CENTRO STUDI LUCA D'AGLIANO

WWW.DAGLIANO.UNIMI.IT

CENTRO STUDI LUCA D'AGLIANO
DEVELOPMENT STUDIES WORKING PAPERS

N. 466

September 2020

As Long as They are Cheap
Experimental Evidence on the Demand for Migrant Workers

Mauro Caselli *
Paolo Falco **

* School of International Studies & Department of Economics and Management, University of Trento
** Department of Economics, University of Copenhagen

As long as they are cheap

Experimental evidence on the demand for migrant workers*

Mauro Caselli† and Paolo Falco‡

September 1, 2020

Abstract

How does demand for migrant vs native workers change with price? We conduct an experiment with 56,000 Danish households (over 2 percent of all households in the country), who receive an advertisement from a cleaning company whose operators vary randomly across areas but meet the same quality standards and have equal customer ratings. When the operator has a migrant background, we find that demand is significantly lower than when the operator is a native. The gap, however, is highly sensitive to price, with demand for the migrant increasing steeply as the price falls. For an hourly pay close to the 25^{th} percentile of the earnings distribution in similar occupations (24 USD per hour), demand for the migrant is one-fifth of the demand for the native. A 25 percent reduction in the price makes the gap in demand disappear.

JEL codes: C93, J23, J61, J71.

Keywords: Migrants, discrimination, experiment, labour market integration, consumer preferences.

*The trial was pre-registered in the AEA RCT Registry (ID: AEARCTR-0005301) and reviewed by the ethical committe of the Department of Economics at the University of Copenhagen. We are grateful to Anik Ashraf, Stefano Caria, Mette Foged, Marco Piovesan, Alexander Sebald, Stefano Schiavo, Jean-Robert Tyran, and Sarah Zaccagni for their useful comments, and to Amalia Charlotte Pape for outstanding research assistance. We thank the Department of Economics at the University of Copenhagen for financial support and the staff of the leaflet distribution company Flyermdeling ApS for their precious collaboration. All errors are our own.

†School of International Studies & Department of Economics and Management, University of Trento: mauro.caselli@unitn.it

‡Department of Economics, University of Copenhagen: paolo.falco@econ.ku.dk

1 Introduction

Migration flows are an increasingly divisive issue across the world, causing deep cleavages within and between countries, and threatening the stability of international relations. The long-debated construction of a wall between Mexico and the US, and the tensions among European countries on how to handle migration into the EU are among the most vivid examples of such tensions. In any electoral campaign, openness to migrants is a controversial topic and political parties supporting more restrictive migration policies have been gaining strength across the world (Otto and Steinhardt, 2014; Barone et al., 2016; Halla et al., 2017; Edo et al., 2019; Caselli et al., 2020a;b).

The stand of advanced countries vis-á-vis migration is characterised by an apparent contradiction. On the one hand, an increasing share of voters in many countries show a strong opposition to welcoming migrants. On the other hand, there are entire sectors of the economy that rely heavily on migrant workers, who are willing to accept lower wages and skill downgrading (Card et al., 2012; Dustmann et al., 2013; Docquier et al., 2014; D'Amuri and Peri, 2014).

This paper uses a large-scale experiment conducted in Denmark to estimate differences in the demand for migrant vs native workers, holding constant their quality and the task to be performed. Furthermore, by means of experimental variation in prices, we are able to estimate compensating differentials that sway people away from their preferred type.[1] The trial was pre-registered.[2]

The experiment consists of distributing leaflets that advertise low-skill services (cleaning) of two workers who differ in their ethnic background *but meet the same quality stan-*

[1] Throughout the paper, the term "migrant" refers to workers with a migrant background (i.e., those who migrated to Denmark from other countries, as well as their descendants).
[2] AEA RCT Registry (ID: AEARCTR-0005301).

dards to 56,000 Danish households (over 2 percent of all households in Denmark). To signal equal levels of quality, we adopt an innovative design whereby we present the two workers as operators of the same Danish cleaning company and we report similar ratings from previous customers. This allows us to provide a credible signal of quality that is equally reliable for different types of workers. The fact that the services provided are very basic further reduces potential disparities in perceived quality. Using this setup, we investigate how demand for the two workers changes as we randomly vary their hourly price.[3] Furthermore, to investigate how demand for the two workers varies with their quality, we randomly change the customer ratings. The background of the two workers is signalled by their names as in much of the existing literature (Bertrand and Duflo, 2017). In particular, we follow recent studies on Denmark and choose a common Muslim name for the migrant in order to signal belonging to some of the largest migrant groups in the country (e.g., Hedegaard and Tyran, 2018).

Our results show that, on average, demand for the migrant is significantly lower than demand for the native. The gap, however, is highly sensitive to price, with demand for the migrant increasing steeply when customers are offered a lower price. For an hourly price close to the 25^{th} percentile of the earnings distribution (approx. 24 USD an hour), demand for the migrant is only one-fifth of the demand for the native. A 25% reduction in the price closes the gap. We also run an additional test by giving one group of customers an explicit choice between the two workers at a given (low) price.[4] Upon doing that, we find that the overwhelming majority chooses the native. This corroborates the conclusion that at any given price customers have a preference for native labour. When the only available option is to hire a migrant, many choose not to hire unless the price is sufficiently enticing. Finally, when we vary the customer

[3] As the rest of the literature, we measure demand using callback rates.

[4] We indicate that both workers are active in the local area and we provide their direct phone numbers to measure callbacks.

ratings of the two workers, we find suggestive evidence that the native is preferred to the migrant even when the latter has a higher quality rating.

We also demonstrate that our rating system is an effective tool to convey salient information about workers' quality and that such information plays a crucial role in customers' decisions. To this end, we devise an additional treatment consisting of a leaflet offering a choice between *two native workers* who differ in their rating. We find that the worker with the higher rating is strictly preferred to the other worker by *all* customers. This is strong evidence in support of the conclusion that our leaflets convey clear and salient signals about workers' quality.

This paper makes significant contributions to the literature on labour market discrimination.This literature has typically been unable to quantify the monetary value of discriminatory attitudes due to the challenge of exogenously varying prices (see early work by Bertrand and Mullainathan, 2004, and recent surveys by Rich, 2014; Bertrand and Duflo, 2017; Neumark, 2018; Baert, 2018). The study closest in spirit to our own is by Hedegaard and Tyran (2018), who run a field experiment with juveniles, but do not elicit demand for labour. Rather, they focus on workers' willingness to pay (in terms of foregone earnings) to avoid a co-worker with a migrant background. Our interest is in the employer-employee relationship, which is a crucial source of discrimination in the labour market and has been at the core of the literature. Furthermore, by working outside the lab with a large-scale experiment, we contribute to the literature in another important respect. The target of mail-out interventions is typically the (selected) population of firm managers. We target the general population of Denmark.

In addition, by presenting workers as employees of the same firm (with common quality standards and equal customer ratings), we make important inroads into holding quality constant and minimising the role of statistical discrimination, an often cited chal-

lenge in the literature (Bertrand and Duflo, 2017). We also go a step further and exploit our setup to vary workers' quality exogenously. This allows us to study how differences in quality affect hiring of migrants and natives. It also allows us to address the concern that differences in the variance of worker characteristics between groups poses a potential challenge for the interpretation of results from correspondence studies. We follow the recommendation advanced by Neumark (2012) and make variation in applicants' quality part of our design.

We also contribute to the literature that examines the consequences of migration for the labour markets of receiving countries. A common finding in this literature is that migrant inflows tend to worsen employment outcomes for previous migrants, but not for natives, indicating that the two are not perfect substitutes (e.g., Cortes, 2008; Ottaviano and Peri, 2012; Dustmann et al., 2013; Foged and Peri, 2016; Clemens et al., 2018). We corroborate this result by showing that even when a migrant and a native are employed by the same firm, perform the same tasks, and meet the same quality standards, they are perceived as imperfect substitutes by customers.

