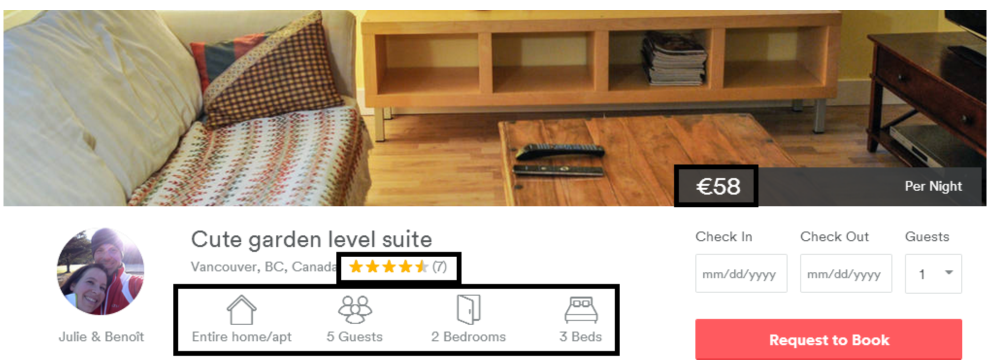
Online Appendix

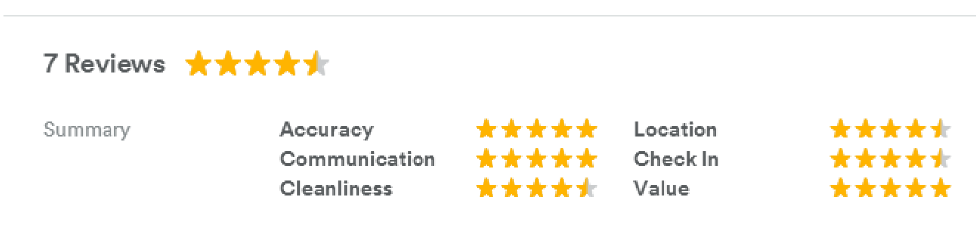
Can Information Reduce Ethnic Discrimination?
Evidence from Airbnb
Morgane Laouénan, Roland Rathelot

A Additional Figures

Figure A1: Example of a listing's dashboard, with the most salient information



Information on listings' ratings



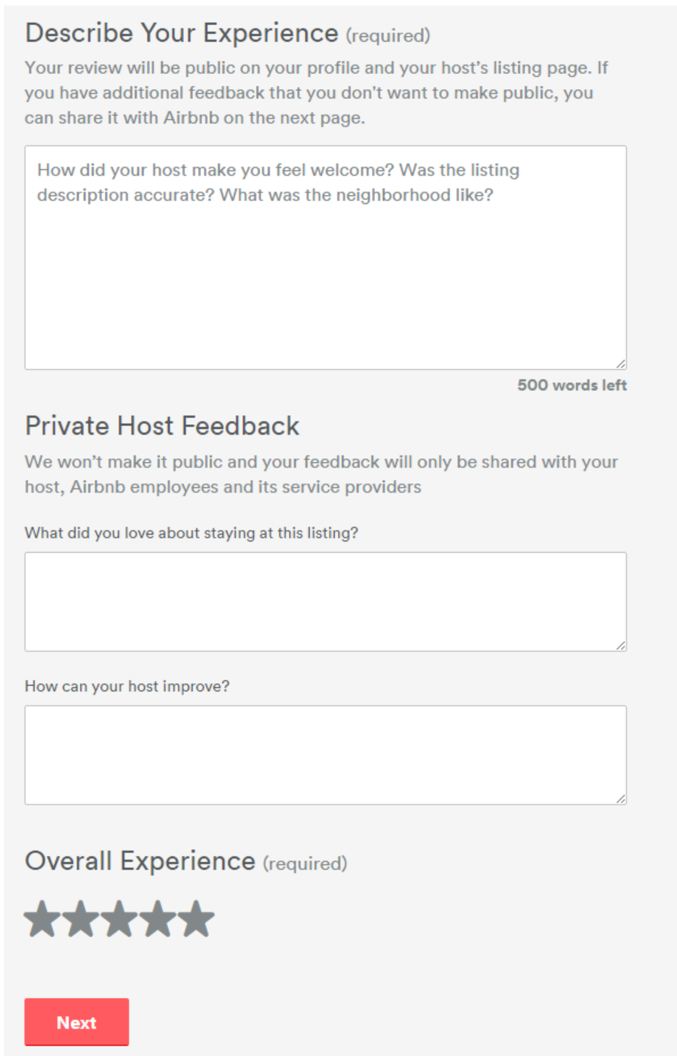
For illustrative purposes only, screenshot of the Airbnb platform captured by the author on May 2016.

