

Electronic Companion for “Do Noisy Customer Reviews Discourage Platform Sellers? Empirical Analysis of an Online Solar Marketplace”

Note #1: The inverted U-shaped relationship between the installer’s activity level and its competitors’ rating dispersion can be explained as follows. When the rating dispersion of competitors is low, the installer might attribute this dispersion to competitor-related issues rather than systemic market level issues (e.g., polarized customer taste). Thus, the installer might interpret any increase in its competitors’ rating dispersion as damaged competitor reputation, which increases the installer’s likelihood of winning customers. Because of the installer’s larger likelihood of winning, any increase in the competitors’ rating dispersion can motivate the installer to make more proposals. In contrast, when the rating dispersion of competitors is high, the installer might attribute this situation to highly differentiated customer taste in the market. Hence, the installer can perceive any increase in competitors’ rating dispersion as an increased likelihood of a negative customer. To mitigate this risk, the installer might be more selective in its project choices and reduce its activity level.

A. Details of Defining Local Markets

The OPTICS routine is an unsupervised machine learning algorithm that identifies density-based clusters in spatial data. It is an extension of various commonly-used clustering algorithms (Kanagala and Krishnaiah 2016). Among others, an important advantage of the OPTICS algorithm is that it does not require fixing the number of clusters before running the algorithm as in k -means clustering method; rather, it generates clusters based on the relative locations and densities of data points. In contrast to k -means, it allows clusters to be non-convex - a possible scenario as business activities can be shaped by geography (rivers, mountains, etc.) or distribution of population. Because of these and many other advantages, OPTICS has been applied in various contexts, ranging from political science (Davidson 2019) to geography (Teimouri et al. 2016). To our knowledge, in the OM literature, there is no prior work that uses this clustering technique.

In light of these, we create the geographic division of local markets with the following steps. First, we collected the 5-digit zipcode of every installer in the marketplace. We then converted each zipcode to the representative coordinates based on the data provided by the US Census Bureau (2019). This transformation is necessary to run the OPTICS algorithm on the location data. Then, we run the OPTICS algorithm to generate clusters. The OPTICS algorithm uses two key hyperparameters: (1) number of samples in a neighborhood for a point to be considered as a core point ($min_samples$) and (2) distance between two samples for one to be considered as in the neighborhood (eps). A cluster is identified by first identifying a *core sample* where there are more than $min_samples$ number of points within eps , followed by adding other points which are within eps from the *core sample*. We set the value of $min_samples$ at 2 to indicate that a minimum of 2 installers is required to form a *core sample* which is subset of a cluster. Based on our conversations with practitioners, we set 90 kilometers as the distance input parameter eps , and the OPTICS algorithm generated 37 clusters.

(We also checked the robustness of our results by trying different *eps* values from 80 - 100 kilometers in the OPTICS algorithm. Our insights remain to be valid with these alternative parameters.) Each of these clusters geographically defines a local market boundary. Figure 2 illustrates the centroid of each cluster, which represents the *centroid* of each local market. Considering the clusters with active installers, the mean (median) number of the active installers in a month is 5.99 (4) in a local market.

B. BERT

We now elaborate the BERT model we used to *vectorize* the text reviews. BERT is a natural language processing (NLP) model that transforms texts into numeric vectors while also preserving the meaning of texts. It belongs to the category of NLP methods that performs word embedding. In literature, in different contexts than ours, text data are commonly vectorized based on word counts, ignoring the semantics and word ordering (see, e.g., Hoberg and Phillips (2016) and Loughran and McDonald (2011)). However, our context involves texts that are informal writings and often contain emotions. Simply capturing word frequencies does not provide accurate results if similar emotions can be expressed with synonymous words. Thus, our analysis requires a vectorization that preserves the information and sentiment of the text reviews despite the use of synonyms and/or different styles. The BERT model achieves that. Specifically, the BERT model has two distinct advantages. First, it understands the semantics. For example, consider the 3 sentences:

Sentence 1: they did a good job. Sentence 2: they did an awful job. Sentence 3: they did a great job.

Considering the meaning of the sentences, we expect the distance between sentences 1 and 3 to be smaller than the distance between 2 and 3 or 1 and 2, i.e., $D(1,3) < D(2,3)$ and $D(1,3) < D(1,2)$. The BERT model vectorization enables just that; it projects “good” and “great” to vectors that are closer to each other. In this example, with BERT, we have $D(1,3) < D(1,2) < D(2,3)$. This level of distinction is not feasible without word embedding (e.g., by simply using a word counter vectorizer). Second, the BERT model takes word ordering into account. For example, the two sentences “The food was good, not bad at all” and “The food was bad, not good at all” have the opposite meanings. Common vectorization methods (e.g., bag-of-words approach) are not able to capture this difference as words and number of counts are the same in both sentences. But, the BERT model can easily differentiate between these two sentences.