Finally, a related and growing literature investigates the impacts of migrant inflows on voters' attitudes and political outcomes in receiving countries (e.g., Hainmueller and Hopkins, 2014; 2015; Barone et al., 2016; Akcigit et al., 2017; Halla et al., 2017; Haaland and Roth, 2017; Caselli et al., 2020a;b). We contribute to the debate by shedding new light on the trade-offs voters are willing to make between workers' origins and the cost of hiring them.

The remainder of the paper is structured as follows. Section 2 describes the experimental design. Section 3 outlines our randomisation strategy and describes our outcome variable. Section 4 presents our empirical strategy and discusses our results. Section 5 concludes.

2 Experimental design

We deliver leaflets advertising the cleaning services of two male workers who differ in their origins (native vs. migrant background, as signalled by their names) to a random sample of 56,000 Danish households across rural and urban areas.[5] This amounts to over 2 percent of the country's population and more than 10 percent of the population in the municipalities we cover. The hourly rate charged by the workers is randomly varied (with a high and a low price) to elicit a demand schedule. In addition, we vary workers' quality, as signalled by the ratings of previous customers. Differences in callback rates between different types of workers are our primary outcome of interest. In this section, we discuss the different aspects of the experimental design. A detailed description of all the variants of the leaflet is provided in Appendix B.

2.1 The leaflets

The leaflets advertise the services of a Danish company that provides home-cleaning services. In order to propose different workers to different households, the company indicates that *only a certain worker operates in the area where the household is located.* The direct phone number for that worker is provided on the leaflet. In certain treatment arms (described below), the leaflet indicates that *two* workers are available in the area. Their phone numbers are both included in the leaflet and the household can choose which of the two to call.[6] In order to capture differences in callback rates between urban and rural areas, we use different phone numbers in different areas.

[5] Since our power calculations indicated that we could not run the analysis on both genders with the available sample size, we chose to focus on male workers. Given that migrants are predominantly male, this appears to be an optimal choice, but it will be important to extend the study to women in the future.

[6] The order in which the two appear on the leaflet is randomised.

2.2 Quality assurance

Presenting both workers as employees of the same Danish company helps us ensure that they meet the same quality standards in the eye of customers, partialling out channels of statistical discrimination. We take several steps to ensure this is attained.

First, in the leaflet the company explicitly says that all its workers have undergone careful training and they are all able to perform the tasks competently, effectively, and within the agreed time. This reinforces the message that they are all equally productive. The statement is strengthened by the fact that the services offered (which are listed on the leaflet) are rather basic (e.g., vacuuming, cleaning floors, mowing lawns), leaving limited room for quality differences among workers who have received the same training. Such tasks also require simple instructions, excluding the possibility that language barriers may play an important role. The leaflet also says that if the customer is not satisfied, full reimbursement will be offered (the Danish expression says "full satisfaction or no payment required").

A potential concern is that customers who are allocated different operators may make different inferences about the quality of the overall company, attaching different values to the information provided. To overcome this problem, the leaflet provides the names of all the operators in the company (four in total), but only the phone number of the one that is available in the area. This ensures that both customers who are allocated the native and those who are allocated the migrant have the same impression about the company overall (since the composition of its labour forces is the same). For the sake of balance, the personnel of the company is shown to be 50% migrant and 50% female, as indicated by the names listed.

Finally, the leaflet provides a direct signal of worker quality by reporting ratings from previous customers. These are displayed using a star-rating system on a scale from 0 to

5, like those commonly found online (e.g., Google and Amazon both use such a system). In most of the leaflets, the reported rating for the two workers is 4.9 (the next section describes how we exogenously vary it in certain treatments). The rating is accompanied by an indication of how many reviews were submitted by customers. The number we report is large (above 150) and very similar between the two workers we are comparing. Crucially, we know from one of our robustness checks that the rating system is salient in informing consumer choices (see the discussion in the next section). In a similar fashion, Laouénan and Rathelot (2020) take advantage of the star-rating system used on the Airbnb platform to identify the mechanisms underlying discrimination against ethnic-minority hosts.

2.3 Changing workers' price

The leaflets indicate an hourly price, which we vary randomly between a high value of 160 Danish Kroner (DKK) per hour, approximately equal to 24 USD, and a low price of 120 DKK (18 USD). To induce a sufficient number of calls and have adequate statistical power, and to make the service affordable to as large a share of households as possible, both our prices had to be in the lower part of the price distribution for services of this kind. A sensible choice for our high price seemed to be the mid-point between the median and the bottom (i.e., the 25^{th} percentile) of the distribution of hourly earnings for workers in comparable occupations.Our low price, on the other hand, is close to the minimum wage, as established by collective agreements in Denmark.[7] We further corroborated this choice by confronting our chosen prices with advertisements for comparable services available online.[8]

[7]Denmark does not have a national minimum wage.
[8]For instance, on *www.happyhelper.dk*.

2.4 Changing workers' quality

We also use the star-rating system described above as an instrument to induce exogenous variation in worker quality. For this purpose, we use the leaflets in which the company indicates two operators who are available in the area, giving customers a choice between the two. By showing customers different ratings for the two workers, we can test how this affects demand, holding the price constant (at the low level of 120 DKK). In one variant of the leaflet, the native has a higher rating (4.9) than the migrant (3.6). The ratings are swapped in a second version. We also have a third version in which the two workers have the same rating. Finally, in a fourth version of this treatment we use two popular native names ("Peter" and "Jens") and give them the high and low rating respectively. This is a robustness check to test whether our customer ratings are salient and we wish to test this when discrimination based on origin is not at play.

2.5 Indicating migrant background

To signal the migrant background of a worker, we follow a large literature relying on first names. We obtain data from the Danish statistical office (*Statistics Denmark*) on the most popular names among Danish residents with and without a migrant background. We choose the relevant migrant group based on size and in light of the existing literature on discrimination in the Danish labour market. In particular, we follow Hedegaard and Tyran (2018), who conduct a field experiment with juveniles to investigate ethnic prejudice in the workplace using Danish-sounding and Muslim-sounding names. Based on these criteria and on the available data, the chosen names are *Peter* and *Mohammad*.[9]

[9] Dahl and Krog (2018) also use Danish-sounding and Muslim-sounding names to investigate ethnic discrimination in the recruitment process in Denmark. Both Peter and Mohammad are among the names they use.

3 Sampling and outcome variable

3.1 Sampling and randomisation

We distribute the leaflets to 56,000 households, which constitute over 2% of all the households in Denmark.[10]. We randomly assign different versions of the leaflet to different areas of Copenhagen (local administrative units called "roder") and different towns across the neighbouring region of Sjælland. This allows us to explore heterogeneity between urban and rural areas. The leaflets are distributed by a professional distribution company that delivers them to all the households residing within a given geographic area. Randomisation is conducted by area, rather than at the individual household level, to avoid contamination across treatments.

The towns we target have at least 1,200 and up to 5,000 inhabitants. This is to ensure that a town could be covered in its entirety (a more feasible approach than giving the distribution company a precise list of addresses), while allowing us to cover a sufficiently large number of towns. Close to one third of the population in Sjælland lives in towns of such size.

The randomisation method explicitly seeks balance across a set of available covariates between towns and neighbourhoods in different treatment groups.[11] All the data used for the randomisation are obtained from Statistics Denmark.

[10] The total number was 2,728,132 in January 2020: https://www.dst.dk/en/Statistik/emner/befolkning-og-valg/husstande-familier-boern/husstande.

[11] The covariates include population size, share of men, and average age for areas in Copenhagen, and population size, average age, employment rate, and share of university graduates for areas outside Copenhagen where the set of available covariates was larger.

