Online Appendix for:

Better residential than ethnic discrimination!

Reconciling audit and interview findings in the Parisian housing market

François Bonnet,[*] Etienne Lalé,[†] Mirna Safi,[‡] Etienne Wasmer[§]

A Audit study design

This appendix provides an extensive description of our audit study.

A.1 Background and prior methods

Our audit study was largely inspired by prior audits that have been conducted elsewhere than in France. The example of the United States is by far the most influential. Since the late 1970s, the U.S. Department of Housing and Urban Development (HUD) has been monitoring series of nationwide paired-testing audit studies in both rental and sales markets. The objectives, methods and results of these studies are carefully documented in reports published by the HUD's office of development and research. They define the state of the art in the implementation of audit methods and provide thorough discussions of the statistical procedures used to analyze audit data.[1] This section provides a brief description of these canonical audit studies to help the reader understand better how our own audit study was designed and implemented.

Nationwide audit studies monitored by the HUD have been conducted once every decade since the first study launched in 1977. Over time, the scope and methods have evolved, but there is unity in several features that can be summarized as follows:

1. Experiment: In its very essence, a housing audit study is an experiment where two testers (one from the majority group and the other from a minority group) are matched, trained to appear com-

[*]CNRS, UMR Pacte.
[†]Department of Economics, University of Bristol.
[‡]Department of Sociology, Sciences Po, OSC, CNRS.
[§]Sciences Po, Department of Economics and LIEPP.

[1]Of course, these paired-testing audit studies are not the only housing audit studies in the United States. Many smaller scale housing audits also exist and have been conducted since the first national audit study of 1977; see for instance Yinger (1986) for an analysis of an early local audit study conducted in Boston in 1981.

parable in several observable dimensions and then put into contact with real estate and rental agents (Turner, 1992).

2. Measurement: The outcome (or treatment) received by each tester can be encoded and subsequently used in statistical analysis. Systematic comparison of these outcomes allows detecting discrimination against one tester of the pair. Ultimately, differential treatments can be summarized by statistics. It should be underlined that what is being measured is often subject to debates (Heckman, 1998; Yinger, 1998), and also that interpretability is not the unique purpose of these statistics. Indeed, the HUD aims at producing figures that can be compared across regions and over time; interpretation set aside, those allow measurement of geographic differentials and/or time trends.

3. Selection of testing areas: Cost and logistic considerations suggest that metropolitan areas are best designated for conducting audit studies. Since the HUD aims at producing representative estimates, it uses Census data to draw samples of metropolitan testing sites and to achieve high coverage of areas with minorities. The 2012 national audit study for instance covers 28 metropolitan sites (Turner et al., 2013).

4. Selection of advertisements: Advertisements are drawn in accordance with routine search behaviors of housing applicants. Historically, audit studies have relied on advertisements found in major newspapers (Turner, 1992; Yinger, 1998). The 2012 audit study adapted its protocol to reflect new practices brought about by technology changes over the past decades; it sought to take into account housing searches via cell phone, the Internet, etc. (Turner et al., 2013).

5. Selection of the minority group: The minority group is chosen in order to test for potential discriminatory actions against that specific group. For instance the 1977 audit study tested for discrimination against Black housing applicants only. The 1989 audit study extended the experiment to Hispanic housing applicants and the 2000 audit study to Hispanic, Asian and Native American housing applicants (Turner, 1992; Turner et al., 2013).

The nationwide audit studies just described are tailored to address specific concerns in relation to race or ethnicity in the U.S., and to assess enforcement of the federal law (the 1968 Fair Housing Act). Our paired-testing audit study contrasts in scope and objective and we can summarize the major differences as follows:

- We focus on a single metropolitan area. Our goal is not produce nationally-representative figures on housing discrimination. Instead, we test whether minority home-seekers systematically receive less favorable treatments within a geographic area where market tightness induces landlords to be very selective about applicants.

- We focus on the first possible contact between applicants and rental agents, i.e. phone conversations, and we do not organize face-to-face meetings. This choice was not dictated by budget limitations only; as explained below, phone conversations are well designated to capture potential discriminatory actions that we seek to study.

- We focus on the private rental housing market. Buying a home is a relatively more complex process and it is clear that phone conversations would be less effective in detecting potential discriminatory actions involved in this process. Moreover, the scenario of our audit study would have been less plausible in the context of home buying, particularly for apartments located within the city of Paris.

Our paired-testing audit nevertheless retains some of the unifying characteristics listed above. Broadly speaking, our study reproduces the experiment described in (1) and (2): we matched an individual from a majority (white) group with an individual from a minority group, put them into contact with real estate agents and compared the treatment each of them received. How the unifying traits (3) and (4) reflected in our audit study is explained in Sections A.2 and A.3 below: those provide detailed information about the targeted metropolitan area and ads that we sampled. The remaining unifying trait (5) is discussed in Sections A.4 and A.5 which describe the selection of testers and their fictitious identities.