Information on listings' amenities

The Space	Property type: **House** Accommodates: **5** Bedrooms: **2** Bathrooms: **1**	Beds: **3** Check In: **3:00 AM** Check Out: **11:00 AM**
Amenities	**Kitchen** **Internet** **TV** **Essentials** **Heating** ~~Air Conditioning~~ **Washer** **Dryer** **Free Parking on Premises** **Wireless Internet** **Cable TV** ~~Breakfast~~ ~~Pets Allowed~~ **Family/Kid Friendly** ~~Suitable for Events~~	~~Smoking Allowed~~ ~~Wheelchair Accessible~~ ~~Elevator in Building~~ ~~Indoor Fireplace~~ ~~Buzzer/Wireless Intercom~~ ~~Doorman~~ ~~Pool~~ ~~Hot Tub~~ ~~Gym~~ ~~Smoke Detector~~ ~~Carbon Monoxide Detector~~ ~~First Aid Kit~~ ~~Safety Card~~ ~~Fire Extinguisher~~
Prices	Extra people: **No Charge** Security Deposit: **$92** Weekly Price: **$412 /week**	Monthly Price: **$1373 /month** Cancellation: **Moderate**

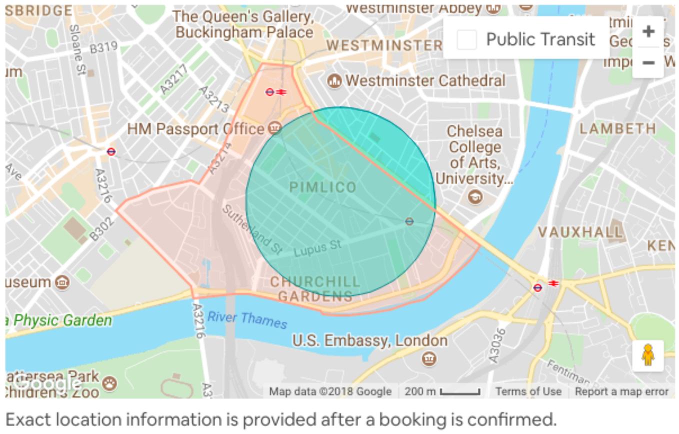
For illustrative purposes only, screenshot of the Airbnb platform captured by the author on May 2016.

Peer-reviewing system



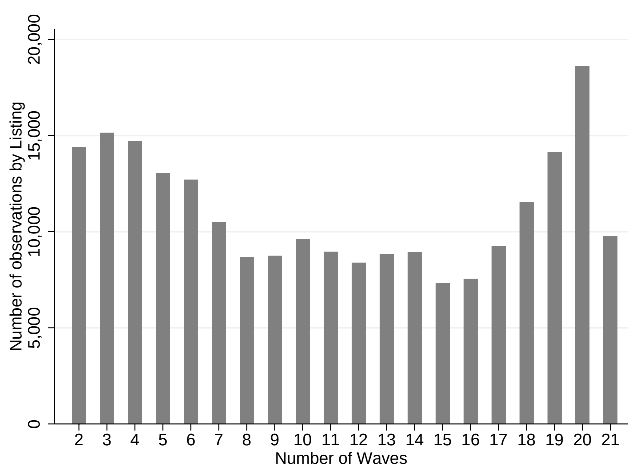
For illustrative purposes only, screenshot of the Airbnb platform captured by the author on May 2016.

