reputation: Evidence from an Online Marketplace," *The Review of Economic Studies*, 2012, *79* (4).

_ , _ , and _ , “Market structure, reputation, and the value of quality certification,” *American Economic Journal: Microeconomics*, 2015, *7* (4), 83–108.

Fan, Ying, Jiandong Ju, and Mo Xiao, “Reputation premium and reputation management: Evidence from the largest e-commerce platform in China,” *International Journal of Industrial Organization*, 2016, *46*, 63–76.

Fradkin, Andrey, Elena Grewal, and David Holtz, “The determinants of online review informativeness: Evidence from field experiments on Airbnb,” Technical Report, Working Paper 2017.

(GAO-12-102R), Government Accountability Office, “Prior Experience and Past Performance as Evaluation Criteria in the Award of Federal Construction Contracts.”

Greenstone, Michael, Paul Oyer, and Annette Vissing-Jorgensen, “Mandated disclosure, stock returns, and the 1964 Securities Acts amendments,” *The Quarterly Journal of Economics*, 2006, *121* (2), 399–460.

Grossman, Gene M and Henrik Horn, “Infant-industry protection reconsidered: the case of informational barriers to entry,” *Quarterly Journal of Economics*, 1988, *103* (4), 767–787.

Grossman, Sanford J, “The informational role of warranties and private disclosure about product quality,” *The Journal of Law and Economics*, 1981, *24* (3), 461–483.

Hopenhayn, Hugo and Maryam Saeedi, “Optimal Quality Ratings and Market Outcomes,” Working Paper 2019.

Hui, Xiang, Maryam Saeedi, and Neel Sundaresan, “Adverse Selection or Moral Hazard: An Empirical Study,” *Journal of Industrial Economics*, 2018.

_ , _ , Zeqian Shen, and Neel Sundaresan, “Reputation and regulations: evidence from eBay,” *Management Science*, 2016, *62* (12), 3604–3616.

Jullien, Bruno and In-Uck Park, “New, like new, or very good? Reputation and credibility,” *The Review of Economic Studies*, 2014, *81* (4), 1543–1574.

Klein, Benjamin and Keith B Leffler, “The role of market forces in assuring contractual performance,” *Journal of Political Economy*, 1981, *89* (4), 615–641.

Klein, Tobias J, Christian Lambertz, and Konrad O Stahl, "Market transparency, adverse selection, and moral hazard," *Journal of Political Economy*, 2016, *124* (6), 1677–1713.

Leland, Hayne E., "Quacks, Lemons and Licensing: a Theory of Minimum Quality Standards," *Journal of Political Economy*, 1979, *87* (6), 1328–1346.

Luca, Micheal, "Reviews, reputation, and revenue: The case of Yelp. com," *Harvard Business School Working paper 12*, 2011, *16*, 12–016.

Mian, Atif and Amir Sufi, "The effects of fiscal stimulus: Evidence from the 2009 cash for clunkers program," *Quarterly journal of economics*, 2012, *127* (3), 1107–1142.

Nosko, Chris and Steven Tadelis, "The limits of reputation in platform markets: An empirical analysis and field experiment," Technical Report, National Bureau of Economic Research 2015.

Shapiro, Carl, "Premiums for high quality products as returns to reputations," *Quarterly journal of economics*, 1983, *98* (4), 659–679.

Stahl, Konrad and Roland Strausz, "Certification and market transparency," *The Review of Economic Studies*, 2017, *84* (4), 1842–1868.

Tadelis, Steven, "Reputation and feedback systems in online platform markets," *Annual Review of Economics*, 2016, *8*, 321–340.

Zervas, Georgios, Davide Proserpio, and John Byers, "A first look at online reputation on Airbnb, where every stay is above average," 2015.