A.2 Targeted metropolitan area

Our audit study was done in Paris and in several cities of the region Île-de-France. We focused on the private rental housing market, which accounts for 35% of the housing stock in Paris, and for 21% of the housing stock in the whole region Île-de-France (figures for the year 2006 excerpted from Jankel and Salembier, 2008).[2]

Within the city of Paris, we sampled Internet postings on the popular website `http://www.seloger.com` for housing units located in almost all *arrondissements* (administrative district) of Paris and obtained observations for 14 out of the 20 *arrondissements*. Outside the city of Paris, we restricted the study to cities with direct connections by public transportation (metro or RER) to Paris. The data resulting from our audit study thus include housing units located inside Paris, in some cities from the "inner ring" (*départements* 92, 93 and 94) and in some cities from the "outer ring" (*départements* 77, 78, 91 and 95) – see Figure A1.

Table A1 displays a set of statistics characterizing the housing vacancies of our audit study. We also report descriptive statistics about the size and rent of privately rented apartments in the Paris region. We targeted apartments for couples without kids. As shown in the table, the surface areas in our dataset fall well in the range of one- to two-room apartments as measured by representative surveys. Rents per square meter are also remarkably in line with the data; in particular, housing vacancies in our dataset reproduce the observed disparities across *départements*. Overall, the table suggests that housing vacancies that form the basis of our empirical work adequately match the housing stock of the Paris region.

Our focus on small-sized apartments advertised on the Internet was motivated by some further considerations. First, these are the apartments typically looked for by middle-class households such as our

[2]Within the European Union, France has a slightly lower home ownership rate compared to the European average : 58% against 65 in the EU in 2007 ; the rate increased in France since then and is about 62% in 2013.

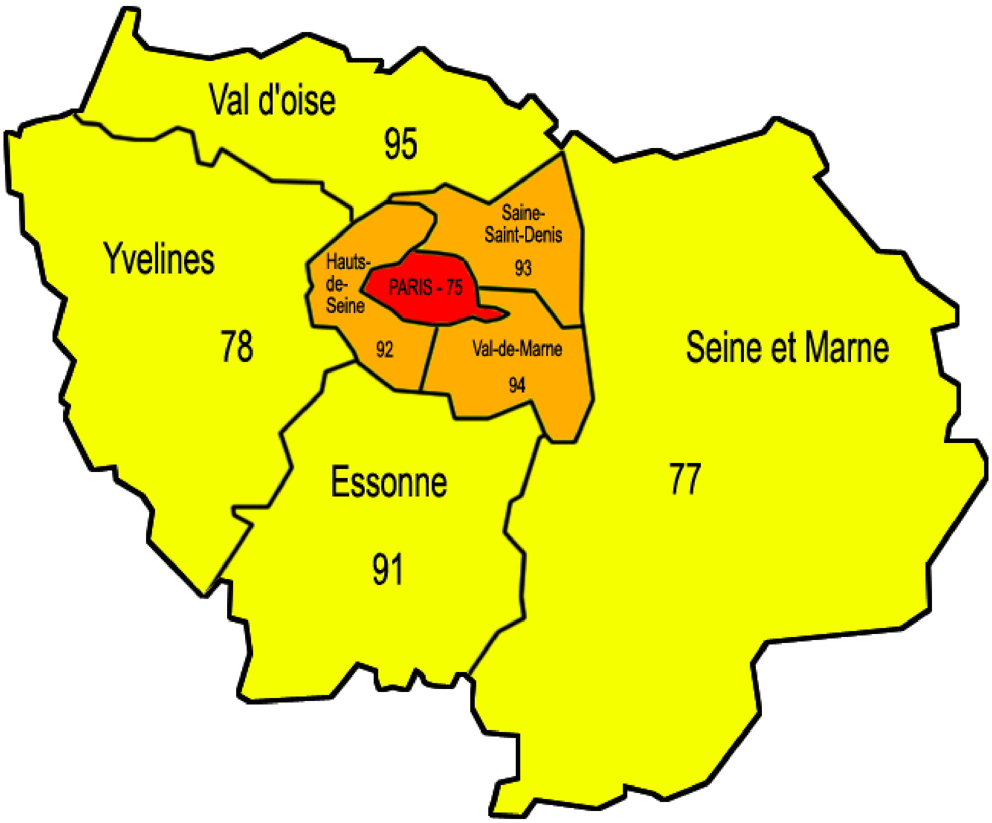

Figure A1. Map of région Île-de-France

fictitious applicants. Second, the information provided on Internet postings is typically comprehensive. In the phone conversations, testers could thus limit their questions to availability and viewing. This has two advantages: to reduce the array of possible conversation outcomes and to minimize risks of suspicious real estate agents. The conclusion from our protocol is that this experimental design was sufficient to capture possible discriminatory actions at this early stage of the housing search process.

