

Pre-analysis plan

Does perceived labor market competition increase prejudice between refugees and their local hosts? Evidence from Uganda and Ethiopia

Authors: Julie Bousquet, Anna Gasten, Marvin Kadigo, J-F Maystadt, Colette Salemi

Abstract

Prejudice towards out-group members has been documented in many settings. In refugee-host settings, existing literature has focused on quantifying discriminatory views of the local population towards refugees. Our study will analyze to what extent these prejudicial attitudes are fueled by concerns about labor market competition. In other words, does the stereotype of refugees “stealing jobs from natives” influence how communities perceive those seeking refuge in their country? Considering the dual nature of interactions, we will also analyze whether refugees hold prejudicial attitudes towards the local population or other refugees. To which extent are they discriminating against the out-group based on perceived labor market competition?

In our randomized survey experiment, we use a narrative vignette about a fictitious jobseeker to elicit inherent attitudes towards in-group/out-group members from $N = 4,000$ host community members (nationals who live near refugee populations) and $N = 4,000$ refugees. Two main characteristics of this fictitious jobseeker are randomized across respondents: their occupation (same vs. different than respondent's) and their citizenship status (national vs. refugee). We estimate the impact of the randomly allocated narrative on the respondent's reported attitudes about the hypothetical jobseeker and quantify the degree of perceived labor market competition depending on the narrative the respondent has been exposed to.

1. Data

1.1 Sampling Frame

The survey experiment is part of a large survey on refugee-host interactions and labor market integration that will be conducted in Uganda and Ethiopia. In Ethiopia, four main reporting domains have been identified: the city of Addis Ababa, the city of Jigjiga, the town of Qebribeyah, and the Qebribeyah refugee camp. In Uganda, we focus on the area of Isingiro around the Nakivale refugee settlement, as well as Kampala. The target populations are both nationals and refugee households residing in the mentioned locations and camp refugees.

The specific groups that we draw the respondents from are listed in the table below:

Table 1: Survey participants

No	Country	Rural/ Urban	Location name	Enumeration areas	N
1	Uganda	Rural	Isingiro District	Isingiro district: 75 Nakivale settlement: 40	1,000 hosts, 1,000 refugees
2	Uganda	Urban	Kampala District	Kampala: 150	1,000 hosts, 1,000 refugees ¹
3	Ethiopia	Rural	Somali Region	Jigjiga: 45 Qebribeyah: 35	1,000 hosts, 1,000 refugees
4	Ethiopia	Urban	Addis Ababa	Addis Ababa: 150	1,000 hosts, 1,000 refugees ¹

For the listing and allocation of targeted individuals in Ethiopia, we use the sampling frame provided by the Ethiopian Central Statistical Agency (CSA) that was constructed as part of the census planned for 2020 but not carried out due to the COVID-19 pandemic and deteriorating security conditions in the country. In Uganda, we use the sampling frame provided by the Ugandan Bureau of Statistics (UBOS). The primary selection units (PSUs) are enumeration areas (EA), based on the frame provided by CSA and UBOS. In the first stage of sampling, a first set of PSUs are selected with probability proportionate to size (PPS) sampling where size is the number of households in each EA. In order to secure a sufficient number of refugee households in the sample, a second set of PSUs are selected, consisting of PSUs bordering the PSUs in the first set that contain a sufficiently high number of households with at least one refugee.

Outside of camps, refugee populations are generally found in small numbers, particularly in urban settings such as that of Addis Ababa and Kampala. These settings can be characterized as constituting “rare populations,” where sampling frames are not readily available. To this end, we employed adaptive sampling for refugees living in urban areas that refers to a procedure where we draw the sample during

¹ Since we do not know how many refugees will be found ex ante, we are using an adaptive sampling frame for refugees in urban areas. We draw the sample during the field activities and rely on expertise from locals on where refugees can be found. The number of observations can slightly differ from numbers indicated in the table.

the field activity, as opposed to a standard sampling strategy where individuals are surveyed based on a listing exercise.