Information on listings' locations



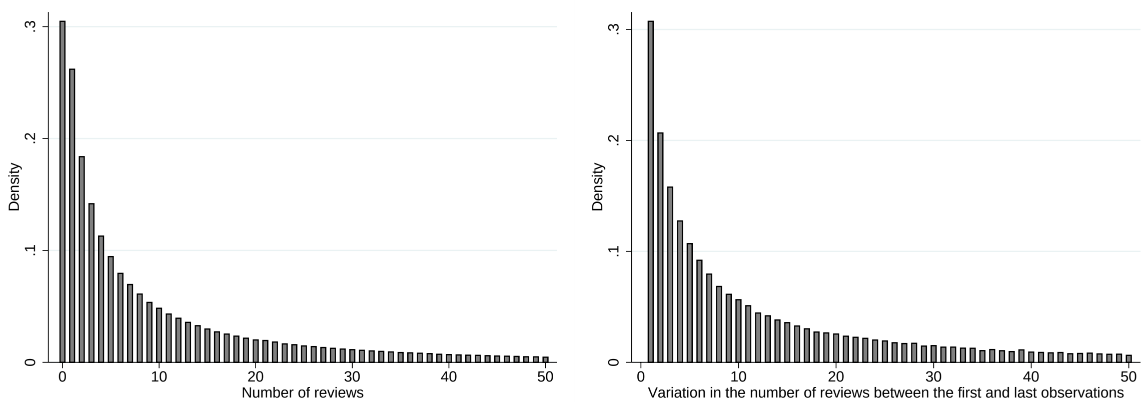
In this example, the listing is shown to be located in the neighbourhood of Pimlico, in London, and the area of the .6-mile-radius circle is almost entirely in that neighbourhood. For illustrative purposes only, screenshot of the Airbnb platform captured by the author on September 2018.

Figure A2: Number of observations by listing



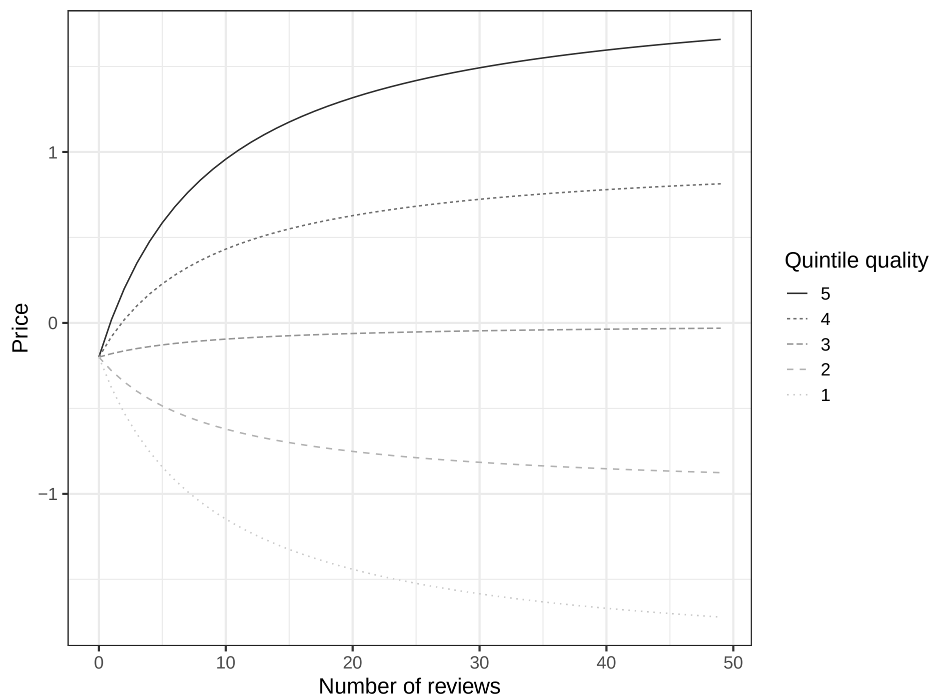
Notes: This figure shows the number of observations by listing depending on the number of waves (x-axis). It starts at 2 waves as we restrict the sample to listings that have gained at least one review over the observation period.

Figure A3: Distribution of the number of reviews (left) and of the longitudinal variation in the number of reviews within a property (right)



Notes: The left figure shows the distribution of the number of reviews. The right figure shows the distribution of the longitudinal variation in the number of reviews within a property. Both figures are right truncated with a maximum of 50 reviews. The sample is restricted to listings that have gained at least one review over the observation period.

Figure A4: Illustration of the conceptual framework: Prices with the number of reviews, by unobservable quality



Note: This illustrative graph displays $(K\nu - \rho/5)/(K+\rho)$ as function of K, where ν takes values in $\{-2, 1, 0, 1, 2\}$ and $\rho = 8$.