C. Further Details on §7.8

Note #2: We define the proportion of no-rating transactions for installer i up to (and including) month t as $Proportion_No-Rating_Self_{i,t} = \frac{Matches_{i,t-1} - Reviews_{i,t}}{Matches_{i,t-1}}$, where $Matches_{i,t-1}$ is installer i 's total number of winning quotes up to (and including) period $t - 1$ and $Reviews_{i,t}$ represents installer i 's total review count up to (and including) period t . Note that $Matches_{i,t-1}$ and $Reviews_{i,t}$ have different time indices to capture time between a transaction and the associated online review post. We also construct the

market-level average proportion of no-rating transactions - $Average_Proportion_No_Rating_Mkt_{m,t}$ - by taking the average of $Proportion_No_Rating_Self_{i,t}$ over all active installers in the focal market m in period t .

Note #3: Including $Proportion_No_Rating_Self_{i,t}$ and $Average_Proportion_No_Rating_Mkt_{m,t}$ as control variables in the installer- and market-level regressions respectively is one way to test the robustness of our findings with respect to the proportion of no-rating transactions. As mentioned in the main text, to perform an alternative robustness check, we incorporate the proportion of no-rating transactions in the main entropy variables without altering their interpretations. To do so, we simply construct the *weighted average rating entropy* $\mathbb{W}\mathbb{A}$ [Rating Entropy]:

$$\begin{aligned} \mathbb{W}\mathbb{A} [\text{Rating Entropy}] &= [\text{Rating Entropy} | \text{Transactions with Ratings}] \times \text{Fraction of Transactions with Ratings} \\ &\quad + [\text{Rating Entropy} | \text{Transactions with no Ratings}] \times \text{Fraction of No-Rating Transactions,} \end{aligned}$$

where $[\text{Rating Entropy} | \text{Transactions with no Ratings}]$ is taken as zero. Using this and the fact that $\text{Fraction of Transactions with Ratings} = 1 - \text{Fraction of No-Rating Transactions}$, we specifically construct the following two new variables:

$$\begin{aligned} \mathbb{W}\mathbb{A} [\text{Rating_Entropy_Self}_{i,t}] &= [\text{Rating_Entropy_Self}_{i,t} | \text{Transactions with Ratings}_{i,t}] \times (1 - \text{Proportion_No-Rating_Self}_{i,t}) \\ \mathbb{W}\mathbb{A} [\text{Rating_Entropy_Mkt}_{m,t}] &= [\text{Rating_Entropy_Mkt}_{m,t} | \text{Transactions with Ratings}_{m,t}] \times (1 - \text{Average_Proportion_No-Rating_Mkt}_{m,t}). \end{aligned}$$

The variable $\mathbb{W}\mathbb{A} [\text{Rating_Entropy_Self}_{i,t}]$ represents the weighted average rating entropy of installer i considering all its ratings up to and including month t , and the variable $\mathbb{W}\mathbb{A} [\text{Rating_Entropy_Mkt}_{m,t}]$ represents the weighted average rating entropy of market m considering all the market's ratings up to and including month t . The variables $Proportion_No_Rating_Self_{i,t}$ and $Average_Proportion_No_Rating_Mkt_{m,t}$ are as defined earlier in the main body of the paper; $Transactions\ with\ Ratings_{i,t}$ (respectively, $Transactions\ with\ Ratings_{m,t}$) stands for installer i 's (respectively, market m 's) transactions with ratings up to and including month t . We did not construct the weighted average of competitors' rating entropy because competitors' proportion of no-rating response is not visible to an installer.

Rating	Review
1	The company is selling unregistered securities and making double you money claims to unsuspecting investors.
2	Proposal was competitive but they didn't show up for a scheduled appointment for an in-home consultation.
3	It didn't work out. Let's leave it at that.
4	My project is still under review. Company was responsive to my inquiry.
5	They did what they said they would at the price quoted. What more can you say?

Table EC.1 Excerpts of reviews given by customers along with respective ratings

Variables	Mean	Standard Deviation	Min	Max
1 star Rating %	2.258	10.819	0.000	100.00
2 stars Rating %	0.641	6.380	0.000	100.00
3 stars Rating %	0.535	4.378	0.000	48.65
4 stars Rating %	2.872	11.332	0.000	100.00
5 stars Rating %	93.695	16.990	0.000	100.00

Distribution is taken across all installers over their entire horizons

Table EC.2 Summary statistics of rating distribution

Variables	Mean	Standard Deviation	Min	Max
% of Monthly Proposals for an Installer in a Market	14.728	22.889	0.000	100.000
% of Monthly Matches for an Installer in a Market	11.611	24.482	0.000	100.00

