



SCHOOL  
FOR ADVANCED  
STUDIES  
LUCCA

## 2021 RSAI World Congress

COVID-19 and Guests' Preferences in  
Short-Term Rentals: Evidence from Madrid

**Author:** Alberto Hidalgo, Massimo  
Riccaboni, Armando Rungi and F.J.  
Velázquez.

25/05/2021

- ▶ **Motivation:** The impact of COVID-19 on the travel behavior remains undetermined due to the novelty in the magnitude and intensity of this new pandemic.
- ▶ **Hot topic:** The hospitality sector has been by far one of the most hit by the COVID-19 crisis. Yet, its impact has been uneven, being sharing-home accommodations the less affected.
- ▶ **Research question:** Has COVID-19 implied a potential change in guests' preferences towards social distancing attributes provided by short-term rentals?

- ▶ **Main findings:** Guests prefer smaller, well-equipped listings to preserve social distancing in pandemic times.
- ▶ **Novel topic:** We provide, for the first time, empirical evidence of the effect of COVID-19 on the change of guests' tastes towards social distancing attributes in the hospitality sector.
- ▶ **New methodology:** This study contributes to the literature of hedonic pricing model in short-term rentals by proposing a new method to estimate hedonic price models: the Generalized Additive Model.

- ▶ **Data**
- ▶ **Methodology**
- ▶ **Results**
- ▶ **Conclusions**

► **Data:**

- **Unit of analysis:** Airbnb listings for the city of Madrid.
- **Source:** Web scraped data from Inside Airbnb.
- **Time period:** August 2019 and August 2020.
- **Sample:** Whole (all listings appeared either in Aug 2019 and Aug 2020) and restricted (listings present in the platform for Aug 2019 - Aug 2020).
- **Key Variables:** Airbnb price rate, the room type (entire apartment vs shared/private room), well-equipped kitchen (oven, dishwasher and refrigerator) and the first dimension of PCA for the number of rooms, beds and capacity of the listing.

|                           | Whole sample | Restricted sample |
|---------------------------|--------------|-------------------|
| Variable                  | N = 14,420   | N = 3,190         |
| Average price, 2019       | 76.45        | 77.92             |
| Average price, 2020       | 57.75        | 56.77             |
| Aggregate price variation | -.24         | -.27              |
| Supply, August 2019       | 11,423       | 1,595             |
| Supply, August 2020       | 2,997        | 1,595             |
| Supply variation          | 0.73         | -                 |

Table II: Sample descriptive statistics for active Airbnb listings, city of Madrid, August 2019 and August 2020.

| Variable                                    | Description   | Mean   | SD    |
|---|---|--------|-------|
| <b>Structural Airbnb characteristics</b>    |   |        |       |
| Accommodates                                | Max n° of people in each listing                                      | 3.89   | 1.95  |
| Bedrooms                                    | N° of bedrooms in each listing  | 1.53   | .84   |
| Beds  | N° of beds in each listing  | 2.24   | 1.32  |
| Social distancing size                      | First dimension of PCA for accommodates, bedrooms and beds variables. | -      | -     |
| Shared room                                 | 1 if it is a shared room, 0 otherwise                                 | .002   | .05   |
| Private room                                | 1 if it is a private room, 0 otherwise                                | .20    | .40   |
| Entire unit                                 | 1 if it is an entire apartment, 0 otherwise                           | .80    | .40   |
| Social distancing room type                 | 1 if it is a private or shared room, 0 otherwise                      | .20    | .5    |
| Oven  | 1 if the listing has an oven  | .49    | .50   |
| Dishwasher                                  | 1 if the listing has a dishwasher                                     | .37    | .48   |
| Refrigerator                                | 1 if the listing has a refrigerator                                   | .77    | .42   |
| Social distancing kitchen                   | 1 if the listing has an oven, a dishwasher and a refrigerator.        | .29    | .46   |
| <b>Quality and host attributes</b>          |   |        |       |
| Host listings count                         | N° of listings owned/managed by the host                              | 11.87  | 24.9  |
| Age   | N° of years the listing has been on the platform                      | 2.5    | 1.72  |
| Number of reviews                           | N° of reviews in each listing   | 100.03 | 97.59 |
| Review scores rating                        | Average score in each listing (0-100)                                 | 93.41  | 5.36  |
| <b>Environmental and location variables</b> |   |        |       |
| Longitude and latitude                      | Approximate longitude and latitude coordinates of each listing        | -      | -     |