Vacancy rates for apartments are much higher in the region Île-de-France relative to the rest of France. No less than 70% of vacant housing units are located within the Paris metropolitan area (Decondé, 2012). 6.2% of all housing units located in the region Île-de-France are vacant, and the corresponding figures for the city of Paris alone rises to 9.2%. This tightness of the housing market is not caused by a lack of housing applicants (Bonnet et al., 2011). Instead, it reflects the fact that landlords are very cautious in their selection of tenants, probably because the French legal environment makes evictions a long and difficult process. Higher vacancy rates in the region Île-de-France (and the possibility that it results in discriminatory actions) thus make it a relevant case for our audit study .

A.3 Sampling of ads

We collected advertisements for housing units from the website `http://www.seloger.com`. This is a major platform for real estate agents who choose to advertise their vacancies on the Internet. Users can

Table A1. Characteristics of the housing units, in our audit study and overall

	Sample of our audit study				Representative sample		
	N	Rent (€)	Surface (Sq. meter)	Rent per Sq. meter (€)	Surface one-room (Sq. meter)	Surface two-room (Sq. meter)	Rent per Sq. meter (€)
Full sample	250	749	41	19.30			22.46
Median		748	39	18.60			
25% centile		685	33	16.18			
75% centile		815	46	21.52			
Median in Paris	45	750	32	25.00	25	42	28.06
Median in Dept. 77	37	660	40	16.67	27	44	14.00
Median in Dept. 78	37	815	45	18.17	29	48	17.44
Median in Dept. 91	40	706	42	17.32	30	42	15.04
Median in Dept. 92	38	716	33	21.18	27	44	22.92
Median in Dept. 93	2	720	35	20.61	28	42	17.92
Median in Dept. 94	11	750	40	20.29	28	44	19.27
Median in Dept. 95	40	800	50	15.54	28	46	16.09

NOTE: Surfaces are measured in square meters (Sq. meter); Rents are reported in euros (€). Representative figures for the surface area of one-room and two-room apartments are for the year 2011; they are extracted from an extensive report of the *Observatoire des loyers* (Observatory of rents) written by Coz and Prandi (2012). Representative figures for Rents per square meters are from the information-aggregator website `http://www.lacoteimmo.com` which gathers information on rents for apartments advertised online.

sort advertised units along a number of criteria, such as the rent, size, location, etc. Our focus on relatively small-sized apartments, located in specific geographical area and for a selected rent range implies that we could almost exhaust the corresponding units advertised on the website. As shown in Table A1, the latter provided a close-to-representative cross-section of the targeted housing units located in the Paris metropolitan area. Further, it covers a large share of online connections to real estate agencies' advertisements. According to the French audience measurement company Mediamétrie, Se loger was the first online agency in France in 2009-2010 (time of the survey) and was still second in 2012. It had more than 3 millions of monthly single IP connections, amounting to 7% of all internet users in France.[3]

A.4 The testers

Our audit study was carried out by a team of 8 individuals, aged 22 to 28, 2 women and 6 men (hereafter: the "testers"). All testers were highly skilled and were fully informed about the situation, of what was to be tested, and knew that any deviation from the methodology would invalidate the results. Two half-day training sessions were also organized to give testers a better understanding of the objective. They were all explicitly told the following: "We don't know whether there is discrimination along those dimensions.

[3] `http://www.journaldunet.com/ebusiness/le-net/immobilier-mediametrie/seloger.shtml`

Remain neutral, both during the phone call and in reporting the results." They were asked to report a maximum of qualitative information. In particular, there was a variable to indicate "a suspicion of audit", included to report questions by the agent such as "Did you call yesterday?". This occurred overall in 7 cases out of 340 calls during the first phase. All testers were explicitly involved in the design of part of the procedures, notably to determine the final encoding of variables.

Our testers did not have any particular accent, and we instructed them not to fake any during the phone conversations. Our main concern was that accents typically associated to French minority suburbs convey information about both residential and ethnic origins, which we aimed to disentangle in this study. Thus, we suspected that accents would, in general, reinforce the impact of any of our two discriminatory criteria while making the interpretation of the coefficients more difficult. Setting aside accents was also dictated by the fact that the relevant information about residential or ethnic origin was to be revealed at the very beginning of phone conversations, which turned out to be very brief. Accents would have added a complex ingredient to these short talks, and we found this undesirable: any effect of ethnicity would have been jointly attributed to ethnicity itself and to the lack of cultural integration this accent would have revealed.

A.5 Scenarii

Testers were assigned fictitious identities designed to impersonate the average housing applicant in the Paris region. Fictitious identities would include a name, a place of residence, an occupation and an income level, for instance: "Sébastien Fournier (French name) / Kader Boualem (North African name), lives in La Courneuve (deprived suburb) / Versailles (rich suburb), is 31 years-old, works as an accountant and earns a monthly wage of 1,700€". These identities also included a marital status (married with no kid), an occupation and an income for the spouse: typically, "works as a secretary and earns 1,300€ per month". Finally, testers were attributed a geographical area around Paris that would be consistent with their asserted willingness to relocate.

Our choice of fictitious identities was motivated by the following observations:

- According to Charrier et al. (2008), the monthly income of household applying for housing on the Paris private rental housing market was 2,909€ in the census year 2006 (2,638€ for the region Île-de-France as a whole), and average household income was 3,103€ in Paris (3,131€ at the level of the region).