The survey experiment is embedded in a multi-topic survey, with several modules. This questionnaire comprises a household section, with questions on basic population characteristics of all household members. There is also a separate questionnaire for one randomly selected individual (RSI) in working age (18 - 65 years) within each household. The RSI questionnaire includes the experimental component where the narrative vignette and corresponding question items are introduced conditional on the characteristics of the respondent in order to explore attitudes and perceptions towards refugees as well as labor market integration.

Our experiment focuses on respondents from the host and refugee community who are either currently working or unemployed and actively seeking work. The randomization process for the experiment (randomization into four different treatment groups) is stratified by rural/urban and refugee/host identity as well as by country (Ethiopia/ Uganda). Individuals who are out of the labor force (i.e. neither working nor actively searching for employment) do not follow the experimental module. Since the fraction of respondents who are out of the labor force may vary across refugee/ host groups; rural/ urban areas and Ethiopia/ Uganda, we do not expect the same number of experimental participants across all four groups listed in Table 1. The number of participants may thus slightly deviate from our expectations (N) indicated in the table above, due to attrition and individuals who are out of the labor force.

1.2 Narrative vignette

A narrative vignette included in the questionnaire is the central component of our experiment. We induce exogenous variation by randomly assigning respondents to one of four possible narratives (Section 2.1.3). The narratives (example for Uganda - symmetrically for Ethiopia) take the following form:

“[AIDA/ROBERT] is a [GROUP G: Ugandan/ refugee living in Uganda]. [She/He] (has lived in Uganda [her/his] entire life and) moved to [Isingiro district/ Kampala] five years ago. [She/He] has been working as a [OCCUPATION O: Same occupation as respondent/ different occupation] for a long time so [she/he] has a lot of experience in [her/ his] occupation. [She/He] also speaks many Ugandan local languages and English very well. [She/He] enjoys working in this profession and would recommend [her/his] friends to work in the same sector. But while being a [OCCUPATION O: Same occupation as respondent/ different occupation] fulfills [her/him], [she/he] is sometimes very tired after work. Due to difficult circumstances, [she/he] has to change jobs while keeping her/his current profession. So far, she/he has struggled finding a job.”

The vignette appears in the survey after the labor modules but before the modules that include questions on perceptions of refugees. Immediately following the vignette are a set of questions about the fictitious worker, which are partly intended to make the respondent relate to the story about the fictitious worker, thereby reinforcing the elicitation of inherent attitudes, and partly serving as main outcome variables for our analysis.

1.3 Treatment assignment

Our experiment focuses on $N = 4,000$ employed or unemployed² host community members (nationals) and $N = 4,000$ employed or unemployed refugees split across rural and urban Uganda and Ethiopia. Each respondent is randomly assigned to one of four experimental arms:

Table 2: Treatment arms

	Same occupation	Different occupation
In-group	T1	T2
Out-group	T3	T4

- T1: narrative about an in-group member (refugee/ host) working in the same occupation as the respondent (N=1000 refugees; N= 1000 hosts)
- T2: narrative about an in-group member (refugee/ host) working in a different occupation as the respondent (N=1000 refugees; N= 1000 hosts)
- T3: narrative about an out-group member (refugee/ host) working in the same occupation as the respondent (N=1000 refugees; N= 1000 hosts)
- T4: narrative about an out-group member (refugee/ host) working in a different occupation as the respondent (N=1000 refugees; N= 1000 hosts)

1.4. Randomization of text elements within the narrative

The experimental module is programmed dynamically, with auto-filling gaps based on the respondent's own characteristics.

The *name* (Aida/ Robert) of the fictitious individual matches the gender of the respondent. This way, we enhance identification of the respondent with the fictitious individual. The chosen names are neutral with respect to ethnic and religious connotations in the Ethiopian and Ugandan context.

The *reference country* (Uganda/ Ethiopia) as well as the region where the fictitious individual is living (Kampala/ Isingiro district; Addis Ababa/ Somali region) matches the location of the respondent. We expect perceived labor market competition to be more relevant when referring to a fictitious individual in geographical closeness.

Group membership (host/ refugee) is determined ex-ante based on computer-based randomization. It determines whether the narrative refers to a "Ugandan" or "refugee living in Uganda" (symmetrically for Ethiopian respondents: "Ethiopian" or "refugee living in Ethiopia").