Table EC.3 Summary statistics of cluster performance for installers in a market

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Rating_Entropy_Self	1						
(2) Rating_Entropy_Others	-0.093***	1					
(3) Average_Rating_Self	-0.089***	0.069***	1				
(4) Average_Rating_Others	0.026*	-0.523***	-0.017	1			
(5) Experience	0.000	0.143***	0.036**	-0.062***	1		
(6) Price_Difference	-0.012	-0.042***	0.004	0.014	-0.025*	1	
(7) Average_LogRevenue	0.038***	0.090***	0.019	-0.028*	0.327***	0.003	1

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table EC.4 Correlation matrix - installer-level

Variable	VIF	1/VIF
Rating_Entropy_Self	5.33	0.188
Rating_Entropy_Self ²	5.34	0.187
Rating_Entropy_Others	1.70	0.587
Rating_Entropy_Others ²	1.30	0.769
Average_Rating_Self	1.02	0.983
Average_Rating_Others	1.39	0.718
Experience	1.18	0.848
Price_Difference	1.01	0.985
Average_LogRevenue	1.13	0.885

Table EC.5 VIF table: installer-level

Variable	Shapley Value (Estimate)	Percent (Estimate)
Rating_Entropy_Self	0.0361	38.36
Rating_Entropy_Self ²	0.0141	15.00
Rating_Entropy_Others	0.0140	14.88
Rating_Entropy_Others ²	0.0054	5.74
Average_Rating_Self	0.0004	0.42
Average_Rating_Others	0.0028	2.99

Table EC.6 Shapley table: installer-level

Variable	VIF	1/VIF
Rating_Entropy_Mkt	2.78	0.360
Rating_Entropy_Mkt ²	2.15	0.465
Average_Rating_Mkt	1.61	0.621
Average_Experience	1.10	0.907
Price_Difference_Mkt	1.01	0.992
Market_LogRevenue	1.05	0.957

Table EC.7 VIF table: market-level

Variables	(1)	(2)	(3)	(4)	(5)
(1) Rating_Entropy_Mkt	1				
(2) Average_Rating_Mkt	-0.611***	1			
(3) Average_Experience	0.232***	-0.127***	1		
(4) Price_Difference_Mkt	-0.015	-0.018	-0.056	1	
(5) Market_LogRevenue	0.029	-0.001	0.109***	-0.309	1

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table EC.8 Correlation matrix - market-level

Variable	Shapley Value (Estimate)	Percent (Estimate)
Rating_Entropy_Mkt	0.092	28.50
Rating_Entropy_Mkt ²	0.076	23.33
Average_Rating_Mkt	0.011	3.24

Table EC.9 Shapley table: market-level

Variables	N	Mean	Standard Deviation	Min	Max
Text-Length of Reviews of Installers	3387	98.919	104.617	1.000	1992.000
Average Rating of Installers	3387	4.626	1.083	1.000	5.000

Table EC.10 Summary statistics of reviews and corresponding ratings

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Average_Sentiment_Self	1						
(2) Average_Sentiment_Others	-0.035**	1					
(3) Text_Dispersion_Others	-0.048***	0.117***	1				
(4) Text_Dispersion_Self	0.101***	-0.077***	-0.088***	1			
(5) Experience	0.081***	0.062***	0.077***	0.069***	1		
(6) Price_Difference	-0.035**	0.003	-0.010	-0.056***	-0.025*	1	
(7) Average_LogRevenue	0.022	0.022	0.111***	0.071***	0.327***	0.003	1

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table EC.11 Correlation matrix - installer-level text-based analysis

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Average_Sentiment_Self	1							
(2) Average_Rating_Self	0.827***	1						
(3) Average_Sentiment_Others	-0.035**	-0.068***	1					
(4) Average_Rating_Others	-0.067***	-0.017	0.457***	1				
(5) Text_Dispersion_Others	-0.048***	-0.004	0.117***	-0.107***	1			
(6) Rating_Entropy_Others	0.117***	0.069***	-0.094***	-0.523***	0.177***	1		
(7) Text_Dispersion_Self	0.101***	0.138***	-0.077***	0.000	-0.088***	0.007	1	
(8) Rating_Entropy_Self	-0.086***	-0.089***	0.069***	0.026*	0.057***	-0.093***	0.306***	1

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table EC.12 Correlation of ratings and text-based measures (installer-level)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Text_Dispersion_Mkt	1						
(2) Rating_Entropy_Mkt	0.099***	1					
(3) Average_Sentiment_Mkt	-0.362***	0.079**	1				
(4) Average_Rating_Mkt	-0.104***	-0.611***	0.226***	1			
(5) Average_Experience	-0.029	0.232***	0.115***	-0.127***	1		
(6) Price_Difference_Mkt	0.097***	-0.015	0.009	-0.018	-0.056	1	
(7) Market_LogRevenue	-0.057	0.029	0.007	-0.001	0.109***	-0.039	1