B Additional tables

Table A1: Number of observations & listings by city

City	Observations		Listings	
	#	share	#	share
Amsterdam	51,189	2.07	6,122	2.77
Barcelona	173,180	7.00	14,529	6.58
Berlin	151,887	6.14	13,948	6.31
Boston	43,637	1.76	4,330	1.96
Chicago	42,990	1.74	4,408	2.00
Florence	67,106	2.71	4,967	2.25
London	264,705	10.70	23,889	10.81
Los Angeles	159,228	6.43	15,182	6.87
Madrid	65,753	2.66	5,359	2.43
Marseille	55,643	2.25	4,921	2.23
Miami	67,373	2.72	6,383	2.89
Milan	85,365	3.45	8,360	3.78
Montreal	69,331	2.80	6,525	2.95
New York City	349,471	14.12	31,717	14.36
Paris	464,493	18.77	39,026	17.66
Rome	152,644	6.17	11,547	5.23
San Francisco	108,144	4.37	10,148	4.59
Toronto	56,843	2.30	5,359	2.43
Vancouver	45,569	1.84	4,219	1.91

Notes: The table shows the number of observations (column 1), its share (column 2) and the number of listings (column 3), and its share (column 4) for each of the 19 cities included in our dataset. The sample is restricted to listings that have gained at least one review over the observation period. The total number of observations is 2,474,551 and the total number of listings is 220,939.

Table A2: Collection dates of waves

Wave	Collection date
1	15 June 2014
2	8 July 2014
3	28 July 2014
4	11 August 2014
5	25 August 2014
6	8 September 2014
7	25 September 2014
8	15 October 2014
9	5 November 2014
10	25 November 2014
11	15 December 2014
12	7 January 2015
13	13 January 2015
14	3 February 2015
15	4 March 2015
16	25 March 2015
17	13 April 2015
18	4 May 2015
19	26 May 2015
20	15 June 2015
21	11 November 2017

Table A3: Summary statistics: Property & host characteristics

	Full Sample	Listings that have gained at least one review over the period
Type of property		
Entire property	0.665	0.705
Flat	0.802	0.843
House	0.064	0.106
Loft	0.016	0.019
Size		
Person capacity	3.148	3.211
Number of bedrooms	1.252	1.244
Number of bathrooms	1.162	1.153
Type of bed		
Couch	0.005	0.006
Airbed	0.003	0.003
Sofa	0.026	0.033
Futon	0.009	0.012
Real bed	0.958	0.946
Amenities		
Cable TV	0.290	0.346
Wireless	0.901	0.899
Heating	0.876	0.887
AC	0.395	0.380
Elevator	0.341	0.340
Wheelchair accessible	0.077	0.098
Doorman	0.080	0.096
Fireplace	0.077	0.080
Washer	0.697	0.697
Dryer	0.402	0.388
Parking	0.200	0.179
Gym	0.072	0.064
Pool	0.063	0.054
Buzzer	0.293	0.386
Hot Tub	0.069	0.069
Services		
Breakfast served	0.111	0.091
Family/Kids friendly	0.466	0.448
Suitable for events	0.045	0.052
Rules & Extras		
Additional people	0.469	0.646
Price per additional people	7.389	7.911

(Continued on next page)

Table A3: Summary statistics: Property & host characteristics

Smoking allowed	0.133	0.144
Pets allowed	0.125	0.131
Host Characteristics		
Has multiple properties	0.356	0.345
Member since 2008	0.001	0.001
Member since 2009	0.006	0.009
Member since 2010	0.019	0.033
Member since 2011	0.063	0.107
Member since 2012	0.126	0.209
Member since 2013	0.166	0.263
Member since 2014	0.198	0.291
Member since 2015	0.068	0.075
Number of languages spoken	0.851	1.408
Superhost	0.023	0.053
Verified email	0.620	0.960
Verified offline	0.320	0.525
Verified phone	0.281	0.428
Number of Facebook friends	153.567	237.714
Number of words in description	217.000	240.168
Number of words in profile	49.560	49.822
Number of pictures	13.058	13.921
Number of pictures by professionals	0.979	0.703
N	663,090	220,939

Notes: The left column displays the mean of each characteristics in the full sample, while the right column focuses on the sub-sample of listings that have gained at least one review over the observation period (between the first and the last waves).