3.2 Measuring demand

The main outcome of interest is the callback rate for different types of workers under different conditions (area, price, rating, etc.). We measure this using two instruments. First, each manipulation is associated with a different phone number and we monitor the different lines over a period of about 2 months following the distribution of the leaflets.[12] Second, we set up a website where customers can review basic information about the company as well as leave messages and express their interest. In order to detect which leaflet customers received (and which worker they are interested in hiring), we ask them to quote a unique discount code included in the leaflet upon contacting us via the website. This allows us to attribute the contacts we receive via the website to requests for a specific worker. All customers who wrote through the website except for one quoted the code.[13]

Upon receiving the calls, we used an automated voicemail message saying that the worker was unavailable and would call back if some availability opened up in the following days. A similar message was sent by email to the customers who expressed an interest via the website. This allowed us to minimise the time cost for respondents who may have called back in the absence of that message.[14]

When counting the calls, we need to account for the possibility that the same customer may call multiple times and, in the case of leaflets with two phone numbers, that the

[12] The COVID-19 crisis hit during the course of the experiment and restrictions on economic activities were imposed from March 13, 2020. Our main results cover the period from January 23 to March 12, 2020. However, when we run the analysis over the full period we originally planned to cover (up to May 15, as indicated in our pre-registered plan), our results do not change (see Table 4 in Appendix A). Similarly, stopping the analysis at the the time when the very first case of COVID-19 was detected in Denmark (February 27) does not change our conclusions.

[13] Since we manipulated prices and quality levels between leaflets, the website did not contain this information.

[14] Our approach is consistent with John List's take on this issue, which is that if the research benefits society, and confers anonymity and just treatment to all subjects, the lack of informed consent seems defensible (Couture and List, 2008).

same number may call both of them. In most of the analysis, we will use the most conservative approach and count multiple calls from the same number as one contact. We also drop contacts from numbers that called both workers on the same leaflets, as that indicates a less clear-cut preference. In the next section, we will discuss different approaches and test the robustness of our results.

4 Results

In this section we show how demand for the migrant and the native workers changes as we vary the experimental conditions. For each manipulation, we show the difference in demand between the two workers *as a fraction of the demand for the native worker*. This provides an immediate indication of the gap (and of its statistical significance). Our results are summarised in Table 1.

<< **Table 1 here** >>

Our first result is that, on average, demand for the migrant is significantly lower than demand for the native. Across all the treatments in which the two have the same customer rating, the migrant receives 40 calls in total, while the native receives 73 calls, a gap of 45% (as shown in Figure 1). The gap is similar in urban and rural areas.

<< **Figure 1 here** >>

Our second result is that the gap in demand between native and migrant labour is highly sensitive to price (Figure 2). At the high price (160 DKK per hour, approximately 24 USD), demand for the migrant is one-fifth of the demand for the native. Lowering

the price by 25% (from 160 DKK to 120 DKK per hour) makes the gap disappear.[15]

The narrowing of the gap as the price falls is the result of demand for the migrant rising steeply while demand for the native does not change significantly (see Figure 4 in Appendix A). The most likely explanation for the low elasticity of the native's demand is that our high price is in fact rather low by Danish standards (a choice we had to make to incentivise callbacks and attain sufficient statistical power). Hence, it is reasonable to expect that lowering the price further may not lead to significantly higher demand for the native. The fact that for migrants demand does increase, on the other hand, is compatible with the idea that hiring a migrant entails an extra cost for consumers and the price drop constitutes a compensating differential.[16]

<< **Figure 2 here** >>

Next, we turn to the treatments in which we keep the price fixed (at 120 DKK, the low price) and manipulate the performance rating of the two workers. To increase statistical power, we carry out this part of the experiment by including the two workers on the same leaflet and giving customers the possibility to choose between them (two phone numbers are provided).[17] In one version of the treatment we assign a higher rating to the migrant (4.9/5 vs 3.6/5 for the native). Viceversa, in the other treatment the native has the higher rating. To verify consistency with the previous part of the analysis (where we only had one worker per leaflet and all workers had the same customer rating), we also have a version in which the two workers have the same rating.

[15] At the low price, demand for the migrant is only 5% lower than demand for the native and the difference is not statistically significant.

[16] An alternative hypothesis is that a very low price conveys a negative signal when associated with native workers whose reference prices are typically higher, while it is perceived as "normal" for a migrant. If this were the case, however, one may expect demand for the migrant to be higher than demand for the native at the low price. This is not what we observe. In fact, when we have two workers on the same leaflet at the low price (and we offer customers an explicit choice between them), as discussed next, the vast majority prefers the native.

[17] In this case, customers are told that *two* operators are active in their area.

We find that when the native has a higher rating, the demand gap is 100%. No customer prefers the migrant. By contrast, when the migrant has a higher rating, demand for him is still lower than demand for the native, albeit not significantly so.[18]

Similarly, when the two workers have the same rating, customers have an overwhelming preference for the native (Figure 3). This helps us refine the interpretation of the evidence above. The fact that demand for the migrant and for the native reached similar levels as the price dropped did not mean that at low prices people are indifferent between the two workers. It means that many people who would not hire a migrant for 160 DKK choose to do so when the migrant costs 120 DKK. However, *if given a choice between the two workers at the low price*, the most preferred option is by far the native.

<< **Figure 3 here** >>

To sum up, the results reveal that a large share (80% percent) of those who hire a native for 160 DKK would not hire a migrant for the same price, but appear to be persuaded to do so when the price drops to 120 DKK. The majority of those consumers, however, would still prefer a native, if given the opportunity to choose between a native and a migrant, even at the low price. The migrant remains the least preferred option by a considerable margin even when he has a higher performance rating than the native (albeit the difference is not statistically significant).

[18] It should be noted that this part of the analysis was designed to be exploratory. As a result we devoted a smaller sample size to it. Furthermore, statistical precision is hampered by the fact that having a lower rating for one of the workers leads to a decrease in the total number of calls received. This is to be expected, since average worker quality in the firm as a whole was lower. Differences in relative demand between the two workers are nonetheless marked.

Do customers pay attention to the customer rating?

The performance-rating system plays an important role in our experiment, as it allows us to signal quality and hold it constant between the two workers. Our approach follows in the footsteps of Hedegaard and Tyran (2018), who reveal past performance as a signal of current productivity. It also resembles the use of the Airbnb star-rating system in Laouénan and Rathelot (2020), who study ethnic discrimination in the rental market.

In order to test whether respondents take the rating into account and how salient the information is for their decisions, in one of the treatments we gave customers the choice between *two native* workers, one of whom has a higher star rating than the other.[19] In this case, where the ethnic background of the worker does not play a role, we find that consumer preferences are overwhelmingly in favour of the worker with a higher star rating (Table 1). In fact, the less skilled worker is not strictly preferred by *any* customer (i.e., the only calls he receives are from numbers that also call the other worker). This is a very strong indication that the star-rating system was effective in conveying information about worker quality that was salient in people's decisions.

Some exploratory evidence on the impact of attention discrimination

One hypothesis advanced in the literature is that migrants may be subject to attention discrimination from customers – a mechanism whereby knowledge of minority status impacts customers' level of attention to information (e.g., about skills) and leads to discrimination (Bertrand and Duflo, 2017).

Our evidence is consistent with this possibility, but identifying this channel is not a core objective of our design. Nonetheless, we can explore a feature of our experiment

[19] As for the first native, we picked one of the most common Danish male names for the second native (Jens).

to provide some exploratory evidence in this regard. Some respondents used the company's website rather than a phone call to expressed a preference for either worker. The site requires respondents who are interested in contacting the workers to fill out a contact form, where they have to indicate a discount code (used to identify which leaflet they received), as well as the name of the preferred worker. One could argue, therefore, that respondents who went through the website needed to pay greater attention to the information provided and had a chance to reflect more carefully about their choice, which was made more salient by having to fill out the form. This could potentially mitigate the problem of attention discrimination. Yet, when we run the analysis based only on the emails received via the website, all the main results are confirmed, despite the loss of precision due to the smaller number of contacts (Table 2 in Appendix A). Since we could not experimentally vary access to the website, this is only suggestive evidence indicating that attention discrimination is not playing a major role in driving our results.

Robustness to different methods of counting calls

Finally, we check the robustness of our results to the method we use to deal with multiple calls from the same number. Indeed, one could count every call separately and independently of the number calling (total) or count multiple calls from the same number as just one call (no multiple calls). With regards to the treatments with two names on the same leaflet, one could include all calls or only count calls from numbers directed to just one of the two workers (single preference only). The total number of calls we obtain when we apply each of these methods (plus the number of contacts received via the website) is shown in Table 6 in Appendix B.