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table EC.13 Correlation matrix - market-level text-based analysis

Variables	N	Mean	Standard Deviation	Min	Max
Text_Dispersion_Self	4,384	0.000	0.070	-0.099	0.282
Text_Dispersion_Others	4,384	0.000	0.027	-0.122	0.244
Average_Sentiment_Self	4,384	0.633	0.238	-0.827	0.958
Average_Sentiment_Others	4,384	0.681	0.092	-0.599	0.958
Text_Dispersion_Mkt	748	0.000	0.035	-0.061	0.236
Average_Sentiment_Mkt	748	0.675	0.107	0.059	0.886

Table EC.14 Summary statistics of text-based variables for installer and market-level

Variables	(I) Installer's Activity Level	(II) Installer's Activity Level	(III) Installer's Activity Level	(IV) Installer's Activity Level
Text_Dispersion_Self	1.266* (0.023)	1.255* (0.025)	1.213* (0.034)	1.294* (0.022)
Text_Dispersion_Self ²	-27.023*** (0.000)	-27.229*** (0.000)	-21.619*** (0.000)	-22.007*** (0.000)
Text_Dispersion_Others	-2.425* (0.024)	-2.303* (0.032)	-2.121* (0.046)	-2.216* (0.037)
Text_Dispersion_Others ²	-40.919*** (0.000)	-42.512*** (0.000)	-31.980** (0.005)	-31.824** (0.005)
Average_Rating_Self	-0.262*** (0.000)			-0.280*** (0.000)
Average_Rating_Others	-0.032 (0.771)			-0.030 (0.796)
Average_Sentiment_Self		-0.667** (0.001)	-0.643** (0.002)	
Average_Sentiment_Others		0.228 (0.433)	0.315 (0.304)	
Rating_Entropy_Self			1.947*** (0.000)	1.867*** (0.000)
Rating_Entropy_Self ²			-3.976*** (0.000)	-4.091*** (0.000)
Rating_Entropy_Others			0.415* (0.032)	0.331* (0.095)
Rating_Entropy_Others ²			-2.595*** (0.000)	-2.599*** (0.000)
Experience	0.674*** (0.000)	0.683*** (0.000)	0.547*** (0.000)	0.544*** (0.000)
Price_Difference	0.027 (0.746)	0.029 (0.733)	0.049 (0.561)	0.054 (0.521)
Average_LogRevenue	0.013 (0.220)	0.013 (0.218)	0.015 (0.143)	0.016 (0.139)
Constant	3.030* (0.019)	1.886* (0.100)	2.204* (0.054)	3.510* (0.008)
Fixed Effects	Yes	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes	Yes
Observations	4187	4187	4187	4187
Adjusted R ²	0.612	0.612	0.618	0.619

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.15 Installer-level analysis with variables derived from text analysis

Variables	(I) Market Transaction	(II) Market Transaction	(III) Market Transaction	(IV) Market Transaction
Rating_Entropy_Mkt			0.859*** (0.000)	0.903*** (0.000)
Rating_Entropy_Mkt ²			-1.505*** (0.000)	-1.491*** (0.001)
Average_Rating_Mkt	-0.128 (0.312)			-0.067 (0.620)
Average_Sentiment_Mkt		-0.952** (0.009)	-0.869* (0.018)	
Average_Experience	0.030 (0.810)	-0.018 (0.883)	-0.067 (0.588)	-0.027 (0.828)
Price_Difference_Mkt	-0.031 (0.887)	-0.016 (0.940)	-0.084 (0.697)	-0.093 (0.668)
Market_LogRevenue	-0.002 (0.888)	-0.001 (0.918)	-0.006 (0.684)	-0.006 (0.649)
Text_Dispersion_Mkt	-0.023 (0.990)	-0.615 (0.741)	-2.396 (0.206)	-1.792 (0.344)
Text_Dispersion_Mkt ²	-15.669 (0.172)	-12.586 (0.273)	-3.867 (0.738)	-6.988 (0.546)
Constant	0.037 (0.996)	0.330 (0.966)	0.119 (0.988)	-0.365 (0.963)
Fixed Effects	Yes	Yes	Yes	Yes
Weighted State Dummies	Yes	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes	Yes
Observations	616	616	616	616
Adjusted R ²	0.756	0.758	0.764	0.761

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.16 Market-level analysis with variables derived from text analysis

Installer-level dynamic panel			
	Order	z	<i>p</i> -value
H_0 : No correlation between $\Delta_{i,t}$ and $\Delta_{i,t-1}$		-24.643	0.000
H_0 : No correlation between $\Delta_{i,t}$ and $\Delta_{i,t-2}$		0.412	0.680
Sargan Test for overidentifying restriction		$\chi^2 = 997.639$	0.079
Market-level dynamic panel			
	Order	z	<i>p</i> -value
H_0 : No correlation between $\Delta_{i,t}$ and $\Delta_{i,t-1}$		-13.148	0.000
H_0 : No correlation between $\Delta_{i,t}$ and $\Delta_{i,t-2}$		-0.712	0.477
Sargan Test for overidentifying restriction		$\chi^2 = 403.203$	0.058