Table A4: Log daily rate

	(1)	(2)
Shared Flat	-0.828***	-0.715***
	(0.004)	(0.007)
Person Capacity (> 2)	0.164***	0.175***
	(0.004)	(0.005)
# bedrooms	0.273***	0.293***
	(0.003)	(0.004)
# bathrooms	0.167***	0.143***
	(0.004)	(0.005)
Flat	-0.154***	-0.179***
	(0.009)	(0.013)
House or Loft	-0.159***	-0.061***
	(0.010)	(0.014)
Couch	-0.193***	-0.165***
	(0.014)	(0.015)
Airbed	-0.192***	-0.125***
	(0.027)	(0.025)
Sofa	-0.175***	-0.166***
	(0.006)	(0.009)
Futon	-0.158***	-0.116***
	(0.011)	(0.010)
Cable TV	0.141***	0.098***
	(0.003)	(0.004)
Wireless	0.033***	0.021***
	(0.005)	(0.006)
Heating	-0.019***	0.001
	(0.005)	(0.007)
AC	0.134***	0.113***
	(0.004)	(0.006)
Elevator	0.093***	0.084***
	(0.003)	(0.005)
Wheelchair Accessible	-0.039***	-0.007

(Continued on next page)

Table A4: Log daily rate

	(0.004)	(0.005)
Doorman	0.102***	0.036***
	(0.005)	(0.007)
Fireplace	0.158***	0.117***
	(0.005)	(0.005)
Washer	-0.021***	0.020***
	(0.004)	(0.006)
Dryer	0.146***	0.094***
	(0.003)	(0.004)
Parking	-0.133***	0.028***
	(0.004)	(0.005)
Gym	0.062***	0.042***
	(0.007)	(0.009)
Pool	0.083***	0.082***
	(0.007)	(0.012)
Buzzer	0.050***	0.008**
	(0.003)	(0.003)
Hot Tub	0.012**	0.010
	(0.005)	(0.006)
Breakfast served	0.005	0.033***
	(0.004)	(0.005)
Family/Kids Friendly	0.014***	0.033***
	(0.003)	(0.003)
Suitable for events	0.072***	0.062***
	(0.006)	(0.008)
Additional People	-0.034***	-0.013***
	(0.002)	(0.002)
Price per Additional People	0.001***	-0.001***
	(0.000)	(0.000)
Cancellation Policy	0.040***	0.015***
	(0.002)	(0.002)
Smoking Allowed	-0.123***	-0.093***
	(0.004)	(0.004)

(Continued on next page)

Table A4: Log daily rate

Pets Allowed	-0.024***	-0.027***
	(0.004)	(0.004)
Host has multiple properties	0.050***	0.024***
	(0.003)	(0.004)
Member since 2009	0.145***	0.118***
	(0.019)	(0.021)
Member since 2010	0.121***	0.097***
	(0.015)	(0.016)
Member since 2011	0.098***	0.087***
	(0.014)	(0.015)
Member since 2012	0.077***	0.070***
	(0.014)	(0.015)
Member since 2013	0.076***	0.066***
	(0.014)	(0.015)
Member since 2014	0.051***	0.048***
	(0.014)	(0.014)
Member since 2015	0.052***	0.047***
	(0.014)	(0.015)
Superhost	0.023***	0.014***
	(0.005)	(0.005)
Verified Email	-0.022***	-0.000
	(0.007)	(0.007)
Verified Offline	0.013***	0.005*
	(0.003)	(0.003)
Verified Phone	0.003	0.004
	(0.007)	(0.012)
Nber of Facebook friends	0.000***	0.000
	(0.000)	(0.000)
Nber of words in Description	-0.000***	-0.000***
	(0.000)	(0.000)
Nber of words in Profile	-0.000***	-0.000***
	(0.000)	(0.000)
Nber of Languages	-0.005***	-0.005***

(Continued on next page)

Table A4: Log daily rate

	(0.001)	(0.001)
Nber of words in Rules	-0.000***	-0.000**
	(0.000)	(0.000)
Nber of pictures	0.003***	0.003***
	(0.000)	(0.000)
Nber of pictures taken by professionals	0.001***	0.002***
	(0.000)	(0.000)
Nber of picture changes	-0.034***	-0.037***
	(0.002)	(0.002)
City-wave FE	Yes	Yes
Neighbourhood FE	No	Yes
Block FE	No	Yes
Adj R^2	0.627	0.733
N obs.	2,474,551	2,474,551

Notes: OLS regression on the daily log-price. In column (2), neighbourhood and block fixed effects are included in the estimation. Robust standard errors clustered at the property level.

Table A5: Distribution of the last rating

	Obs	Share
3.5 stars	9,560	4.39%
4 stars	26,943	12.37%
4.5 stars	85,047	39.06%
5 stars	96,178	44.17%

Notes : The sample corresponds to listings for which last rating is observed. Listings with less than 3.5 stars are included in the first row.