As mentioned above, the main results presented so far are all based on the more conservative methods. That is, we count multiple calls from the same number as one. In

addition, for leaflets with two names, we drop the calls from numbers that contact both workers (i.e., we only use single-preference contacts), as those do not indicate a clearcut preference. When we include all the calls as a check, however, our conclusions do not change (Table 3 in Appendix A). We confirm the existence of a sizeable gap between the demand for the migrant and the native, which disappears as the price drops. We also confirm that even when the quality of the migrant is higher, demand for him remains below demand for the migrant (though the results become less precise, as expected since we introduce less clearcut preferences from customers that call both workers on the leaflet).

5 Conclusions

While anti-migration sentiments sweep through the world, large sectors of the economy in many countries rely on the availability of low-paid migrant labour. The rhetoric of "us first" appears to falter in the face of the economic gains that cheap labour brings. Using a large-scale randomised controlled trial with over 2% of all Danish households, this article quantifies the gap in the price of labour that sways employers of domestic workers from their preferences for natives, holding constant workers' quality and the task to be performed. We find that for an hourly pay close to the 25^{th} percentile of the earnings distribution demand for a migrant is one-fifth of the demand for the native. A 25% reduction in the wage makes a migrant as attractive as a native.

This is the first study to our knowledge that experimentally documents demand differences for migrant and native labour on such a large scale and with such a broad population. Moreover, our design makes important inroads into controlling for worker quality and excluding channels of statistical discrimination, which have been notoriously difficult to isolate. By presenting the two workers as employees of the same company

with equal customer ratings, we minimise uncertainty about their relative quality and hence the scope for inference based on statistical discrimination.

Going forward, it will be important to investigate whether existing biases can be reduced by means of policy interventions that aim to correct well-documented misperceptions about migration (e.g., European Commission (2018)). Recent experimental work on the US tests the effectiveness of information campaigns to correct such misperceptions (e.g., Grigorieff et al., 2018; Hopkins et al., 2019) and investigates their impact on people's support for immigration (Haaland and Roth, 2017). In addition, a study by Laouénan and Rathelot (2020) shows that richer information about the quality of a listing on the Airbnb platform can close the ethnic price gap in rentals. The results of our work highlight the importance of testing interventions that may impact employer choices.

References

Akcigit, U., Grigsby, J., and Nicholas, T. (2017). Immigration and the rise of American ingenuity. *American Economic Review*, 107(5):327–331.

Baert, S. (2018). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In Gaddis, S. M., editor, *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, chapter 3, pages 63–77. Springer International Publishing.

Barone, G., D'Ignazio, A., de Blasio, G., and Naticchioni, P. (2016). Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives' voting behavior. *Journal of Public Economics*, 136(C):1–13.

Bertrand, M. and Duflo, E. (2017). Field Experiments on Discrimination. In Duflo, E. and Banerjee, A., editors, *Handbook of Field Experiments*. Elsevier.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal?: A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4):991–1013.

Card, D., Dustmann, C., and Preston, I. (2012). Immigration, wages, and compositional amenities. *Journal of the European Economic Association*, 10(1):78–119.

Caselli, M., Fracasso, A., and Traverso, S. (2020a). Globalization and electoral outcomes: Evidence from Italy. *Economics & Politics*, 32(1):68–103.

Caselli, M., Fracasso, A., and Traverso, S. (2020b). Globalization, robotization and electoral outcomes: Evidence from spatial regressions for Italy. *Journal of Regional Science*, forthcomin.

Clemens, M. A., Lewis, E. G., and Postel, H. M. (2018). Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion. *American Economic Review*, 108(6):1468–1487.

Cortes, P. (2008). The effect of low-skilled immigration on U.S. prices: Evidence from CPI data. *Journal of Political Economy*, 116(3):381–422.

Couture, P. and List, J. A. (2008). Informed Consent in Social Science. *Science*, 322(5902):672a–672a.

Dahl, M. and Krog, N. (2018). Experimental evidence of discrimination in the labour market: Intersections between ethnicity, gender, and socio-economic status. *European Sociological Review*, 34(4):402–417.

D'Amuri, F. and Peri, G. (2014). Immigration, jobs, and employment protection: Evidence from europe before and during the great recession. *Journal of the European Economic Association*, 12(2):432–464.

Docquier, F., Ozden, Ç., and Peri, G. (2014). The labour market effects of immigration and emigration in OECD countries. *Economic Journal*, 124(579):1106–1145.

Dustmann, C., Frattini, T., and Preston, I. P. (2013). The effect of immigration along the distribution of wages. *Review of Economic Studies*, 80(1):145–173.

Edo, A., Giesing, Y., Oztunc, J., and Poutvaara, P. (2019). Immigration and electoral support for the far-left and the far-right. *European Economic Review*, 115:99–143.

European Commission (2018). Special Eurobarometer 469 Report Integration of immigrants in the European Union. Technical report, European Commission.

Foged, M. and Peri, G. (2016). Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics*, 8(2):1–34.

Grigorieff, A., Roth, C., and Ubfal, D. (2018). Does Information Change Attitudes Towards Immigrants? Representative Evidence from Survey Experiments.

Haaland, I. and Roth, C. (2017). Labor Market Concerns and Support for Immigration.

Hainmueller, J. and Hopkins, D. J. (2014). Public attitudes toward immigration. *Annual Review of Political Science*, 17.

Hainmueller, J. and Hopkins, D. J. (2015). The hidden american immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science*, 59(3):529–548.

Halla, M., Wagner, A. F., and Zweimüller, J. (2017). Immigration and Voting for the Far Right. *Journal of the European Economic Association*, 15(6):1341–1385.

Hedegaard, M. S. and Tyran, J. R. (2018). The price of prejudice. *American Economic Journal: Applied Economics*, 10(1):40–63.

Hopkins, D. J., Sides, J., and Citrin, J. (2019). The muted consequences of correct information about immigration. *Journal of Politics*, 81(1):315–320.

Laouénan, M. and Rathelot, R. (2020). Can Information Reduce Ethnic Discrimination? Evidence from Airbnb.

Neumark, D. (2012). Detecting discrimination in audit and correspondence studies. *Journal of Human Resources*, 47(4):1128–1157.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3):799–866.