The *occupation level* (same occupation/ different occupation) is determined ex-ante based on computer-based randomization.

² Among the respondents who are currently unemployed, only those who are actively searching for a job are asked to participate in the experimental module. This approach has two reasons: (1) Measuring perceived labor market competition is less relevant for those individuals who are out of the labor force. By excluding them from the experimental group, we are reducing variance and noise in our measures and improve power. (2) Our narrative is constructed by inserting an occupation name that either matches or not the occupation of the respondent. For individuals who are out of the labor force, it is not possible to identify a relevant "individual" occupation.

The following steps are implemented for those individuals who are assigned to T1 or T3 (same occupation):

(1a) If the respondent is currently working: The enumerator is asked to shorten the string occupation indicated in the survey to a shorter string (1-2 words). E.g. "Primary school teacher for Maths" -> "Teacher"

(1b) If the respondent is currently not employed, but actively searching for a job: The enumerator asks what occupation the respondent is searching for and shortens the occupation to a 1-2 word string.

(2) This shortened string is used as occupation of the fictitious individual in the narrative.

The following steps are implemented for those individuals who are assigned to T2 or T4 (different occupation):

(1a) If the respondent is currently working: The enumerator asks whether this occupation requires at least secondary level of schooling. (NO -> A; Yes -> B)

(1b) If the respondent is currently not employed, but actively searching for a job: The enumerator asks what occupation the respondent is searching for and whether this occupation requires at least secondary level of schooling. (NO -> A; Yes -> B)

(2) One element is randomly drawn as occupation of the fictitious individual from the lists below. The skill level is expected to match the schooling level of the respondent, to avoid hierarchical judgements about the fictitious individual.

List A: Below secondary schooling	List B: Above secondary schooling
Farmer	Lawyer
Shopkeeper	Doctor
Waiter	Teacher
Cleaner	Banker
Security officer	Architect

(3) The enumerator is shown both the respondent's own occupation and the one randomly drawn by the computer out of the list above. He is asked to confirm that the two occupations are indeed different. If not: Step (2) and (3) are repeated. If yes: The randomly drawn occupation string is used as occupation of the fictitious individual in the narrative.

2. Fieldwork

2.1. Instruments

Survey data are collected with CSPro 7.5.1. The questionnaire is executed by enumerators who will read out the narrative in a vivid way, but are trained to stick to the wording above. Since the experimental module is part of the RSI-questionnaire, only one individual per household is answering the questions individually. We are varying only a few words within the narrative across our four treatment groups, thereby expecting a low risk of spill-overs across households within the same enumeration area (EA) which might bias our results. For the same reason, we expect individuals to be ignorant of the randomized structure within the survey, thereby keeping experimenter demand effects low.

2.2. Data Collection and Processing

The intervention starts mid-January 2022 and is expected to end by the end of June 2022. Data collection is executed by local teams and only shared with the supervising and funding institutions (FAFO and the World Bank) during the time of the data collection. All researchers involved in the design and analysis of the experiment will be granted access to the raw data only after the data collection ends, most likely from July 2022 onwards.

3. Empirical analysis

3.1. Balancing tests

We will test for balance between treatment arms, with respect to basic socio-economic characteristics (country, rural/ urban, age in years, gender, education, marital status, household size and composition, household financial status, labor market characteristics). Specifically, we will run t-tests of equality of means and Kolmogorov-Smirnov tests of equality of distribution.

3.2. Hypotheses

H1: Individuals discriminate against members of the out-group, compared to members of the in-group.

Comparing pooled measures of discrimination (see section “Primary Outcomes”) for treatment groups T1 and T2 to treatment groups T3 and T4 allows us to quantify discriminatory views of the host population towards refugees and vice versa - independently from their labor market status.

H2: Preconceived notions of labor market competition may have adverse effects on the perception of the other. By comparing pooled measures of discrimination for treatment groups T1 and T3 to treatment groups T2 and T4, we are able to analyze whether higher perceived labor market competition (proxied by a shared occupation) is linked to higher levels of individual discrimination and labor market discrimination (see section “Primary Outcomes”).