Table EC.17 Dynamic panel specification checks

Variables	(I) Installer's Activity Level	(II) Installer's Activity Level	(III) Installer's Activity Level	(IV) Installer's Activity Level
Installer Activity Level _t	0.538*** (0.000)	0.528*** (0.000)	0.527*** (0.000)	0.519*** (0.000)
Rating_Entropy_Self			1.677** (0.007)	1.600* (0.010)
Rating_Entropy_Self ²			-2.93* (0.013)	-2.797* (0.015)
Rating_Entropy_Others		0.521* (0.041)		0.428* (0.090)
Rating_Entropy_Others ²		-1.616* (0.041)		-1.510* (0.056)
Average_Rating_Self	0.073 (0.394)	0.078 (0.373)	0.089 (0.343)	0.089 (0.350)
Average_Rating_Others	-0.138 (0.293)	-0.056 (0.684)	-0.101 (0.444)	-0.041 (0.768)
Experience	0.426 (0.132)	0.287 (0.251)	0.315 (0.234)	0.203 (0.388)
Price_Difference	-0.048 (0.759)	-0.034 (0.832)	-0.042 (0.777)	-0.028 (0.855)
Average_LogRevenue	0.003 (0.772)	0.004 (0.737)	0.003 (0.772)	0.004 (0.741)
Fixed Effects	Yes	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes	Yes
Observations	3872	3872	3872	3872

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.18 Robustness check - installer-level dynamic panels

Variables	(I) Market Transaction	(II) Market Transaction
Rating_Entropy_Mkt		0.628** (0.002)
Rating_Entropy_Mkt ²		-1.176*** (0.000)
Average_Rating_Mkt	0.090 (0.385)	0.049 (0.506)
Market_LogRevenue	-0.011 (0.478)	-0.013 (0.408)
Average_Experience	-0.010 (0.867)	-0.023 (0.692)
Price_Difference_Mkt	0.197 (0.503)	0.101 (0.728)
Market_Transaction _t	0.229** (0.001)	0.213** (0.003)
Market_Transaction _{t-1}	0.112* (0.012)	0.111* (0.010)
Market_Transaction _{t-2}	0.118* (0.014)	0.115* (0.015)
Fixed Effects	Yes	Yes
Seasonality	Yes	Yes
Observations	423	423

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.19 Robustness check - market-level dynamic panels

Variables	(I) Installer's Activity Level	(II) Installer's Activity Level	(III) Installer's Activity Level	(IV) Installer's Activity Level
Rating_Entropy_Others_1	0.624** (0.004)			
Rating_Entropy_Others_2	-2.024*** (0.000)			
Rating_Entropy_Self_1		1.570*** (0.000)		
Rating_Entropy_Self_2		-3.112*** (0.000)		
Rating_Entropy_Others_1			0.820** (0.001)	
Rating_Entropy_Others_2			-1.923* (0.029)	
Rating_Entropy_Others_3			-1.317• (0.083)	
Rating_Entropy_Self_1				2.498*** (0.000)
Rating_Entropy_Self_2				-3.542*** (0.000)
Rating_Entropy_Self_3				-2.071• (0.072)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes	Yes
Observations	4187	4187	4187	4187
Adjusted R ²	0.606	0.610	0.607	0.611

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.20 Alternative inverted-U testing: spline regressions (installer-level)

Variables	(I) Market Transaction	(II) Market Transaction
Rating_Entropy_Mkt_1	0.833*** (0.000)	
Rating_Entropy_Mkt_2	-1.088* (0.013)	
Rating_Entropy_Mkt_1		2.602*** (0.000)
Rating_Entropy_Mkt_2		-3.909** (0.002)
Rating_Entropy_Mkt_3		-0.575• (0.074)
Controls	Yes	Yes
Fixed Effects	Yes	Yes
Weighted State Dummies	Yes	Yes
Seasonality	Yes	Yes
Observations	619	619
Adjusted R ²	0.742	0.756

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.21 Alternative inverted-U testing: spline regressions (market-level)