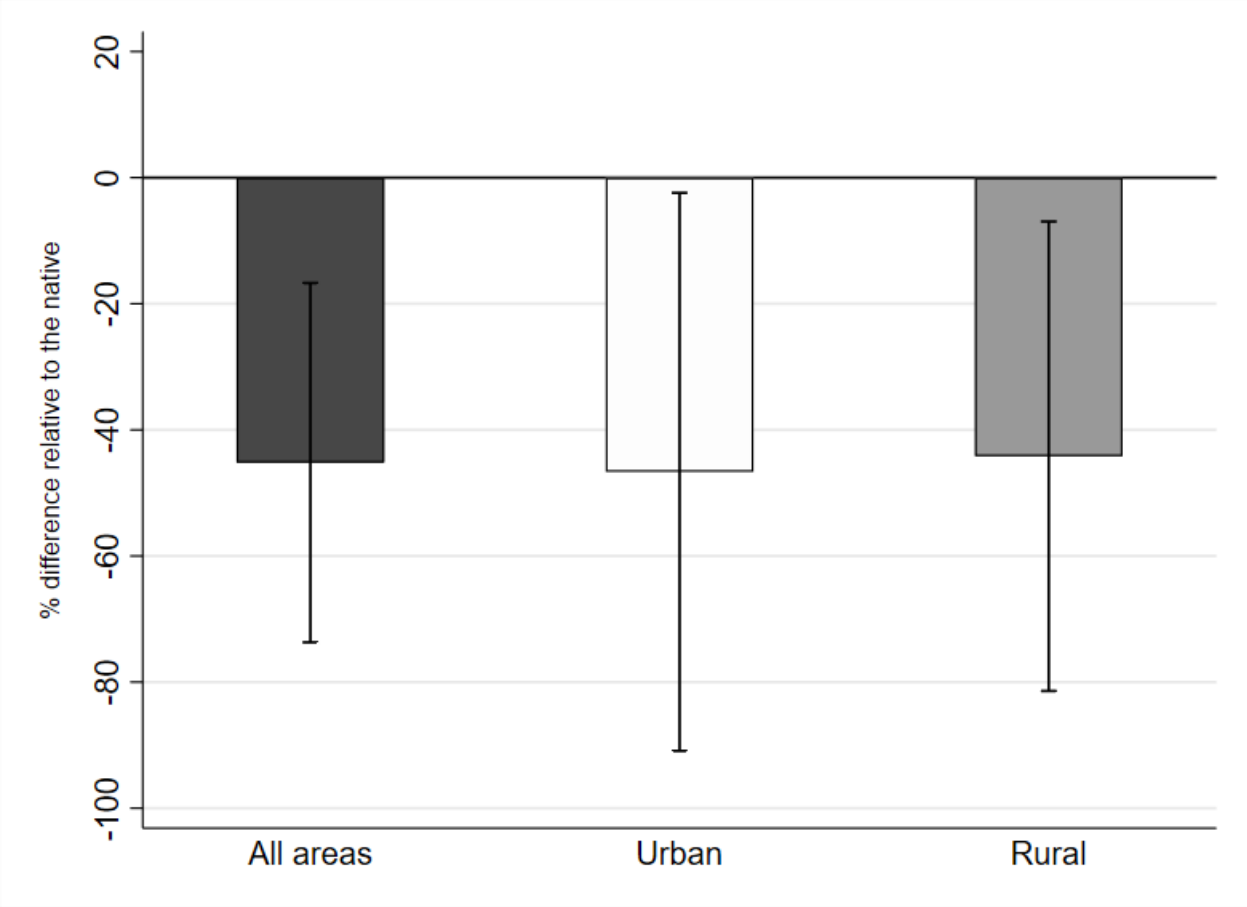
Ottaviano, G. I. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1):152–197.

Otto, A. H. and Steinhardt, M. F. (2014). Immigration and election outcomes – Evidence from city districts in Hamburg. *Regional Science and Urban Economics*, 45:67–79.

Rich, J. (2014). What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000. *IZA Discussion Papers*, (No. 8584).

Figures and Tables

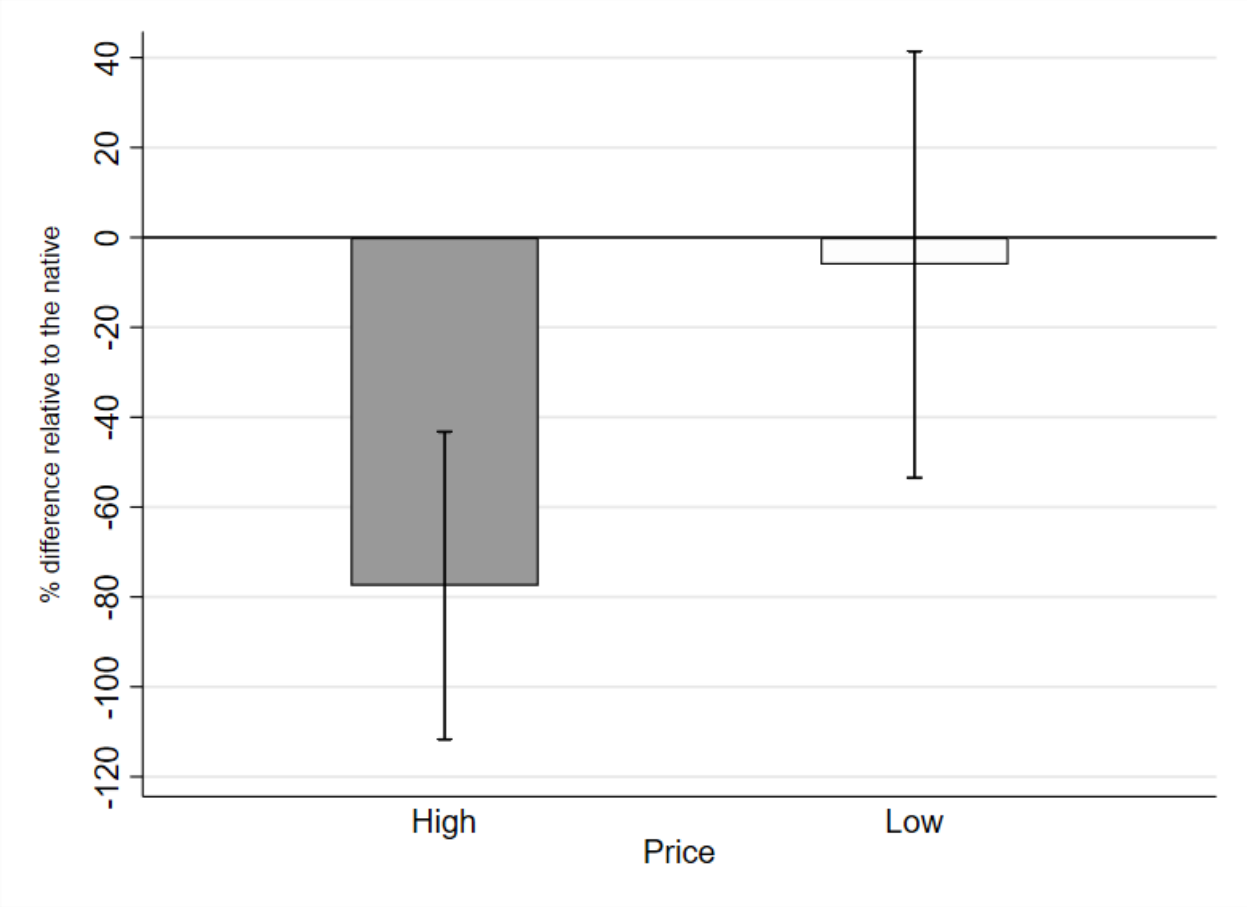
Figure 1: Demand for a migrant worker *relative to a native*



Notes: The figure shows percentage differences in callback rates for the migrant relative to the native by area. Whiskers indicate 95% confidence intervals. ***Urban*** includes callbacks to leaflets distributed in Copenhagen. ***Rural*** includes callbacks to leaflets distributed outside Copenhagen. The customer rating is the same for both workers in all leaflets and we pool the two prices. Only leaflets with one worker per leaflet are considered (leaflets 1-8). Callbacks are calculated on the basis of the third column (no multiple calls) of Table 6, i.e., only one callback per contact is recorded and contacts that express multiple preferences for different workers are excluded.