- The same study reports that professionals (*Cadres et professions intellectuelles supérieures*) account for one-third of the population working in Paris, and is even higher among those applying for housing on the rental market (as high as 52%). This is followed by intermediary professions (*Professions intermediaires*, i.e. clerical workers with university education) which account for 14% of the population in the Paris metropolitan area. The occupations for the fictitious applicant and his/her partner were selected to reflect these figures.

With regards to asserted ethnic and residential origins, those were selected as follows:

- Names were chosen to signal origin from either France or North Africa. We followed the methodology of the International Labour Office described by Cediey and Foroni (2008). Specifically, we used the following "North-African" fictitious names: Rachid Ammelal, Mehdi Belbouab, Hafsa Belhadj, Farid Boukhrit, Kader Boualem, Habib El Bekkali and Fathia Laouadi; for "white French" ones: Marion Denis, Pascal Dubois, Sébastien Fournier, Sébastien Pialoux, Julie Morvan, Julien Roche and Gilles Rousseau.

- The fictitious current locations were chosen among cities of the Île-de-France region that are either notoriously deprived, or privileged. Some of the selected deprived cities had experienced the 2005 urban riots. Cities used to signal deprived neighbourhoods were: Bagneux, Bondy, Gennevilliers, Saint Ouen, Sarcelles, Trappes and Villeneuve Saint-Georges; cities used to signal privileged neighbourhoods were Le Raincy, Levallois, Meudon, Nogent sur Marne, Sceaux, Versailles and Vincennes.
  We collected and reported in Table A2 some summary statistics about each of the city of respondents in our audit study. Further, what matters in the selection process is the typical representations of these neighbourhoods (good vs. bad) as, for instance, conveyed by the media after the 2005 urban riots. The true statistics are (sadly) less relevant in decision making than perceptions. As an anecdote, a recent IPSOS poll indicate that in France, respondents overestimate by almost a factor 4 the share in population of Muslims in France (estimated to be 31% against an actual population of 8% according to `http://www.economist.com/blogs/graphicdetail/2015/01/daily-chart-2`). A professional real estate agent from the Paris region could not ignore the social connotations of the chosen cities.

Although our team included two women, we made no attempt to test for discrimination by gender. Indeed, given the limited number of calls that could be made, it appeared desirable to limit the number of potentially discriminatory combinations and to focus on ethnic and residential stigma. The two female testers of our team would thus always work in tandem. Overall, we noticed no significant differences between the outcomes reported by our pair of female testers relative to male testers.

A.6 Location and Technical Details

Testers were installed in an office provided by our university (Sciences Po). They called from two separate phones. The first one was installed on a fixed line, opened for this testing procedure as for a private household and the number of which started in 01 (Paris and suburbs region). Importantly, it differed from the Sciences Po numbers that all start with the six numbers 01 45 49. The second one was installed on an Internet box provided by the telephone company, the number starting in 09 which is not specific to any region. Both phones had an answering machine. In addition, the first tester was given a cell phone. When rental brokers asked for a phone number to call back, the first tester would give the cell phone

Table A2. Characteristics of cities selected in the audit

	Share of owners (%)	Median income (*)	Unemp. Rate 15-64 (%)	Share of immigrants (%) (**)
a. Cities used to signal deprived neighbourhoods				
Bagneux	25.7	17,508	14.2	21.2
Bondy	43.8	14,088	18.8	28.8
Gennevilliers	19.6	13,614	18.5	30.6
Saint Ouen	20.5	14,505	18.7	32.4
Sarcelles	33.7	12,189	23.2	31.3
Trappes	26.1	13,563	17.4	26.1
Villeneuve Saint-Georges	37.7	13,862	16.7	28.5
b. Cities used to signal privileged neighbourhoods				
Le Raincy	61.9	27,430	9.7	11.0
Levallois-Perret	37.1	30,761	9.5	14.7
Meudon	54.2	28,359	8.5	13.5
Nogent sur Marne	52.7	30,680	9.8	13.0
Sceaux	46.4	35,558	8.8	10.6
Versailles	45.8	30,577	7.9	8.8
Vincennes	48.1	31,408	8.7	13.2

NOTE: All variables are for the year 2011, except the share of immigrants (2012). (*) In euro (€) per consumption unit. (**) French definition of immigrants: persons born abroad and not French at birth (Haut Conseil de l'Intégration).

number and the second tester would give the number starting in 09 (which does not necessary indicates a fixed line). Testers would never call at the exact same time; they would instead work sequentially. Phone calls were given during regular office hours on weekdays. Providing testers with an answering machine appeared necessary in order to refine the measurement of the outcome of the conversations.