Variables	(I) Installer's Activity	(II) Installer's Activity	(III) Installer's Activity
Text_Dispersion_Self		0.625 (0.434)	0.593 (0.471)
Text_Dispersion_Self ²		-26.554*** (0.000)	-25.783*** (0.000)
Text_Dispersion_Others	-2.407* (0.016)	-1.736* (0.087)	-1.589 (0.116)
Text_Dispersion_Others ²	-37.872** (0.003)	-30.952* (0.013)	-30.300* (0.014)
Average_Rating_Self	-0.265** (0.007)	-0.284** (0.005)	
Average_Rating_Others	-0.040 (0.722)	-0.029 (0.794)	
Average_Sentiment_Self			-0.646* (0.014)
Average_Sentiment_Others			0.302 (0.309)
Rating_Entropy_Self	2.372*** (0.000)	1.949*** (0.000)	2.302*** (0.000)
Rating_Entropy_Self ²	-4.359*** (0.000)	-4.174*** (0.000)	-4.061*** (0.000)
Rating_Entropy_Others	0.438* (0.033)	0.310 (0.136)	0.393*** (0.000)
Rating_Entropy_Others ²	-2.859*** (0.000)	-2.572*** (0.000)	-2.571*** (0.000)
Price_Difference	0.012 (0.899)	0.060 (0.535)	0.055 (0.573)
Average_LogRevenue	0.015 (0.177)	0.015 (0.180)	0.015 (0.186)
Experience	0.604*** (0.000)	0.530*** (0.000)	0.533*** (0.000)
Constant	2.491** (0.003)	2.865** (0.001)	1.559** (0.001)
Fixed Effects	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes
Observations	4187	4187	4187
Adjusted R ²	0.615	0.618	0.618

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

The text-based dispersion variables are derived by taking mean, instead of median, of pairwise cosine distances.

Table EC.22 Robustness check - installer-level analysis with text-based dispersion from mean cosine distances

Variables	(I) Market Transaction	(II) Market Transaction	(III) Market Transaction
Rating_Entropy_Mkt			0.746** (0.003)
Rating_Entropy_Mkt ²			-1.319** (0.001)
Average_Rating_Mkt	-0.104 (0.400)		
Average_Sentiment_Mkt		-0.869* (0.034)	-0.807* (0.049)
Average_Experience	-0.008 (0.951)	-0.049 (0.706)	-0.083 (0.535)
Price_Difference_Mkt	-0.002 (0.995)	0.014 (0.959)	-0.049 (0.855)
Market_LogRevenue	-0.004 (0.772)	-0.003 (0.802)	-0.007 (0.591)
Text_Dispersion_Mkt	0.457 (0.741)	-0.145 (0.918)	-1.390** (0.001)
Text_Dispersion_Mkt ²	-73.006*** (0.000)	-68.863*** (0.000)	-58.375** (0.002)
Constant	1.801 (0.107)	1.972* (0.049)	1.940* (0.048)
Fixed Effects	Yes	Yes	Yes
Weighted State Dummies	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes
Observations	619	619	619
Adjusted R ²	0.763	0.765	0.769

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

The text-based dispersion variables are derived by taking mean, instead of median, of pairwise cosine distances.

Table EC.23 Robustness check - market-level analysis with text-based dispersion from mean cosine distances

Variables	(1) Market Transaction	(2) Market Transaction	(3) Market Transaction
Rating_Entropy_Mkt	1.214** (0.002)	1.300*** (0.000)	1.644** (0.002)
Rating_Entropy_Mkt ²	-3.232*** (0.000)	-2.218*** (0.000)	-2.014* (0.025)
Average_Rating_Mkt	0.068 (0.703)	-0.056 (0.799)	0.583• (0.083)
Market_LogRevenue	-0.024 (0.217)	-0.007 (0.754)	-0.054 (0.164)
Average_Experience	-0.197 (0.378)	0.296 (0.101)	0.502 (0.140)
Price_Difference_Mkt	0.391 (0.291)	-0.217 (0.456)	-0.455 (0.250)
Constant	1.588 (0.142)	0.429 (0.705)	-1.479 (0.420)
Fixed Effects	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes
Observations	390	306	146
Adjusted R ²	0.779	0.757	0.743

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Column (I): restrict sample to reviews count >= 8

Column (II): resample to bi-monthly

Column (III): resample to quarterly

Table EC.24 Robustness check: subset with only high review count or altered sampling frequency (bi-monthly and quarterly)

Variables	(I) Installer's Activity Level	(II) Installer's Activity Level	(III) Installer's Activity Level
Rating_HHI_Self	24.449*** (0.000)		24.085*** (0.000)
Rating_HHI_Self ²	-15.698*** (0.000)		-15.478*** (0.000)
Rating_HHI_Others		25.974*** (0.000)	25.929*** (0.000)
Rating_HHI_Others ²		-16.829*** (0.000)	-16.776*** (0.000)
Average_Rating_Self	-0.062 (0.603)	-0.187• (0.089)	-0.045 (0.705)
Average_Rating_Others	-0.070 (0.486)	0.328** (0.008)	0.296* (0.016)
Experience	0.625*** (0.000)	0.600*** (0.000)	0.392** (0.008)
Price_Difference	-0.034 (0.746)	-0.048 (0.648)	-0.072 (0.485)
Average_LogRevenue	0.019• (0.095)	0.017 (0.131)	0.014 (0.207)
Constant	-7.571*** (0.000)	-9.584** (0.000)	-18.313*** (0.000)
Fixed Effects	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes
Observations	3877	3877	3877
Adjusted R ²	0.620	0.624	0.629