Figure 2: Demand for a migrant worker *relative to a native* at different prices

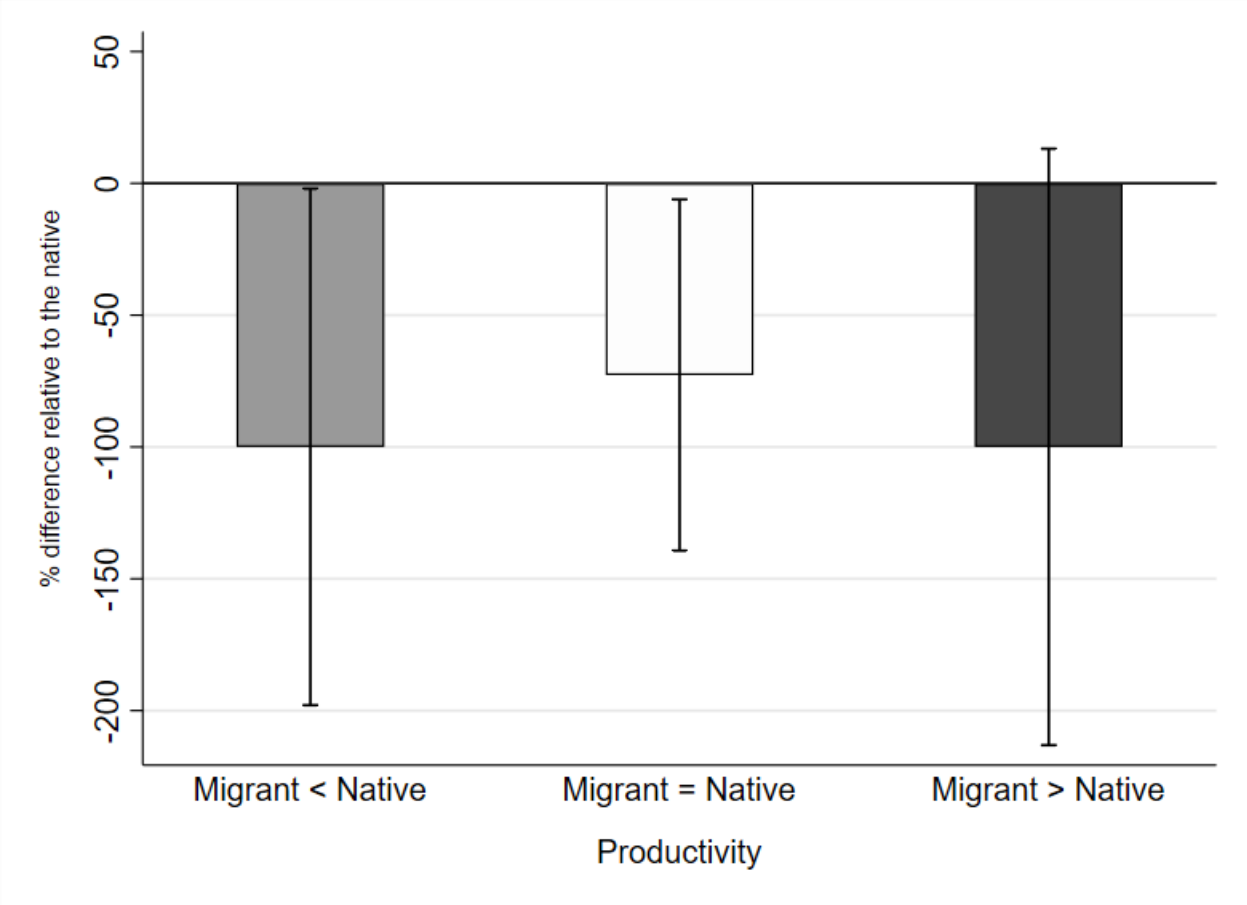
(holding worker quality constant)



Notes: The figure shows percentage differences in callback rates for the migrant relative to the native by price. Whiskers indicate 95% confidence intervals. ***High*** includes callbacks to leaflets with the high price (DKK 160). ***Low*** includes callbacks to leaflets with the low price (DKK 120). The customer rating is the same for both workers in all leaflets and we pool the urban and rural areas. Only leaflets with one worker per leaflet are considered (leaflets 1-8). Callbacks are calculated on the basis of the third column (no multiple calls) of Table 6, i.e., only one callback per contact is recorded and contacts that express multiple preferences for different workers are excluded.

Figure 3: Demand for a migrant worker *relative to a native* when worker quality differs

(holding the price constant)



Notes: The figure shows percentage differences in callback rates for the migrant relative to the native by quality level (i.e., past performance measured through customer ratings). Whiskers indicate 95% confidence intervals. ***Migrant*** $<$ ***Native*** includes callbacks to leaflets where the customer rating of the migrant is lower than the one of the native. ***Migrant*** $=$ ***Native*** includes callbacks to leaflets where the two workers have the same customer rating. ***Migrant*** $>$ ***Native*** includes callbacks to leaflets where the customer rating of the migrant is higher than the one of the native. Only leaflets where customers have a choice between two operators whose quality can differ are considered (leaflets 9-14). Callbacks are calculated on the basis of the third column (no multiple calls) of Table 6, i.e., only one callback per contact is recorded and contacts that express multiple preferences for different workers are excluded.

Table 1: Differences in callback rates between migrant and native

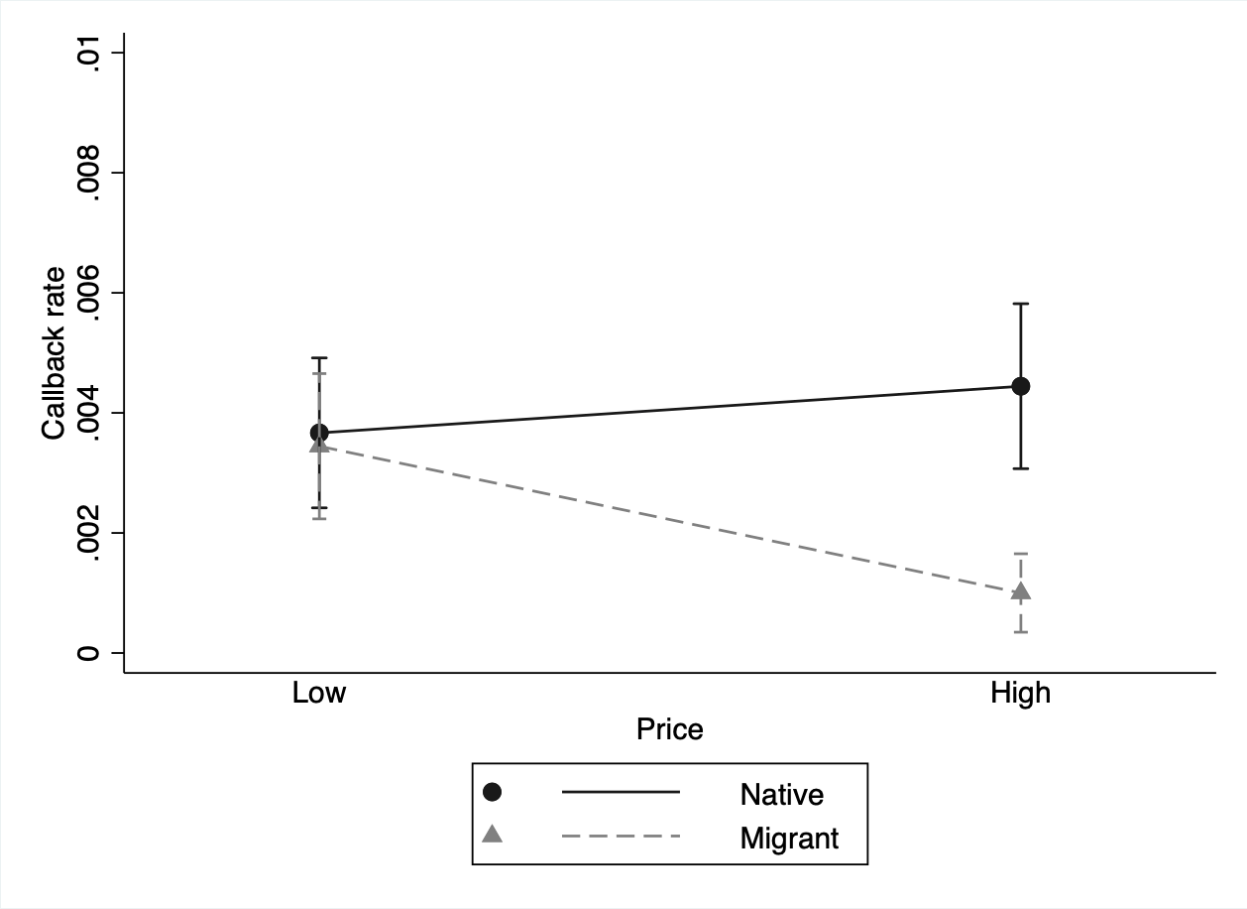
A. Pooling both prices, same worker quality				
	Migrant	Native	Difference	S.E.
All areas	0.0022	0.0041	-0.0018***	0.0006
Urban	0.0016	0.0030	-0.0014**	0.0007
Rural	0.0030	0.0054	-0.0024**	0.0010
B. Different prices, same worker quality				
	Migrant	Native	Difference	S.E.
Low Price	0.0034	0.0037	-0.0002	0.0009
High Price	0.0010	0.0044	-0.0034***	0.0008
C. Same price, different worker quality				
	Migrant	Native	Difference	S.E.
Migrant < Native	0.0000	0.0008	-0.0008**	0.0004
Migrant = Native	0.0006	0.0022	-0.0016**	0.0007
Migrant > Native	0.0000	0.0006	-0.0006*	0.0003
D. Same price, different quality between two natives (robustness)				
	Native 1	Native 2	Difference	S.E.
Native 1 > Native 2	0.0026	0.0000	0.0026***	0.0007