Table A6: Number of neighbourhoods & blocks by city

City	# neighbourhoods	# Blocks
Amsterdam	45	101
Barcelona	70	82
Berlin	88	404
Boston	42	250
Chicago	75	242
Florence	18	102
London	150	838
Los Angeles	115	1267
Madrid	67	166
Marseille	61	615
Miami	80	430
Milan	25	155
Montreal	53	242
New York City	189	527
Paris	64	116
Rome	44	107
San Francisco	169	495
Toronto	115	286
Vancouver	34	307
Total	1,504	6,732

Notes: The definition of neighbourhoods directly comes from Airbnb while blocks are created via the approximate coordinates of the listing.

C Ethnic differences in the exit rate

In this section, we look at the issue of differential selection in the sample across ethnic groups and find that minority hosts are not more likely to leave the market than the majority. We consider that a listing i leaves the market at t if it is present at t, and not present anytime after t, and define $q_{it} = 1$ and 0 for $s \neq t$. Within the period of observation, 65,358 majority hosts (31.6%) and 4,777 minority hosts (33.6%) leave the platform. We regress q_{it} on a minority dummy, and control for property characteristics, ratings, neighbourhood fixed-effects, block fixed-effects and price.

Table A7 shows that the exit rate is similar for both groups when controlling for property characteristics, ratings, neighbourhood and block fixed-effects, price of the listing and number of reviews.

Table A7: Probability to leave the market at wave t

	(1)	(2)	(3)
Minority host	0.0004	0.0003	0.0003
	(0.0005)	(0.0005)	(0.0005)
Log-price		-0.0043***	-0.0053***
		(0.0003)	(0.0003)
Number of reviews			-0.0001***
			(0.0000)
Adj R^2	0.04	0.04	0.04
N obs.	2,474,551	2,474,551	2,474,551

Notes: OLS regressions of the probability to leave the market at wave t. Covariates include, aside from the ones mentioned in the table, neighbourhood fixed effects, block fixed-effects, property characteristics and ratings. Robust standard errors clustered at the property level.

D Pictures from which host ethnicity cannot be measured

Hosts can choose whether to post a picture of themselves on their host profiles. Popular alternative choices are pictures of their properties, pets, furniture, landscapes, etc. We identify pictures for which it was impossible to say anything about the ethnicity of anyone in the picture. In our data, there are 17% of such listings. If minorities are aware of the existence of discrimination on the platform, they might more often obfuscate their skin colour.

In this appendix, we try to understand the choice leading hosts to post or not their pictures. First, is the price set by no-person-picture hosts higher in neighbourhoods where the share of blacks is high? First, how do no-person-picture hosts set their price? Second, does the probability of having a no-person picture depend on the share of Blacks in the neighbourhood?

Table A8 first shows that, controlling for listing characteristics, hosts with a listing located in a neighbourhood with more Black hosts are not more or less likely to post a picture of themselves (Column 1). This result is at odds with a model of strategic hosts anticipating discrimination. Column 2 shows that, controlling for neighbourhoods and characteristics, hosts post very similar prices whether they choose to publish their pictures or not. Column 3 shows that the pattern does not seem to vary much with the ethnic composition of the neighbourhood. If anything, in areas with more Black hosts, the hosts that do not post their pictures have lower prices than those posting their pictures.

Table A8: Behaviour of hosts posting non-person pictures

	Non-person picture	Log-price	
	(1)	(2)	(3)
Local share of Blacks	0.007		
	(0.018)		
Non-person picture		0.002	0.004
		(0.003)	(0.004)
Non-person picture $\times$ share Blacks			-0.078
			(0.064)
Neighbourhood FE	No	Yes	Yes
Adj R^2	0.036	0.713	0.713
N obs.	2,466,726	2,466,726	2,466,726

Notes: OLS regressions. Aside from those mentioned in the Table, controls include city-wave FE, and property characteristics (see Table A4). Specifications in Columns 2 and 3 include neighbourhood FE and block FE, not in Column 1. Robust standard errors clustered at the listing level.

E Using a non-normal prior distribution of quality with a discrete signal

Assume that $\nu \sim \mathcal{B}(\alpha_\nu, \beta_\nu)$ (a Beta distribution). A Beta distribution looks more similar to the measures of quality that we have empirically: it is bounded and can be really skewed.