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.25 Installer-level analysis (Activity vs. HHI)

Variables	(I) Market Transaction	(II) Market Transaction
Rating_HHI_Mkt		15.178*** (0.000)
Rating_HHI_Mkt ²		-10.663*** (0.000)
Average_Rating_Mkt	-0.194 (0.143)	0.100 (0.446)
Average_Experience	0.160 (0.240)	0.532** (0.008)
Price_Difference_Mkt	-0.372 (0.211)	-0.537 (0.134)
Market_LogRevenue	-0.007 (0.616)	-0.015 (0.359)
Constant	1.366• (0.053)	-5.292*** (0.000)
Fixed Effects	Yes	Yes
Weighted State Dummies	Yes	Yes
Seasonality	Yes	Yes
Observations	619	467
Adjusted R ²	0.735	0.771

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.26 Market-level analysis (Transaction vs. HHI)

Variables	(I) Installer's Activity Level	(II) Installer's Activity Level	(III) Installer's Activity Level
Rating_Bin	-0.221*** (0.000)	-0.223*** (0.000)	-0.145 (0.122)
Rating_Bin × Rating_Entropy_Self		0.094 (0.742)	0.762 (0.237)
Rating_Bin × Rating_Entropy_Self ²			-1.908 (0.247)
Rating_Entropy_Self	2.228*** (0.000)	2.178*** (0.000)	2.067*** (0.000)
Rating_Entropy_Self ²	-3.870*** (0.000)	-3.821*** (0.000)	-3.586*** (0.000)
Rating_Entropy_Others	0.630*** (0.000)	0.626*** (0.000)	0.622*** (0.000)
Rating_Entropy_Others ²	-3.104*** (0.000)	-3.009*** (0.000)	-2.988*** (0.000)
Experience	0.619*** (0.000)	0.623*** (0.000)	0.611*** (0.000)
Price_Difference	0.009 (0.915)	0.010 (0.905)	0.019 (0.824)
Average_LogRevenue	0.019• (0.066)	0.019• (0.066)	0.019• (0.065)
Constant	1.449 (0.191)	1.435• (0.197)	1.444 (0.194)
Fixed Effects	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes
Observations	4187	4187	4187
Adjusted R ²	0.614	0.614	0.614

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.27 Installer-level analysis with rating-bins

Variables	Installer's Activity Level
Average_Rating_Self	-0.195* (0.017)
Average_Rating_Others	-0.043 (0.724)
Rating_Entropy_Self	2.289*** (0.000)
Rating_Entropy_Self ²	-3.898*** (0.000)
Rating_Entropy_Others	0.317 (0.122)
Rating_Entropy_Others ²	-2.497*** (0.000)
Experience	0.523*** (0.000)
Price_Difference	-0.016 (0.854)
Average_LogRevenue	0.015 (0.178)
Proportion_No-Rating_Self	0.396*** (0.001)
Constant	2.284• (0.086)
Fixed Effects	Yes
State Dummies	Yes
Seasonality	Yes
Observations	3872
Adjusted R ²	0.628

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.28 Installer-level analysis with proportion of no-rating transactions as a control

Variables	Market Transaction
Average_Rating_Mkt	-0.047 (0.739)
Rating_Entropy_Mkt	0.868*** (0.001)
Rating_Entropy_Mkt ²	-0.913* (0.031)
Average_Experience	-0.056 (0.655)
Price_Difference_Mkt	0.295 (0.212)
Market_LogRevenue	-0.001 (0.937)
Average_Proportion_No-Rating_Mkt	0.573*** (0.000)
Constant	3.959 (0.644)
Fixed Effects	Yes
Weighted State Dummies	Yes
Seasonality	Yes
Observations	536
Adjusted R ²	0.801

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.29 Market-level analysis with proportion of no-rating transactions as a control

Variables	Installer's Activity Level
Average_Rating_Self	-0.242** (0.002)
Average_Rating_Others	-0.063 (0.607)
WA [Rating_Entropy_Self]	1.990*** (0.000)
WA [Rating_Entropy_Self] ²	-4.308*** (0.000)
Rating_Entropy_Others	0.367• (0.073)
Rating_Entropy_Others ²	-2.584*** (0.000)
Experience	0.665*** (0.000)
Price_Difference	0.003 (0.975)
Average_LogRevenue	0.013 (0.221)
Constant	2.362• (0.076)
Fixed Effects	Yes
State Dummies	Yes
Seasonality	Yes
Observations	3872
Adjusted R ²	0.626

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.30 Installer-level analysis using the weighted average installer-level rating entropy