A.7 Timing

A typical session consisted in a sequence of phone calls (i) to vacancies that had been contacted before and (ii) to new vacancies. The first tester reported the day and hour of the phone call as well as an identifier of the vacancy so that the second tester could call back the same vacancy after a "long enough" period of time had elapsed. As the duration between the two calls increased, the outcome of the second call was more likely to be negative for reasons unrelated to discrimination: for instance, another applicant might have called after the first tester. It was thus necessary to limit the time between calls as much as possible. However, the audit was more likely to be suspected if the interval was too short, so testers were encouraged not to wait too long to make the second call. We found a period of half a day to one day to be a good compromise between these constraints, and testers were able to follow this procedure in most cases.

We insisted on switching the order in which phone calls were made, so that the tester responsible with the role of minority applicant would call first as often as the other tester. Every four calls, testers thus had to invert the order. In the resulting dataset, the order is perfectly balanced across pairs of calls and there is hence no systematic association with the fictitious identities of the testers.

A.8 Encoding

As indicated in the body of the paper, the outcomes of conversations were coded as follows:

Code	Outcome
1	Apartment is already rented, nothing else available
2	Caller is asked to send a written application with personal details
3	Real estate agent will call back, but no return call
4	Apartment is already rented but something else is available
5	Real estate agent plans a group visit
6	Real estate agent plans an individual visit

The outcomes are rank-ordered. We decided on this method of coding after several days of trials and about 50 calls. Code 4 was considered inferior to 5 and 6 because it may indicate steering. Code 3 could have been considered inferior to Code 2, but the testers themselves reported that Code 2 was frequently a polite way to send applicants off. Code 3 on the other hand might be due to a random event, especially during rush hours and because of the high turnover rates for Paris apartments. Finally, Code 3 was changed appropriately into 4, 5 or 6 when real estate agents called back as promised or left a message on the answering machine of each tester.

When testers were redirected towards a different apartment (Code 4), we were unable to test whether this could be due to steering because testers were usually not given sufficiently many details about the alternative offer. We note that this was located in the same city in most cases and there are hence reasons to assume that the other unit was a proper substitute. However, because we cannot test this assumption directly, we offer robustness checks in Appendix B with respect to Code 4.

A.9 Additional information

We conducted about 50 experimental phone calls before opting for our final setup. Two constraints appeared worthy of consideration. First, although there were, in principle, four scenarios that we would have liked to test (good name/good location, bad name/good location, etc.), we could not ask four testers to contact the same advertised dwelling. To avoid detection of the audit, it was necessary to impose a delay of at least 24 hours between phone calls for the same vacancy. With four scenarios, this would have required almost a week to contact the same vacancy, which, given the high rate of turnover of the housing vacancies contacted, was not feasible. Second, the typical phone conversation was extremely short (often less than one minute), which limited the number of “signals” (geographic origin or ethnic origin) that could be revealed within the phone conversation.

Throughout the duration of the experiment, we had frequent debriefings with our team. We asked tester to report subjective information about each phone calls in order to monitor them more efficiently. Eventually, we were able to check whether this information (which we encoded in our final dataset) affected the variability of the outcome across testers and/or across phone conversations and found no systematic relationship.

A.10 External validity and replication

Due to budget limitations, we were not able to extend our audit study to other major metropolitan areas of France, such as Lyon or Marseille. The Paris region is much bigger, more diverse, and overall more economically developed than provincial urban agglomerations. It is hence a natural candidate to conduct experiments aimed at assessing the extent and nature of discrimination in France. Further research in the rest of France would also be informative especially to uncover regional patterns, and would provide valuable material for comparisons.

We noted above that our experimental design appeared sufficient to capture possible discriminatory actions at the early stage of the housing search process. We hope to have provided enough details in this appendix to allow for replication of the study. Our design could also be adapted to specific settings in other countries. Finally, the detailed depiction of our experiment provided in this appendix should allow direct comparison to the methods implemented in other audits.

B Additional tables and results

Cross-tabulation of the outcomes

As shown in Figure 1 in the paper, outcome 6 (individual visit planned) and 3 (we will call you back but no return call) were the most frequent outcomes of the phone conversations. In Tables B1 and B2, we cross-tabulate the outcomes. This is instructive because if we were to analyze only the phone conversations which resulted in these two outcomes only, the detection of discriminatory practices would come from rows and columns 3 and 6 in the Tables. Fortunately for the purpose of identification, we notice that when one candidate was told that he/she would be called back, the other candidate was told the same thing in only half of all cases. Furthermore, the table indicates that this tended to be more favorable to the majority candidate in the first procedure. Indeed, when the majority candidate was to be called back, the minority candidate obtained an individual viewing in 23% of all cases (row 3, column 6 in Table B1), whereas when the minority candidate was told to expect a phone call back, the majority candidate managed to arrange an individual visit in 39% of all cases (row 6, column 3 in Table B1). In the second procedure on the other hand, the odds were more balanced: these percentages become 27% (row 3, column 6 in Table B2) and 17% (row 6, column 3 in Table B2), respectively. This reiterates the finding that discriminatory practices affected the outcomes under the first procedure, but not in the second procedure.