A single rating being a discrete signal, let's assume that we can model it as a draw in a $Binomial(n, \nu)$, where n depends on how much information a single rating contains (to what extent it is discrete). A rating takes values in $0 \ldots n$.

The pdf of the posterior distribution, given the observation of a rating r can be written as:

$$f(\nu|r) = \frac{P(r|\nu)f(\nu)}{\int P(r|\nu)f(\nu)d\nu}$$

Working on the numerator, we have:

$$P(r|\nu)f(\nu) = \binom{n}{r}\frac{\nu^r(1-\nu)^{n-r}\nu^{\alpha_\nu - 1}(1-\nu)^{\beta_\nu - 1}}{B(\alpha_\nu, \beta_\nu)}$$

where $B(.,.)$ is the beta function. This simplifies to:

$$P(r|\nu)f(\nu) = \binom{n}{r}\frac{\nu^{\alpha_\nu - 1 + r}(1-\nu)^{\beta_\nu - 1 + n - r}}{B(\alpha_\nu, \beta_\nu)}$$

Because $f(\nu|r)$ is a density, we know it is of integral one and thus should be equal to the density of a $\mathcal{B}(\alpha_\nu + r, \beta_\nu + n - r)$. We can also prove it by computing the integral of $P(r|\nu)f(\nu)$ wrt ν and computing $f(\nu|r)$ explicitly.

The expectation of ν conditional on r is therefore equal to:

$$E(\nu|r) = \frac{\alpha_\nu + r}{\alpha_\nu + \beta_\nu + n}$$

Now, suppose that we have K signals instead of just one. I also rescale the signal between 0 and 1 (which is the range of ν) and define $\bar{r} = \sum_k r_k/(nK)$, $\tilde{\alpha}_\nu = \alpha_\nu/n$ and $\tilde{\beta}\nu = \beta\nu/n$. We can show that the expectation depends only on $\bar{r}$:

$$E(\nu|\bar{r}, K) = \frac{\hat{\alpha}_\nu + K\bar{r}}{\hat{\alpha}_\nu + \hat{\beta}_\nu + K}$$

Dividing everything by n rescales the signal between 0 and 1 (which is the range of ν) and we obtain an expression that is exactly identical, up to a change in notations, to the one with normal distributions.

$$\mathbb{E}(\nu|\bar{r}, K, m) = \frac{\rho\bar{v} + K\bar{r}}{\rho + K}$$

with $\alpha_\nu = \rho\bar{v}$ and $\alpha_\nu + \beta_\nu = \rho$.

F Proofs for the identification results

F.1 Accurate beliefs

We start from the equation (2), reproduced here:

$$p = p_0 - \lambda\gamma m + \lambda\alpha w + \lambda\beta\zeta + \lambda\beta\frac{Kr + \rho\bar{v}_m}{K + \rho}$$

Assuming that we know ρ, the regression line of p_{it} conditional on $\mathcal{I}_{it}$, an information set made of $\frac{K_{it}}{K_{it}+\rho}$, $m_i\frac{K_{it}}{K_{it}+\rho}$, $\bar{r}_i\frac{K_{it}}{K_{it}+\rho}$, characteristics X_{it} and listing fixed effects μ_i:

$$\mathbb{E}(p_{it}|\mathcal{I}_{it}) = \mathbb{E}(p_0 - \lambda\gamma m + \lambda\alpha w_{it} + \lambda\beta\zeta_{it}|\mathcal{I}_{it}) + \mathbb{E}(\lambda\beta\frac{K_{it}r_{it} + \rho\bar{v}_m}{K_{it} + \rho}|\mathcal{I}_{it})$$

By assumption, the first term $\mathbb{E}(p_0 - \lambda\gamma m + \lambda\alpha w_{it} + \lambda\beta\zeta_{it}|\mathcal{I}_{it})$ is equal to a linear combination of the fixed effects and the observable characteristics.

$$\mathbb{E}(p_{it}|\mathcal{I}_{it}) = \mu_i + X_{it}\beta_x + \lambda\beta\mathbb{E}(\frac{K_{it}r_{it}}{K_{it} + \rho}|\mathcal{I}_{it}) + \lambda\beta\mathbb{E}(\frac{\rho\bar{v}_m}{K_{it} + \rho}|\mathcal{I}_{it})$$