Variables	Market Transaction
Average_Rating_Mkt	-0.063 (0.656)
WA [Rating_Entropy_Mkt]	0.600* (0.038)
WA [Rating_Entropy_Mkt] ²	-1.068* (0.025)
Average_Experience	-0.055 (0.667)
Price_Difference_Mkt	0.096 (0.682)
Market_LogRevenue	0.002 (0.889)
Constant	6.798 (0.436)
Fixed Effects	Yes
Weighted State Dummies	Yes
Seasonality	Yes
Observations	536
Adjusted R ²	0.792

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.31 Market-level analysis using the weighted average market-level rating entropy

Variables	(I)	(II)	(III)
	Installer's Activity Level (Market = 4)	Installer's Activity Level (Market = 36)	Installer's Activity Level (Market = 17)
Rating_Entropy_Self	2.343** (0.009)	4.680*** (0.000)	5.139* (0.025)
Rating_Entropy_Self ²	-6.289** (0.002)	-15.580*** (0.000)	-10.009* (0.011)
Rating_Entropy_Others	-2.546** (0.005)	-4.183*** (0.000)	2.448* (0.009)
Rating_Entropy_Others ²	-4.907 (0.303)	-18.024*** (0.000)	6.438 (0.243)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes
Seasonality	Yes	Yes	Yes
Estimation Sample	693	647	391
Adjusted R ²	0.626	0.653	0.703

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.32 Installer-level analysis for individual markets

Variables	Installer's Activity Level
Rating_Entropy_Self	2.480*** (0.000)
Rating_Entropy_Self ²	-4.430*** (0.000)
Rating_Entropy_Others	0.558** (0.005)
Rating_Entropy_Others ²	-2.976*** (0.000)
Controls	Yes
Fixed Effects	Yes
Market Dummies	Yes
Seasonality	Yes
Observations	4187
Adjusted R ²	0.614

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.33 Installer-level analysis with market dummies (instead of state dummies)

Variables	Installer Transaction
Average_Rating_Self	-0.019 (0.597)
Average_Rating_Others	-0.028 (0.602)
Rating_Entropy_Self	0.362* (0.011)
Rating_Entropy_Self ²	-0.597* (0.033)
Rating_Entropy_Others	0.062 (0.489)
Rating_Entropy_Others ²	-0.678*** (0.000)
Experience	0.169** (0.351)
Price_Difference	0.039 (0.291)
Average_LogRevenue	0.003 (0.567)
Constant	0.157 (0.778)
Fixed Effects	Yes
State Dummies	Yes
Seasonality	Yes
Observations	3426
Adjusted R ²	0.490

p-value in parentheses: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1

Table EC.34 Installer-level analysis (Matching vs. Entropy)

Variables	N	Mean	Std. Dev.	Min	Max
Rating_Entropy_Self	4,384	0.101	0.221	0.000	1.210
Rating_Entropy_Others	4,384	0.230	0.182	0.000	1.071
Average_Rating_Self	4,384	4.536	1.304	0.000	5.000
Average_Rating_Others	4,384	4.878	0.202	1.000	5.000
Experience	4,384	1.812	0.890	0.000	3.714
Price_Difference	4,384	-0.032	0.397	-2.171	3.212
Average_LogRevenue	4,384	10.805	0.400	9.570	16.258
Installer_Activity	4,384	2.417	1.692	0.000	7.257

Table EC.35 Summary statistics - installer-level (non-demeaned entropy)

Variables	N	Mean	Std. Dev.	Min	Max
Rating_Entropy_Mkt	748	0.199	0.236	0.000	1.055
Market_LogRevenue	748	15.795	2.221	10.380	22.303
Average_Rating_Mkt	748	4.869	0.242	3.000	5.000
Average_Experience	748	1.474	1.085	0.000	3.332
Price_Difference_Mkt	748	-0.014	0.126	-0.504	1.312
Market_Transaction	748	0.965	1.030	0.000	4.522

Table EC.36 Summary statistics - market-level (non-demeaned entropy)

References

Davidson M (2019) What Neighborhood Is This? Estimating Material Based Neighborhood Boundaries Using Single-Family Homes. <https://ssrn.com/abstract=3485994> Working paper, University of Michigan at Ann Arbor.

Kanagala HK, Krishnaiah VJR (2016) A comparative study of K-Means, DBSCAN and OPTICS. 2016 International Conference on Computer Communication and Informatics (ICCCI), 1–6 (IEEE).

Loughran, T, McDonald, B (2011) When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*. 66(1): 35-65.

Teimouri M, Indahl UG, Tveite H (2016) A method to detect inactive periods in animal movement using density-based clustering. *Appl. Geogr.* 73:102–112.

US Census Bureau (2019) Gazetteer files. <https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2018.html>. Accessed June 12, 2023.