Table I: Variable description for active Airbnb listings in the restricted sample, city of Madrid, August 2019 and August 2020.

## Generalized Additive Model

$$g(\mu_i) = A_i\beta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \dots$$

- ▶ **Semi-parametric specification:** Generalized Linear Model relaxing the linearity assumption in the predictors. The error term is allowed to follow any exponential family distribution.
- ▶ **Splines:** A spline is a special function defined piecewise by polynomials that are joined together at some specific locations also known as *knots*.  $f_j(x_j) = \sum_{k=1}^K \beta_{j,k} b_{j,k}(x_j)$



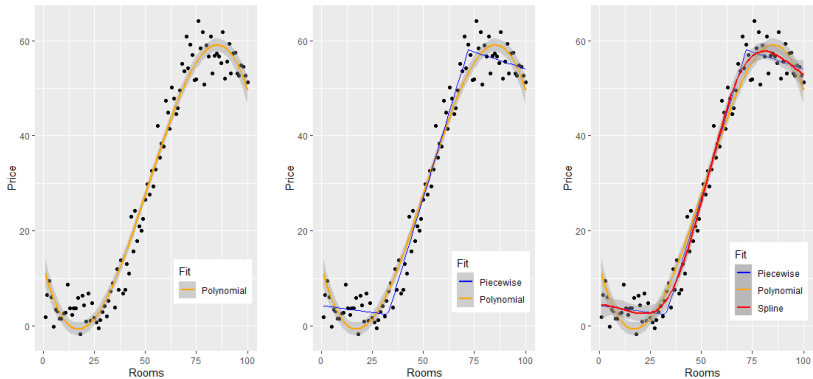


Figure I: Wiggleness fitting

## Penalized maximum likelihood estimation

$$l_p(\beta) = l(\beta) - \frac{1}{2} \sum_j \lambda_j \beta^T S_j \beta \quad \text{where} \quad \beta^T \mathbf{S} \beta = \int_{\mathbb{R}} [f'']^2$$

- ▶ **Choice of  $\lambda$ :** through Generalized Cross Validation (GCV) or Restricted Maximum Likelihood (REML).
- ▶ **Choice of  $k$ :** depends on the model setting and influences the complexity of the model, i.e., strongly non-linear variable require higher values of  $k$ .

## Specification

$$\log(P_{i,t}) = \alpha + \beta X_i + \rho \text{Aug2020}_t + \gamma(\text{Aug2020}_t \times X_i) + f(Z_{i,t}^k) + f(\text{long}_i, \text{lat}_i) + \epsilon_{i,t}$$

- ▶  $X_i$ : social distancing variables (e.g., the size, the room type and the kitchen equipment level).
- ▶  $Z_{i,t}^k$ : smooth terms for continuous variables.
- ▶  $f(\text{long}_i, \text{lat}_i)$ : bivariate smooth term of the spatial coordinates from each listing. Sort of “flexible” fixed effect (Graevenitz and Panduro, 2015).