Robustness check 1: bootstrapping pairs in procedure #1

To assess the sensitivity of our results to the difference in sample sizes across procedures, we replicate the finding of procedure #1 by boostraping pairs of phone calls so as to obtain comparable sample sizes. That is, we randomly select pairs of calls among the 177 initial pairs and re-estimate the generic linear models specified in the paper. We focus on our preferred specifications, i.e. regressions with audit-level fixed effects (columns (IV)–(VII) in Table 3 of the paper). To accord with the sample sizes of the second procedure, we select 77 pairs. The results from the bootstrap procedure based on 500 replications are reported in Table B3. The significance level is reduced but the main coefficient of interest remains significant at the 10 percent level in regressions that control better for differences across individual phone calls (columns (III) and (IV)).

Robustness check 2: sensitivity with respect to outcome no. 4

As we note in Subsection A.8 of the Appendix, we were unable to test whether outcome 4 (Apartment is already rented but something else is available) could reflect steering instead of redirection towards a proper substitute for the initial apartment. Nevertheless, we can offer two series of robustness checks with respect to outcome no. 4:

- First, we can group Code 4 with Codes 1, 2 and 3 ("rather negative outcome") and re-estimate our models. If our assumption that Code 4 indicated a neutral or either positive outcome, then we

Table B1. Cross-tabulation of the outcomes across paired phone calls: Procedure #1 (residence first)

	Minority candidate						
	Already rented, nothing else (1)	Asked to send a file with personal details (2)	We call you back but no recall (3)	Already rented but smtg else available (4)	Group visit planned (5)	Individual visit planned (6)	Total
Majority candidate							
Already rented,	10	0	5	2	2	2	21
nothing else	47.62	0.00	23.81	9.52	9.52	9.52	100.00
(1)	40.00	0.00	8.06	10.53	14.29	3.92	12.14
Asked to send a file	0	1	0	0	0	0	1
with personal details	0.00	100.00	0.00	0.00	0.00	0.00	100.00
(2)	0.00	50.00	0.00	0.00	0.00	0.00	0.58
We call you back	5	0	25	6	1	11	48
but no recall	10.42	0.00	52.08	12.50	2.08	22.92	100.00
(3)	20.00	0.00	40.32	31.58	7.14	21.57	27.75
Already rented but	3	0	6	4	0	5	18
smtg else available	16.67	0.00	33.33	22.22	0.00	27.78	100.00
(4)	12.00	0.00	9.68	21.05	0.00	9.80	10.40
Group visit	2	0	2	0	8	2	14
planned	14.29	0.00	14.29	0.00	57.14	14.29	100.00
(5)	8.00	0.00	3.23	0.00	57.14	3.92	8.09
Individual visit	5	1	24	7	3	31	71
planned	7.04	1.41	33.80	9.86	4.23	43.66	100.00
(6)	20.00	50.00	38.71	36.84	21.43	60.78	41.04
Total	25	2	62	19	14	51	173
	14.45	1.16	35.84	10.98	8.09	29.48	100.00
	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table B2. Cross-tabulation of the outcomes across paired phone calls: Procedure #2 (ethnicity first)

	Minority candidate						
	Already rented, nothing else (1)	Asked to send a file with personal details (2)	We call you back but no recall (3)	Already rented but smtg else available (4)	Group visit planned (5)	Individual visit planned (6)	Total
Majority candidate							
Already rented,	2	0	2	1	0	1	6
nothing else	33.33	0.00	33.33	16.67	0.00	16.67	100.00
(1)	20.00	0.00	8.33	20.00	0.00	3.03	7.79
Asked to send a file	2	1	2	0	0	0	5
with personal details	40.00	20.00	40.00	0.00	0.00	0.00	100.00
(2)	20.00	33.33	8.33	0.00	0.00	0.00	6.49
We call you back	4	1	16	1	0	8	30
but no recall	13.33	3.33	53.33	3.33	0.00	26.67	100.00
(3)	40.00	33.33	66.67	20.00	0.00	24.24	38.96
Already rented but	1	0	0	2	0	0	3
smtg else available	33.33	0.00	0.00	66.67	0.00	0.00	100.00
(4)	10.00	0.00	0.00	40.00	0.00	0.00	3.90
Group visit	0	0	0	0	1	2	3
planned	0.00	0.00	0.00	0.00	33.33	66.67	100.00
(5)	0.00	0.00	0.00	0.00	50.00	6.06	3.90
Individual visit	1	1	4	1	1	22	30
planned	3.33	3.33	13.33	3.33	3.33	73.33	100.00
(6)	10.00	33.33	16.67	20.00	50.00	66.67	38.96
Total	10	3	24	5	2	33	77
	12.99	3.90	31.17	6.49	2.60	42.86	100.00
	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table B3. Robustness check 1: discrimination rates with bootstraped pairs in procedure #1