Notes: The table shows the difference in callback rates (number of callbacks divided by number of leaflets) between the two workers. Callbacks are calculated on the basis of the third column (no multiple calls) of Table 6, i.e., only one callback per contact is recorded and contacts that express multiple preferences for different workers are excluded. In all treatments, the number of leaflets distributed (N) is split equally between the two workers. Panel A pools both prices and only considers leaflets with one operator available in the area (leaflets 1-8). ***Urban*** includes callbacks to leaflets distributed in Copenhagen (N = 20,000). ***Rural*** includes callbacks to leaflets distributed outside Copenhagen (N = 16,000). ***All areas*** includes callbacks to leaflets from both areas (N = 36,000). Panel B separates leaflets by price and only considers leaflets with one available operator (leaflets 1-8). ***Low Price*** includes callbacks to leaflets with the low price (DKK 120; leaflets 1, 3, 5, 7; N = 18,000). ***High Price*** includes callbacks to leaflets with the high price (DKK 160; leaflets 2, 4, 6, 8; N = 18,000). Panel C considers the leaflets where customers have a choice between two operators whose quality can differ. ***Migrant = Native*** includes leaflets where the two workers have the same customer rating (leaflets 9-10; N = 10,000). ***Migrant > Native*** includes leaflets where the customer rating of the migrant is higher than the one of the native (leaflets 11-12; N = 10,000). ***Migrant < Native*** includes leaflets where the customer rating of the migrant is lower than the one of the native (leaflets 13-14; N = 10,000). Panel D (***Native 1 > Native 2***) is a robustness check based on a leaflet with two natives whose ratings differ (leaflets 15-16; N = 10,000). *, ** and *** indicate coefficients significantly different from zero at 10, 5 and 1% level respectively.

Appendix

A Additional Figures and Tables

Figure 4: Callback rates by price



Notes: The figure shows the callback rates for the migrant and native worker by price. Whiskers indicate 95% confidence intervals. ***Low*** includes callbacks to leaflets with the low price (DKK 120). ***High*** includes callbacks to leaflets with the high price (DKK 160). The customer rating is the same for both workers in all leaflets and we pool the urban and rural areas. Only leaflets with one worker per leaflet are considered (leaflets 1-8). Callbacks are calculated on the basis of the third column (no multiple calls) of Table 6, i.e., only one callback per contact is recorded and contacts that express multiple preferences for different workers are excluded.

Table 2: Differences in callback rates
(Only contacts by email)

A. Pooling both prices, same worker quality				
	Migrant	Native	Difference	S.E.
All areas	0.0004	0.0011	-0.0006**	0.0003
Urban	0.0003	0.0008	-0.0005	0.0003
Rural	0.0006	0.0014	-0.0008	0.0005
B. Different prices, same worker quality				
	Migrant	Native	Difference	S.E.
Low Price	0.0008	0.0010	-0.0002	0.0004
High Price	0.0001	0.0011	-0.0010***	0.0004
C. Same price, different worker quality				
	Migrant	Native	Difference	S.E.
Migrant < Native	0.0000	0.0004	-0.0004	0.0003
Migrant = Native	0.0002	0.0002	0.0000	0.0003
Migrant > Native	0.0000	0.0000	0.0000	0.0000
D. Same price, different quality between two natives (robustness)				
	Native 1	Native 2	Difference	S.E.
Native 1 > Native 2	0.0008	0.0000	0.0008**	0.0004

Notes: The table shows the difference in callback rates (number of callbacks divided by number of leaflets) between the two workers. Callbacks are calculated on the basis of the fourth column (emails only) of Table 6, i.e., only emails are recorded. In all treatments, the number of leaflets distributed (N) is split equally between the two workers. Panel A pools both prices and only considers leaflets with one operator available in the area (leaflets 1-8). ***Urban*** includes callbacks to leaflets distributed in Copenhagen (N = 20,000). ***Rural*** includes callbacks to leaflets distributed outside Copenhagen (N = 16,000). ***All areas*** includes callbacks to leaflets from both areas (N = 36,000). Panel B separates leaflets by price and only considers leaflets with one available operator (leaflets 1-8). ***Low Price*** includes callbacks to leaflets with the low price (DKK 120; leaflets 1, 3, 5, 7; N = 18,000). ***High Price*** includes callbacks to leaflets with the high price (DKK 160; leaflets 2, 4, 6, 8; N = 18,000). Panel C considers the leaflets where customers have a choice between two operators whose quality can differ. ***Migrant = Native*** includes leaflets where the two workers have the same customer rating (leaflets 9-10; N = 10,000). ***Migrant > Native*** includes leaflets where the customer rating of the migrant is higher than the one of the native (leaflets 11-12; N = 10,000). ***Migrant < Native*** includes leaflets where the customer rating of the migrant is lower than the one of the native (leaflets 13-14; N = 10,000). Panel D (***Native 1 > Native 2***) is a robustness check based on a leaflet with two natives whose ratings differ (leaflets 15-16; N = 10,000). ** and *** indicate coefficients significantly different from zero at 5 and 1% level respectively.

Table 3: Differences in callback rates
(including multiple calls from same customer)

A. Pooling both prices, same worker quality				
	Migrant	Native	Difference	S.E.
All areas	0.0037	0.0055	-0.0018**	0.0007
Urban	0.0030	0.0038	-0.0008	0.0008
Rural	0.0046	0.0076	-0.0030**	0.0012
B. Different prices, same worker quality				
	Migrant	Native	Difference	S.E.
Low Price	0.0061	0.0052	0.0009	0.0011
High Price	0.0013	0.0058	-0.0044***	0.0009
C. Same price, different worker quality				
	Migrant	Native	Difference	S.E.
Migrant < Native	0.0008	0.0018	-0.0010	0.0007
Migrant = Native	0.0018	0.0038	-0.0020*	0.0011
Migrant > Native	0.0006	0.0010	-0.0004	0.0006
D. Same price, different quality between two natives (robustness)				
	Native 1	Native 2	Difference	S.E.
Native 1 > Native 2	0.0054	0.0018	0.0036***	0.0012

Notes: The table shows the difference in callback rates (number of callbacks divided by number of leaflets) between the two workers. Callbacks are calculated on the basis of the second column (total) of Table 6, i.e., multiple callbacks per contact might be recorded. In all treatments, the number of leaflets distributed (N) is split equally between the two workers. Panel A pools both prices and only considers leaflets with one operator available in the area (leaflets 1-8). ***Urban*** includes callbacks to leaflets distributed in Copenhagen (N = 20,000). ***Rural*** includes callbacks to leaflets distributed outside Copenhagen (N = 16,000). ***All areas*** includes callbacks to leaflets from both areas (N = 36,000). Panel B separates leaflets by price and only considers leaflets with one available operator (leaflets 1-8). ***Low Price*** includes callbacks to leaflets with the low price (DKK 120; leaflets 1, 3, 5, 7; N = 18,000). ***High Price*** includes callbacks to leaflets with the high price (DKK 160; leaflets 2, 4, 6, 8; N = 18,000). Panel C considers the leaflets where customers have a choice between two operators whose quality can differ. ***Migrant = Native*** includes leaflets where the two workers have the same customer rating (leaflets 9-10; N = 10,000). ***Migrant > Native*** includes leaflets where the customer rating of the migrant is higher than the one of the native (leaflets 11-12; N = 10,000). ***Migrant < Native*** includes leaflets where the customer rating of the migrant is lower than the one of the native (leaflets 13-14; N = 10,000). Panel D (***Native 1 > Native 2***) is a robustness check based on a leaflet with two natives whose ratings differ (leaflets 15-16; N = 10,000). *, ** and *** indicate coefficients significantly different from zero at 10, 5 and 1% level respectively.