**Linear model:** To compare the performance of GAM with more standard econometric methods, we perform a linear regression model with zip code spatial fixed effects.

| Linear terms                               | Whole sample, GAM           | Restricted sample, GAM     | Whole sample, linear       | Restricted sample linear    |
|--|-----------------------------|----------------------------|----------------------------|-----------------------------|
| (Intercept)                                | 4.295***<br>(0.005)         | 4.302***<br>(0.012)        | 4.087***<br>(0.092)        | 3.413***<br>(0.109)         |
| Social distancing size                     | 0.153***<br>(0.003)         | 0.151***<br>(0.006)        | 0.151***<br>(0.006)        | 0.150***<br>(0.008)         |
| Social distancing kitchen                  | 0.065***<br>(0.010)         | 0.051*<br>(0.021)          | 0.076***<br>(0.012)        | 0.051<br>(0.032)            |
| Social distancing room type                | -0.623***<br>(0.010)        | -0.727***<br>(0.025)       | -0.619***<br>(0.021)       | -0.720***<br>(0.045)        |
| Aug2020                                    | -0.356***<br>(0.012)        | -0.338***<br>(0.018)       | -0.345***<br>(0.030)       | -0.343***<br>(0.035)        |
| <b>Aug2020×Social distancing size</b>      | <b>-0.028***</b><br>(0.006) | <b>-0.027**</b><br>(0.009) | <b>-0.026**</b><br>(0.007) | <b>-0.026***</b><br>(0.007) |
| <b>Aug2020×Social distancing kitchen</b>   | <b>0.123***</b><br>(0.019)  | <b>0.135***</b><br>(0.028) | <b>0.116***</b><br>(0.027) | <b>0.133**</b><br>(0.043)   |
| <b>Aug2020×Social distancing room type</b> | <b>-0.034</b><br>(0.022)    | <b>0.061</b><br>(0.034)    | <b>-0.033</b><br>(0.041)   | <b>0.062</b><br>(0.068)     |
| Number of reviews                          |                             |                            | -0.001***<br>(0.000)       | -0.001***<br>(0.000)        |
| Review scores rating                       |                             |                            | 0.005***<br>(0.001)        | 0.012***<br>(0.001)         |
| Age  |                             |                            | 0.000*<br>(0.000)          | 0.023**<br>(0.007)          |
| Host listings count                        |                             |                            | -0.000<br>(0.000)          | 0.000<br>(0.000)            |
|  |                             |                            | (0.041)                    | (0.068)                     |
| <b>Smooth terms</b>                        |                             |                            |                            |                             |
| EDF: f(Longitude, Latitude)                | 49.988***<br>(55.272)       | 37.789***<br>(46.056)      |                            |                             |
| EDF: f(Number of reviews)                  | 5.814***<br>(6.811)         | 3.320***<br>(4.160)        |                            |                             |
| EDF: f(Review scores rating)               | 5.823***<br>(6.759)         | 4.800***<br>(5.694)        |                            |                             |
| EDF: f(Age)                                | 2.736***<br>(3.464)         | 4.803***<br>(5.922)        |                            |                             |
| EDF: s(Host listings count)                | 6.469***<br>(7.339)         | 3.300**<br>(3.942)         |                            |                             |
| R <sup>2</sup>                             | 0.625                       | 0.685                      | 0.604                      | 0.671                       |
| Num. obs.                                  | 14421                       | 3190                       | 14421                      | 3190                        |

\*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05. Zip code fixed effects are included but not shown for linear specifications. Errors are clustered by zip code. EDF stands for *Estimated Degrees of Freedom*. They capture the level of wiggleness in the variable. The significance of the smooth terms is a test of deviation from a flat or null function that is constant at 0 over all observed  $Z_i$ . The first dimension of the PCA for the size-related variables, Social distancing size, account for nearly 85% of the explained variance.

Table III: Whole and restricted sample hedonic models, Madrid, August 2019 and 2020.

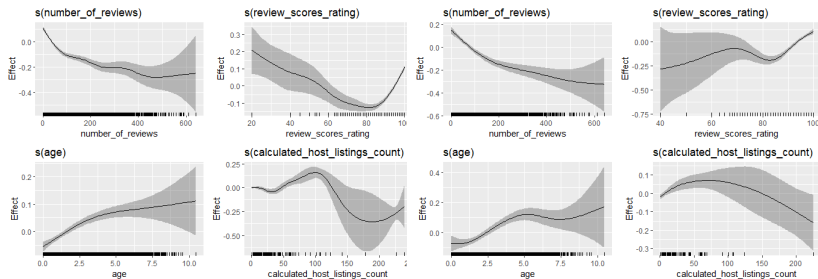


Figure II: Partial plots for the whole and restricted GAM sample respectively.