Procedure #1 (residence first)	(I)	(II)	(III)	(IV)
Deprived suburb	-0.1098	-0.1227	-0.1301	-0.1373
	(0.0703)	(0.0751)	(0.0786)	(0.0793)
Minority name (North Africa)			0.1273	0.1046
			(0.1358)	(0.1411)
Name revealed			0.2415	0.2373
			(0.1465)	(0.1494)
Name revealed x Minority Name			-0.1888	-0.1747
			(0.1706)	(0.1728)
Controls:				
Individual call		Y		Y
Audit fixed effects	Y	Y	Y	Y
N	154	154	154	154
R^2	0.0311	0.0443	0.0668	0.0775

NOTE: Standard errors in parentheses. Standard errors are obtained under a bootstrap procedure where 77 pairs are chosen randomly among the 177 initial pairs. Each column reports the coefficients γ and β estimated along with a specific set of controls. Name revealed is a dummy indicating whether the tester revealed his/her name in the phone conversation. Characteristics at the level of the individual phone calls are: a dummy indicating whether the tester used the phone line with a number starting with 09 (indicating Internet box) and a dummy indicating whether he/she called first.

expect the coefficient on the dummy for the minority candidate to be lower in absolute value. We also expect the fit of the regression to be lower than under the benchmark specification.

- Second, we can run our regressions on the subsample of audit pairs where neither the minority candidate nor the majority candidate was offered to visit a different housing unit. As this effectively reduces the size of the sample, we expect the coefficient to be less precisely, but without any change in sign.

For economy of space, here again we focus on our preferred specifications, i.e. with audit-level fixed effects. Table B4 reports the estimation results under the two audit procedures: columns (I)–(IV) show the results obtained after grouping Code 4 with Codes 1 to 3 and columns (V)–(VIII) show the results obtained after excluding outcome 4.

It is clear that our results are not driven by outcome no 4. Our estimates in the first procedure remain almost unchanged and statistically significant, even with a smaller sample size when audit pairs with Code 4 are excluded. Conversely, no significant change occurs in the second procedure. Overall, this robustness check confirms the analysis of Tables B1 and B2 showing that detection of discrimination (and the lack thereof) relies mostly on outcomes 3 and 6, but not on outcome 4.

Table B4. Robustness check 2: sensitivity with respect to outcome no. 4

Procedure #1 (residence first)	Grouping Code 4 with Codes 1–3 (I)	(II)	(III)	(IV)	Excluding Pairs with Code 4 (V)	(VI)	(VII)	(VIII)
Deprived suburb	**-0.1156**	**-0.1227**	**-0.1268**	**-0.1284**	**-0.1286**	**-0.1324**	**-0.1425**	**-0.1453**
	(0.0449)	(0.0449)	(0.0503)	(0.0500)	(0.0496)	(0.0499)	(0.0545)	(0.0543)
Minority name			0.1046	0.0873			0.1679	0.1639
(North Africa)			(0.0837)	(0.0875)			(0.0926)	(0.0936)
Name revealed			0.1213	0.1246			0.1827	0.1753
			(0.0957)	(0.0964)			(0.1048)	(0.1063)
Name revealed			-0.1632	-0.1584			**-0.2577**	**-0.2483**
x Minority Name			(0.1099)	(0.1096)			(0.1200)	(0.1177)
Controls:								
Individual call		Y		Y		Y		Y
Audit fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
N	346	346	345	345	280	280	279	279
R^2	0.0373	0.0413	0.0543	0.0590	0.0463	0.0532	0.0823	0.0846
R^2 with fixed effect	0.8010	0.8018	0.8077	0.8086	0.8272	0.8285	0.8369	0.8373

Procedure #2 (ethnicity first)	Grouping Code 4 with Codes 1–3 (I)	(II)	(III)	(IV)	Excluding Pairs with Code 4 (V)	(VI)	(VII)	(VIII)
Minority name	0.0260	0.0323	0.0303	0.0367	0.0423	0.0526	0.0391	0.0478
	(0.0524)	(0.0585)	(0.0572)	(0.0674)	(0.0549)	(0.0599)	(0.0597)	(0.0694)
Deprived suburb			0.0375	0.0376			0.0285	0.0269
			(0.0567)	(0.0583)			(0.0606)	(0.0617)
Residence revealed			-0.0295	-0.0131			0.0890	0.0969
			(0.2462)	(0.2747)			(0.2386)	(0.2725)
Residence revealed			-0.0803	-0.0972			-0.1566	-0.1659
x Deprived suburb			(0.2379)	(0.2727)			(0.2422)	(0.2792)
Controls:								
Individual call		Y		Y		Y		Y
Audit fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
N	154	154	154	154	142	142	142	142
R^2	0.0032	0.0081	0.0130	0.0178	0.0085	0.0115	0.0183	0.0222
R^2 with fixed effect	0.8827	0.8833	0.8839	0.8844	0.8890	0.8893	0.8901	0.8905

NOTE: Estimates that are statistically significant at the 5% level are in boldface. Standard errors in parentheses. Standard errors are clustered at the audit level. Each column reports the coefficients γ and β estimated along with a specific set of controls for each procedure. Name revealed (resp. residence revealed) is a dummy indicating whether the tester revealed his/her name in the phone conversation of the first procedure (resp. residential origin in the second procedure). Characteristics at the level of the individual phone calls are: a dummy indicating whether the tester used the phone line with a number starting with 09 (indicating Internet box) and a dummy indicating whether he/she called first.