Table 4: Differences in callback rates between migrant and native
(including calls received during the Covid-19 epidemic)

A. Pooling both prices, same worker quality				
	Migrant	Native	Difference	S.E.
All areas	0.0031	0.0053	-0.0022***	0.0007
Urban	0.0021	0.0040	-0.0019**	0.0008
Rural	0.0044	0.0070	-0.0026**	0.0012
B. Different prices, same worker quality				
	Migrant	Native	Difference	S.E.
Low Price	0.0051	0.0044	0.0007	0.0010
High Price	0.0011	0.0062	-0.0051***	0.0009
C. Same price, different worker quality				
	Migrant	Native	Difference	S.E.
Migrant < Native	0.0000	0.0014	-0.0014***	0.0005
Migrant = Native	0.0010	0.0036	-0.0026***	0.0010
Migrant > Native	0.0000	0.0008	-0.0008**	0.0004
D. Same price, different quality between two natives (robustness)				
	Native 1	Native 2	Difference	S.E.
Native 1 > Native 2	0.0028	0.0000	0.0028***	0.0007

Notes: The table shows the difference in callback rates (number of callbacks divided by number of leaflets) between the two workers. Callbacks are calculated on the basis of the third column (no multiple calls) of Table 6, i.e., only one callback per contact is recorded and contacts that express multiple preferences for different workers are excluded. Callbacks received during the Covid-19 epidemic (from March 13 to May 15) are included in the analysis. In all treatments, the number of leaflets distributed (N) is split equally between the two workers. Panel A pools both prices and only considers leaflets with one operator available in the area (leaflets 1-8). ***Urban*** includes callbacks to leaflets distributed in Copenhagen (N = 20,000). ***Rural*** includes callbacks to leaflets distributed outside Copenhagen (N = 16,000). ***All areas*** includes callbacks to leaflets from both areas (N = 36,000). Panel B separates leaflets by price and only considers leaflets with one available operator (leaflets 1-8). ***Low Price*** includes callbacks to leaflets with the low price (DKK 120; leaflets 1, 3, 5, 7; N = 18,000). ***High Price*** includes callbacks to leaflets with the high price (DKK 160; leaflets 2, 4, 6, 8; N = 18,000). Panel C considers the leaflets where customers have a choice between two operators whose quality can differ. ***Migrant = Native*** includes leaflets where the two workers have the same customer rating (leaflets 9-10; N = 10,000). ***Migrant > Native*** includes leaflets where the customer rating of the migrant is higher than the one of the native (leaflets 11-12; N = 10,000). ***Migrant < Native*** includes leaflets where the customer rating of the migrant is lower than the one of the native (leaflets 13-14; N = 10,000). Panel D (***Native 1 > Native 2***) is a robustness check based on a leaflet with two natives whose ratings differ (leaflets 15-16; N = 10,000). ** and *** indicate coefficients significantly different from zero at 5 and 1% level respectively.

B Leaflet variants in detail

The combination of all the different manipulations we are interested in (name of the worker, price, area, whether one or two workers are available in the area, star ratings, and name order) gives rise to 16 different variants of the leaflet.[20] The different variants were randomly assigned to non-overlapping geographical areas.

The 16 variants can be divided in two groups. The first 8 focus on identifying the effect of changing the price of the worker (”Price” treatment). The second 8 hold the price fixed and focus on identifying the effect of changing the quality of the worker by changing the star rating (”Quality” treatment). Since the main focus of the analysis is on the first part, the number of leaflets distributed for the second half was lower.

All the following details are specified in a pre-analysis plan registered before the trial (AEA RCT Registry ID: AEARCTR-0005301).

B.1 Leaflets 1-8 (the ”Price” treatment)

The main treatment (the ”Price” treatment) consists of 8 types of leaflets. They indicate the phone number of a single worker who is available in the area (either the migrant or the native). The star rating is held constant at 4.9 for both.

Leaflets 1-4 are distributed in urban areas (neighbourhoods of Copenhagen) and are obtained from the combination of two workers (native vs. migrant) and two prices (high and low). Leaflets 5-8 are distributed in rural areas and are identical to the first four except for the phone numbers they display (which is necessary to investigate heterogeneity between rural and urban areas).

[20] Due to budget constraints, we cannot explore all the possible combinations. For instance, the leaflets with the phone numbers of two workers are only distributed in urban areas.

B.2 Leaflets 9-16 (the ”Quality” treatment)

The second treatment (the ”Quality” treatment) consists of an additional 8 types of leaflets, which are all sent to urban areas (for cost-related reasons). These leaflets differ from the first 8 insofar as they offer customers a choice between two different workers (customers are presented with the phone numbers of two operators active in their areas and they can choose which one, if any, they would like to call). The two workers differ in their quality, as signalled by the ratings of previous customers.

In Leaflet 9, native and migrant worker have the same quality level (4.9). In Leaflet 11, the migrant has higher quality (4.9 vs 3.6 for the native). In Leaflet 13, the native has higher quality (4.9 vs 3.6 for the migrant). Leaflet 15 replaces the migrant with another native and serves as a robustness check to measure the impact of differing customer reviews independently of workers' origins. Leaflets 10, 12, 14, and 16 are identical to 9, 11, 13, and 15 respectively, except for the fact that they swap the order in which the names of the workers appear (to exclude order effects).

Table 5 summarises this information and indicates how many leaflets were distributed for each variant. Table 6 indicates how many calls were received for each leaflet.

Table 5: Number of leaflets and other information by treatment

Treatment	Number	Copenhagen	Price (DKK)	Worker 1 (Stars)	Worker 2 (Stars)
1	5,000	Yes	120	Migrant (4.9)	
2	5,000	Yes	160	Migrant (4.9)	
3	5,000	Yes	120	Native (4.9)	
4	5,000	Yes	160	Native (4.9)	
5	4,000	No	120	Migrant (4.9)	
6	4,000	No	160	Migrant (4.9)	
7	4,000	No	120	Native (4.9)	
8	4,000	No	160	Native (4.9)	
9	2,500	Yes	120	Migrant (4.9)	Native (4.9)
10	2,500	Yes	120	Native (4.9)	Migrant (4.9)
11	2,500	Yes	120	Migrant (4.9)	Native (3.6)
12	2,500	Yes	120	Native (3.6)	Migrant (4.9)
13	2,500	Yes	120	Native (4.9)	Migrant (3.6)
14	2,500	Yes	120	Migrant (3.6)	Native (4.9)
15	2,500	Yes	120	Native 1 (4.9)	Native 2 (3.6)
16	2,500	Yes	120	Native 2 (3.6)	Native 1 (4.9)

Notes: The table shows the number of leaflets by treatment, whether the leaflets were sent to Copenhagen (Yes) or outside (No), the price advertised, the name(s) of worker(s) that could be employed, and their ratings (capturing their quality) in parentheses. The name used for Native 1 is Peter, for the Migrant is Mohammad, and for Native 2 is Jens.

Table 6: Number of callbacks (calls and emails) by treatment

Treatment	Callbacks Total	Callbacks No multiple calls	Callbacks Emails only
1	22	9	2
2	8	7	1
3	14	13	5
4	24	17	3
5	33	22	5
6	4	2	0
7	33	20	4
8	28	23	7
9	9	3	1
10	19	11	1
11	3	0	0
12	5	3	0
13	9	4	2
14	4	0	0
15	27	13	4
16	9	0	0

Notes: The table shows the number of callbacks (calls and emails) by treatment. The second column (total) shows the total number of callbacks. The third column (no multiple calls) eliminates multiple calls received from the same number, i.e., it records just one callback per contact, and excludes contacts that express multiple preferences for different workers, i.e., it includes only calls from numbers with a strong preference for just one worker (single preference only). The fourth column (emails only) records only messages arrived to the company's email address via the contact form on its website.