At this stage, it is key that $\mathbb{E}(r_{it}|\mathcal{I}_{it}) = \mathbb{E}(r_{it}|\bar{r}_i)$. In particular, r_{it} does not depend on ethnicity conditional on $\bar{r}_i$.

$$\mathbb{E}(p_{it}|\mathcal{I}_{it}) = \mu_i + X_{it}\beta_x + \lambda\beta\frac{K_{it}}{K_{it} + \rho}\mathbb{E}(r_{it}|\bar{r}_i) + \lambda\beta\frac{\rho\bar{v}_0}{K_{it} + \rho} + \lambda\beta\frac{\rho}{K_{it} + \rho}(\bar{v}_1 - \bar{v}_0)m_i$$

As $\frac{\rho}{K_{it}+\rho} = 1 - \frac{K_{it}}{K_{it}+\rho}$:

$$\mathbb{E}(p_{it}|\mathcal{I}_{it}) = \mu_i + X_{it}\beta_x + \lambda\beta\frac{K_{it}}{K_{it} + \rho}[\mathbb{E}(r_{it}|\bar{r}_i) - \bar{v}_0] - \lambda\beta\frac{K_{it}}{K_{it} + \rho}(\bar{v}_1 - \bar{v}_0)m_i$$

Therefore, regressing p_{it} on $\frac{K_{it}}{K_{it}+\rho}\mathbb{1}\{\bar{r}_i = \bar{r}\}$, for all values $\bar{r}$ in the support of $\bar{r}_i$, and $\frac{K_{it}}{K_{it}+\rho}m_i$, conditional on listing fixed effects and characteristics X_{it}, will identify:

- $\beta_{\bar{r}} = \lambda\beta[\mathbb{E}(r_{it}|\bar{r}) - \bar{v}_0]$ for each value $\bar{r}$ in the support of $\bar{r}_i$.
- $\beta_m = -\lambda\beta(\bar{v}_1 - \bar{v}_0)$.

To finish the proof, note that ρ is identified non-parametrically within listing conditional on $\beta_{\bar{r}}$ and β_m. $\square$

F.2 Inaccurate beliefs

The first part of the proof directly follows the one of the case with accurate beliefs. For the second part, we apply the same reasoning, except that we attempt to characterise the regression line of p_{it} conditional on $\mathcal{I}'_{it}$, an information set equal to $\mathcal{I}_{it}$ minus $\bar{r}_i \frac{K_{it}}{K_{it}+\rho}$. The main difference is that Bayesian updating starts from the wrong bias $\tilde{v}_1$ instead of $\bar{v}_1$ for listings held by minority hosts. We obtain:

$$\mathbb{E}(p_{it}|\mathcal{I}'_{it}) = \mu_i + X_{it}\beta_x + \lambda\beta\mathbb{E}(\frac{K_{it}r_{it}}{K_{it}+\rho}|\mathcal{I}'_{it}) + \lambda\beta\mathbb{E}(\frac{\rho(\bar{v}_0 + m_i(\tilde{v}_1 - \bar{v}_0))}{K_{it}+\rho}|\mathcal{I}'_{it})$$

Now, note that $\mathbb{E}(r_{it}|\mathcal{I}'_{it}) = \mathbb{E}(r_{it}|m_i) = \bar{v}_{m_i} = \bar{v}_0 + m_i(\bar{v}_1 - \bar{v}_0)$.

$$\mathbb{E}(p_{it}|\mathcal{I}'_{it}) = \mu_i + X_{it}\beta_x + \lambda\beta\frac{K_{it}}{K_{it}+\rho}(\bar{v}_0 + m_i(\bar{v}_1 - \bar{v}_0)) + \lambda\beta\frac{\rho}{K_{it}+\rho}(\bar{v}_0 + m_i(\tilde{v}_1 - \bar{v}_0))$$

As $\frac{\rho}{K_{it}+\rho} = 1 - \frac{K_{it}}{K_{it}+\rho}$:

$$\mathbb{E}(p_{it}|\mathcal{I}_{it}) = \mu_i + X_{it}\beta_x + \lambda\beta(\tilde{v}_1 - \bar{v}_1)\frac{K_{it}}{K_{it}+\rho}m_i$$

Therefore, regressing p_{it} on $\frac{K_{it}}{K_{it}+\rho}m_i$, conditional on listing fixed effects and characteristics X_{it}, will identify $\tilde{\beta}_m = \lambda\beta(\tilde{v}_1 - \bar{v}_1)$. $\square$