To rule out the possibility that our results are time and space dependent, we perform the following robustness checks:

- ▶ **Robustness check I: Alternative measures.** we check the robustness of our results under alternative measures of our variables of interests.
- ▶ **Robustness check II: Sep 2020 - Sep 2019.** To check whether there change in guests' preferences is a temporal phenomenon.
- ▶ **Robustness check III: Aug 2019 - Aug 2018.** To discard that our results are driven by an underlying trend in the guests' tastes.
- ▶ **Robustness check IV: Barcelona, Aug 2020 - Aug 2019.** To see whether the shift in implicit prices is a Madrid-driven result.

| Linear terms                          | Accommodates        | Beds                | Bedrooms            | Sum of kitchen amenities |
|---------------------------------------|---------------------|---------------------|---------------------|--------------------------|
| (Intercept)                           | 3.774***<br>(0.026) | 3.960***<br>(0.022) | 3.947***<br>(0.021) | 4.281***<br>(0.017)      |
| Aug2020×Accommodates                  | -0.025**<br>(0.008) |                     |                     |                          |
| Aug2020×Beds                          |                     | -0.028**<br>(0.011) |                     |                          |
| Aug2020×Bedrooms                      |                     |                     | -0.032*<br>(0.016)  |                          |
| Aug2020×Social distancing kitchen sum |                     |                     |                     | 0.031**<br>(0.012)       |
| R <sup>2</sup>                        | 0.680               | 0.661               | 0.672               | 0.681                    |
| Num. obs.                             | 3190                | 3190                | 3190                | 3190                     |

\*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . The GAM specifications are as in (4), but we omit all other variables from the table for the sake of exposition. *Social distancing kitchen sum* identifies the number of kitchen amenities of a listing, i.e., oven, dishwasher, refrigerator and oven.

Table IV: Robustness checks I Alternative measures for the social distancing variables using the restricted sample GAM specification for Madrid.

# Robustness checks

| Linear terms                            | Robustness check II<br>Sep 2020 - Sep 2019 | Robustness check III<br>Aug 2019 - Aug 2018 | Robustness check IV<br>Barcelona, Aug 2020 - Aug 2019 |
|---|--|---|---|
| (Intercept)                             | 4.326***<br>(0.013)                        | 4.256***<br>(0.012)                         | 4.611***<br>(0.017)                                   |
| Social distancing size                  | 0.154***<br>(0.007)                        | 0.152***<br>(0.006)                         | 0.138***<br>(0.009)                                   |
| Social distancing kitchen               | 0.065**<br>(0.021)                         | 0.109***<br>(0.022)                         | 0.017<br>(0.025)                                      |
| Social distancing room type             | -0.706***<br>(0.027)                       | -0.694***<br>(0.027)                        | -0.656***<br>(0.031)                                  |
| Time*                                   | -0.286***<br>(0.019)                       | 0.046**<br>(0.018)                          | -0.335***<br>(0.023)                                  |
| <b>Time×Social distancing size</b>      | <b>-0.024*</b><br>(0.009)                  | <b>0.006</b><br>(0.009)                     | <b>-0.053***</b><br>(0.012)                           |
| <b>Time×Social distancing kitchen</b>   | <b>0.098**</b><br>(0.030)                  | <b>-0.051</b><br>(0.030)                    | <b>0.080*</b><br>(0.035)                              |
| <b>Time×Social distancing room type</b> | <b>0.019</b><br>(0.036)                    | <b>0.036</b><br>(0.036)                     | <b>-0.150***</b><br>(0.041)                           |
| Smooth terms                            |  |   |   |
| EDF: f(Longitude, Latitude)             | 35.011***<br>(43.354)                      | 35.478***<br>(43.799)                       | 45.315***<br>(50.875)                                 |
| EDF: f(Number of reviews)               | 3.122***<br>(3.927)                        | 3.575***<br>(4.472)                         | 4.665*<br>(5.728)                                     |
| EDF: f(Review scores rating)            | 3.282***<br>(4.116)                        | 3.387***<br>(4.241)                         | 3.990***<br>(4.916)                                   |
| EDF: f(Age)                             | 5.734***<br>(6.913)                        | 4.585***<br>(5.666)                         | 4.134*<br>(5.157)                                     |
| EDF: f(Host listings count)             | 7.894***<br>(8.622)                        | 7.625***<br>(8.321)                         | 8.793***<br>(8.985)                                   |
| R <sup>2</sup>                          | 0.680                                      | 0.756                                       | 0.692   |
| Num. obs.                               | 2900                                       | 2010  | 3201  |