C Interviews

This appendix provides a description of our interviews.

We interviewed 29 real-estate agents between June and October 2010. We divided the Paris region into six sub-regions (Paris / inner ring / outer ring, each subdivided in white-rich and minority-poor). Each of the six sub-regions was assigned to a research assistant, who had to interview five real-estate agents from the area (one research assistant only performed four interviews). We opted for this sampling strategy to maximize heterogeneity within the area covered by our audit study.

In each sub-region, we drew a list of agencies from Internet postings on `http://www.seloger.com`, and research assistants were tasked with contacting agencies with a letter explaining the study (that we provided ourselves). There are two reasons why we avoided sampling the agencies we had audited. (1) We have been extremely cautious with avoiding suspicions of audit. Suspicious real-estate agents may purposefully distort their behavior, which would undermine the reliability of our findings. (2) It is impossible to infer discriminatory behavior from one particular pair of audits. Random reasons may explain why a rental unit was available to the white candidate and not to the minority candidate. Only consistent patterns of differential treatment reveal discrimination; these patterns only make sense within a quantitative framework. It would have been a logical stretch to treat these interviewees as discriminatory.

We did not compensate our informants and we guaranteed anonymity. Interviews consistently lasted about 60 minutes and took place at the agency. These were semi-structured interviews, for which we had provided the interview guide. The three main themes of the guide were: (1) context (informant's work history, description of her current job, of the neighborhood where she works); (2) a detailed account of how agents select tenants on behalf of landlords; (3) discrimination. To minimize social desirability concerns, we did not impose the discrimination theme; we let it develop organically from the discussion on selection practices. In such cases, an explicit question was also asked about "residence-based discrimination" in this section. We did not ask our respondents about their own discriminatory practices or their own attitudes toward minorities. All interviews were recorded and then transcribed; notes were taken during the interviews.

Summing the duration of the interviews, we obtained 1,665 minutes of audio recordings. The transcribed interviews single-spaced and written in 12 point font is 495 pages long. Additional information is available from the authors upon request.

Table C1 shows the recurrence of words referring to ethnic/racial criteria, distinguishing whether they were used by the interviewees or the interviewers. Interviewer and respondent used the word "origin" with equal frequency, whereas words indicative of ethnicity were predominantly used by interviewees. On average, the words reported in Table C1 are used more than twice as frequently by the interviewee.

Table C1. Recurrence of words referring to ethnic/racial criteria in the interviews

	Respondent	**Interviewer**	**Ratio**
Turk(s)	23	2	11.5
Africa, African, etc.	36	4	9.0
Black(s)	105	20	5.3
Arab(s)	26	6	4.3
White(s)	37	11	3.4
Race, racial	14	7	2.0
Racist, racism	45	23	2.0
Foreigner(s)	45	29	1.6
Origin	52	56	0.9
Ethnic, ethnicity	19	26	0.7
Total	402	184	2.2
Total per interview	13.9	6.3	2.2

References

Bonnet, François, Etienne Lalé, Mirna Safi, and Etienne Wasmer. 2011. "À la recherche du locataire «idéal»: du droit aux pratiques en région parisienne." *Regards croisés sur l'économie*, 9: 216–227.

Cediey, Eric, and Fabrice Foroni. 2008. *Discrimination in access to employment on grounds of foreign origin in France: A national survey of discrimination based on the testing methodology of the International Labour Office.* ILO.

Charrier, Rémi, Jean-Jacques Guillouet, Philippe Pauquet, and Mathilde Turpin. 2008. "Les conditions de logement en Île-de-France en 2006." *Île de France: à la page*, 298.

Coz, Gaëlle, and Genevieve Prandi. 2012. "Le parc locatif privé de l'agglomération parisienne au 1er janvier 2011." *Se loger: Repères – Collection de l observatoire des loyers*.

Decondé, Claire. 2012. "Première baisse significative de la vacance en Île-de-France depuis plusieurs décennies." *Île de France: à la page*, 381.

Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives*, 12(2): 101–116.

Jankel, Stéphanie, and Laurianne Salembier. 2008. "1996-2006: 10 ans de logement à Paris et en petite couronne." *Île de France: à la page*, 301.

Turner, Margery Austin. 1992. "Discrimination in urban housing markets: Lessons from fair housing audits." *Housing Policy Debate*, 3(2): 183–215.

Turner, Margery Austin, Rob Santos, Diane K Levy, Doug Wissoker, Claudia Aranda, and Rob Pitingolo. 2013. "Housing discrimination against racial and ethnic minorities 2012: Full report."

Yinger, John. 1986. "Measuring racial discrimination with fair housing audits: Caught in the act." *The American Economic Review*, 76(5): 881–893.

Yinger, John. 1998. "Housing discrimination is still worth worrying about." *Housing Policy Debate*, 9(4): 893–927.