H3: Discrimination against members of the out-group is more pronounced when preconceived notions of labor market competition are strong. We are analyzing the interplay between labor market competition and out-group discrimination. We hypothesize that discrimination against the out-group is not happening in a universal way, but depends on the perceived threat of labor market competition arising from this group. We thereby test whether there is targeted discrimination where individuals discriminate against the out-group because of their perception of labor market competition. From a conceptual point, this would imply that the difference in attitudes towards the in-group with the same occupation (T1) and the out-group with the same occupation (T3) is larger than the difference in attitudes towards the in-group with different occupation (T2) and the out-group with different occupation (T4): (T1-T3) is significantly different from (T2-T4).

Our premise is that hosts and refugees may feel more threatened if they face direct competition in the labor market, while they do have more welcoming attitudes when they are relatively far in terms of labor characteristics to the out-group member. Discrimination is then triggered mostly by labor market competition, and not out-group membership per se.

3.3. Estimation strategy

Below, we list the regressions that we will run once the data are available. All results will be reported in the paper. We anticipate running additional regressions during the data analysis stage, in order to explore further the relationships that emerge from running our registered regressions. If so, the associated results will be labeled as such in the paper (Casey, Glennerster and Miguel 2012).

We use two different regression specifications. The first is designed to examine what prejudicial attitudes the respondents have about the fictitious individual described in the narrative. The second aims to examine whether these prejudicial attitudes apply to the entire out-group.

For all regressions we perform our analysis separately for host community members and refugee respondents. Given the large differences across the groups, we believe that the experiment will solicit different responses from refugees vs. hosts and that pooling these two groups will muddle outcomes.

3.3.1. Prejudice towards fictitious out-group member

Does the narrative elicit more self-reported prejudicial attitudes towards the fictitious individual who is an out-group member? Does it matter if this fictitious out-group member works in the same occupation as the respondent? We examine the three hypotheses raised above with the following regression, which will be run separately for the group of refugee and host respondents, given the different channels that are at play for those groups:

$$y_i = \alpha_0 + \alpha_1 OutGroup_i + \alpha_2 SameOcc_i + \alpha_3 (OutGroup_i \times SameOcc_i) + X_i' \gamma + u_i$$

Where

- $OutGroup_i$ is an indicator variable taking the value 1 if the respondent has been assigned to a treatment group with an out-group narrative (T3 or T4)
- $SameOcc_i$ is an indicator variable taking the value 1 if the respondent has listened to a narrative about a fictitious individual sharing his occupation (T1 or T3)
- X_i is a vector of demographic characteristics (e.g. age, gender, educational attainment, employment status, marital status, household size, country, rural/ urban) included to increase precision.

Here, the outcome variable y_i is a prejudice index, based on the questions asked immediately after the narrative about the fictitious individual. These questions are designed so that the respondent reports his attitudes towards the fictitious character in the narrative. These questions are all agree-disagree based on a five-point Likert scale.

The prejudice-index will be based on the following six questions:

1. I would feel comfortable when interacting with Aida/Robert
2. I would get along with Aida/Robert
3. I am comfortable if ... someone like Aida/Robert lives close to me
4. I am comfortable if ... someone like Aida/Robert marries a family member
5. Someone like Aida/Robert can work with me
6. Someone like Aida/Robert can become my supervisor

In a subsequent step, we will disentangle this prejudice index by focusing on three separate dimensions of interactions: The mean of questions 1 and 2 yields the degree of prejudice in the area of social interactions. The mean of questions 3 and 4 measures the level of prejudice in the area of private interactions. The mean of questions 5 and 6 provides a measure for work-related prejudice.

Since we are interested in the role that perceived labor market competition plays in enhancing out-group competition, we will test this channel in further detail. We will construct a measure of perceived competition towards the fictitious individual, based on the following two questions which are part of the experimental questionnaire module. We will analyze whether our narrative about a fictitious individual sharing the same occupation was effective in inducing higher levels of perceived labor market competition and can test whether individuals who report higher levels of perceived labor market competition are indeed more likely to discriminate against the fictitious individual:

7. I don't feel in competition with people like Aida/Robert if I would have to search for a job
8. Ultimately, I fear people like Aida/Robert take away my job

3.3.2. Prejudice towards entire out-group

As an extension of the main analysis described in 3.3.1, we plan to analyze whether anchoring the respondent's attitudes to a narrative with high versus low perceived labor competition has an effect on the respondent's perceptions about the entire out-group.