\*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05 Zip code fixed effects are included but not shown for the linear specifications. Errors clustered by zip code. EDF stands for *Estimated Degrees of Freedom*, as they capture wiggleness of predictors. Significance of smooth terms based on a test of deviation from a flat or null function constant at 0 over all observed  $Z_i$ . The first dimension of the PCA for *Social distancing size* accounts for nearly 85% of the explained variance.

\*Time is a time dummy variable that takes the value of 1 in September 2020 and 0 in September 2019 in Robustness check 1, 1 in August 2019 and 0 in August 2018 in Robustness check 2 and 1 in August 2020 and 0 in August 2019 in Robustness check 3.

Table V: Robustness check results.



- ▶ **Social distancing attributes matters:** Smaller and well-equipped kitchen listings are better positioned to host COVID-19 travellers.
- ▶ **Strong non-linearities in Airbnb price determinants:** Relaxing the linear assumption allows to correct for spatial dependencies in the urban sprawl and reveal turning points in the Airbnb characteristics.
- ▶ **Hotel vs P2P accommodation:** In the short-run, home-sharing platforms are better endowed to face the change in COVID-19 guest behavior.

**Thank you for your attention!**

| Linear terms                       | GAM Aug 2020                | GAM Aug 2019                | Linear Aug 2020             | Linear Aug 2019             |
|------------------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| (Intercept)                        | 3.952***<br>(0.009)         | 4.291***<br>(0.005)         | 3.485***<br>(0.151)         | 4.144***<br>(0.114)         |
| <b>Social distancing size</b>      | <b>0.127***</b><br>(0.005)  | <b>0.154***</b><br>(0.003)  | <b>0.125***</b><br>(0.006)  | <b>0.152***</b><br>(0.006)  |
| <b>Social distancing kitchen</b>   | <b>0.167***</b><br>(0.015)  | <b>0.064***</b><br>(0.010)  | <b>0.184***</b><br>(0.019)  | <b>0.074***</b><br>(0.010)  |
| <b>Social distancing room type</b> | <b>-0.685***</b><br>(0.019) | <b>-0.609***</b><br>(0.010) | <b>-0.690***</b><br>(0.040) | <b>-0.611***</b><br>(0.022) |
| Number of reviews                  |                             |                             | -0.001***<br>(0.000)        | -0.001***<br>(0.000)        |
| Review scores rating               |                             |                             | 0.007***<br>(0.002)         | 0.005***<br>(0.001)         |
| Age                                |                             |                             | 0.041***<br>(0.008)         | -0.002<br>(0.004)           |
| Host Listings Count                |                             |                             | 0.000<br>(0.001)            | -0.000**<br>(0.000)         |
| <b>Smooth terms</b>                |                             |                             |                             |                             |
| EDF: f(Longitude, Latitude)        | 29.702***<br>(37.867)       | 49.455***<br>(55.049)       |                             |                             |
| EDF: f(Number of reviews)          | 4.092***<br>(5.047)         | 5.539***<br>(6.552)         |                             |                             |
| EDF: f(Review scores rating)       | 4.704***<br>(5.675)         | 5.403***<br>(6.374)         |                             |                             |
| EDF: f(Age)                        | 3.886***<br>(4.862)         | 1.500***<br>(1.855)         |                             |                             |
| EDF: f(Host Listings Count)        | 4.285**<br>(5.096)          | 8.807***<br>(8.987)         |                             |                             |
| R <sup>2</sup>                     | 0.637                       | 0.624                       | 0.6004                      | 0.599                       |
| Num. obs.                          | 2998                        | 11423                       | 2998                        | 11423                       |

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . Zip code fixed effects are included but not shown for the linear specifications. Errors were clustered at the zip code level to account for residual spatial correlation. EDF stands for *Estimated Degrees of Freedom*. They capture the level of wiggleness in the variable. The significance of the smooth terms is a test of deviation from a flat or null function that is constant at 0 over all observed  $Z_i$ .

Table A1: Whole cross-sectional sample hedonic models, Madrid, August 2020 and 2019.

| Linear terms                       | GAM Aug 2020                | GAM Aug 2019                | Linear Aug 2020             | Linear Aug 2019             |
|------------------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| (Intercept)                        | 3.974***<br>(0.012)         | 4.286***<br>(0.012)         | 2.424***<br>(0.323)         | 3.887***<br>(0.227)         |
| <b>Social distancing size</b>      | <b>0.125***</b><br>(0.006)  | <b>0.152***</b><br>(0.006)  | <b>0.126***</b><br>(0.008)  | <b>0.149***</b><br>(0.008)  |
| <b>Social distancing kitchen</b>   | <b>0.162***</b><br>(0.020)  | <b>0.090***</b><br>(0.021)  | <b>0.155***</b><br>(0.024)  | <b>0.080***</b><br>(0.029)  |
| <b>Social distancing room type</b> | <b>-0.706***</b><br>(0.025) | <b>-0.672***</b><br>(0.025) | <b>-0.686***</b><br>(0.055) | <b>-0.692***</b><br>(0.041) |
| Number of reviews                  |                             |                             | -0.001***<br>(0.000)        | -0.001***<br>(0.000)        |
| Review scores rating               |                             |                             | 0.020***<br>(0.004)         | 0.007**<br>(0.002)          |
| Age                                |                             |                             | 0.043***<br>(0.010)         | -0.003<br>(0.009)           |
| Host Listings Count                |                             |                             | -0.000<br>(0.001)           | 0.001<br>(0.001)            |
| Smooth terms                       |                             |                             |                             |                             |
| EDF: f(Longitude, Latitude)        | 21.623***<br>(28.234)       | 32.434***<br>(40.715)       |                             |                             |
| EDF: f(Number of reviews)          | 3.318***<br>(4.151)         | 2.641***<br>(3.321)         |                             |                             |
| EDF: f(Review scores rating)       | 3.732***<br>(4.649)         | 3.980***<br>(4.824)         |                             |                             |
| EDF: f(Age)                        | 3.004***<br>(3.785)         | 1.001<br>(1.001)            |                             |                             |
| EDF: f(Host Listings Count)        | 4.798**<br>(5.701)          | 8.340***<br>(8.844)         |                             |                             |
| R <sup>2</sup>                     | 0.678                       | 0.702                       | 0.6574                      | 0.657                       |
| Num. obs.                          | 1595                        | 1595                        | 1595                        | 1595                        |

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . Zip code fixed effects are included but not shown for the linear specifications. Errors were clustered at the zip code level to account for residual spatial correlation. EDF stands for *Estimated Degrees of Freedom*. They capture the level of wiggleness in the variable. The significance of the smooth terms is a test of deviation from a flat or null function that is constant at 0 over all observed  $Z_i$ .

Table All: Restricted cross-sectional sample hedonic models, Madrid, August 2020 and 2019.