We will estimate the following regression exclusively on the subset of respondents who had been **assigned to T3 and T4** (out-group narrative):

$$y_i = \alpha_0 + \alpha_1 \text{SameOcc}_i + X_i' \gamma + u_i$$

Where

- SameOcc_i is an indicator variable taking the value 1 if the respondent has listened to a narrative about a fictitious individual sharing his occupation (T3)
- X_i is a vector of demographic characteristics (e.g. country, rural/urban, age, gender, educational attainment, employment status, marital status, household size, country, rural/urban) included to increase precision.

For our outcome variable we will examine questions from the subsequent module on "Integration and Perception":

9. Would you accept that an (out-group member) lives close to you?
10. Would you accept that an (out-group member) marries a family member?
11. Would you accept that an (out-group member) works with you?
12. Would you accept that an (out-group member) becomes your work supervisor?

Similar to the outcomes described in section 3.3.1., we will construct a measure of prejudice in the area of private interactions based on the mean of questions 9 and 10 and a measure of work-related prejudice based on the mean of questions 11 and 12.

We will also explore for the hosts how perceived labor market competition towards the out-group member further translates into altered acceptance towards refugees, based on the following questions:

13. Refugees take jobs away from locals.
14. Refugees receive too much support compared to nationals.
15. In general, you can trust refugees.

3.4. Heterogeneity analysis

The most obvious sample stratification is to test our hypotheses for separate groups (hosts versus refugees). Other sources of heterogeneity are interesting.

During the past decade, Ethiopia and Uganda have differed considerably in their refugee hosting policies. One of the major differences is that Ethiopia initially prohibited refugees from legally working while Uganda has for long allowed refugees to work. Do these differences in policy translate to different treatment effects? To examine this, we will re-estimate our results for Ugandans and Ethiopians separately.

Heterogeneity by key occupational characteristics is also important. For example, the experiment may lead to different results for employers, own account workers, and wage workers. But our power to examine heterogeneity by such characteristics depends on our sample distribution. To our best ability, we will examine different worker types when we have sufficient power.

If power allows, we will also test to which extent previous interactions with the out-group (contact hypothesis) and reported competition on the labor market (subjective measure) as well as de facto competition on the local labor market (objective measure: density of different occupations within host and refugee groups) link back to our results.

3.5. Clustered standard errors

Following two recent lines of thought in the statistical literature on clustering, we will estimate our regressions without clustered standard errors (Cameron and Miller 2015) and with standard errors clustered at the level of the respondent's enumeration area (Abadie et al. 2017).

Randomization is implemented at the individual level. Treatment assignment, our key regressor, is not expected to be correlated with individual characteristics of the respondents and participants are not assigned to treatment groups in specific clusters. Within-cluster correlation of the regressor is thus likely to be close to zero. Consequently, there is no need to cluster our standard errors (Cameron and Miller, 2015).

A more recent line of thought, summarized by Abadie et al. (2017) confirms that our standard errors will be correct without clustering, for the convenience sample we have (Blattmann, 2015). However, when making inference about the population of interest, we might have to cluster standard errors at the level of the enumeration area (EA). According to Abadie et al. (2017), whether or not clustering standard errors is necessary is essentially a design issue: Accounting for cluster-robust standard errors is thereby either justified by (1) the sampling design or (2) the experimental design. In our specific case, clustering is not an experimental design issue: Randomization is implemented at the individual and participants are not

assigned to treatment groups in specific clusters (e.g. all individuals of one EA). However, according to Abadie et al. (2017), our sampling design requires clustering standard errors if we want to infer results about the entire population: Our sampling strategy is based on a two-stage sampling procedure, where in the first stage a random set of PSUs (EAs) are randomly drawn and in the second stage a random set of households is selected within each PSU. This sampling procedure implies that there might be clusters in the population of interest that are not represented in the sample.

When inferring results on the *population*, we will cluster standard errors at the level of the enumeration area. Our sample consists of households stemming from about 495 enumeration areas, a large-enough number to implement cluster-robust standard errors.

3.6. Attrition

Since our respondents never find out their treatment assignments, and since respondents' experiences completing the survey do not depend on treatment assignment, we have no reason to expect the attrition rates to differ across groups. Given random treatment assignment, we also have no reason to expect that the groups will not be balanced. But mischance can always influence attrition and balance. In the unlikely event that these issues arise, we will do the following:

1. We will examine balance across treatment groups using the demographic controls included in X_i . If groups are unbalanced, X_i will account for these differences across groups.
2. If attrition rates differ across the two groups, we will use a bounds analysis to estimate the range of possible average treatment effects, following Lee (2009).

3.7. Power calculations

Our sample size was determined by the stakeholders implementing the survey. We perform a power calculation to ensure that this sample is of sufficient size to detect the effect of our experiment.

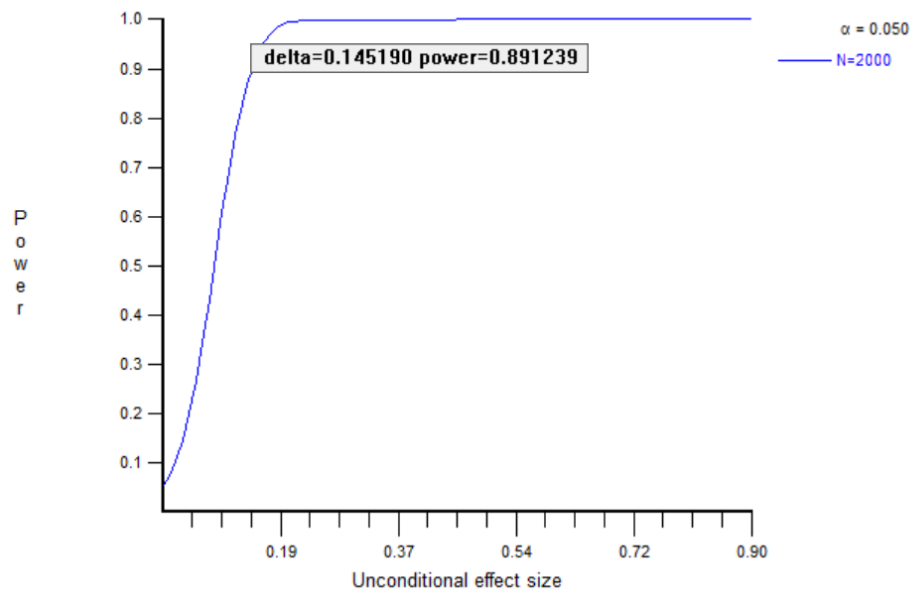
One approach to estimating the minimum detectable size for this specification is to use a simulation. But a simulation requires that we make assumptions about the sample properties. When an experiment has been well explored in past literature, researchers can use prior estimates to make more reliable predictions about the sample properties. Given the novelty of our experiment, there is no pre-existing literature that we can reliably draw from. Consequently, a simulation may not provide us with a reliable sense of power.

Instead, we rely on data subsetting. Recall that our sample is split between refugees (N=4,000) and hosts (N=4,000). Studying these populations separately will be important, so our baseline regressions will be run on sample size N=4,000. Our study contains four treatment arms (N=1,000 for each arm), and we are primarily interested in the coefficient of the interaction term. Our main regression will have the same power to detect a significant outcome on the coefficient of interest as in the case of 2,000 observations from 2 study arms to run the following regression:

$$y_i = \beta_0 + \alpha\beta_1 T_i + e_i$$

where $T_i = 1$ if the respondent receives a narrative about an out-group member in the same occupation and $T_i = 0$ if the respondent receives the opposite narrative (in-group member in different occupation).

We can use this reduced form to run a basic power calculation with $N=2,000$. The graph below shows the result of this power calculation using Optimal Design software.



Our power calculation suggests that we have 90% power to detect an effect of 0.15 SDs, and 80% power for an effect of 0.13 SDs.

When we are performing our analysis, we can also examine power ex-post. In the event that we estimate null results that we think require additional scrutiny (for example, in our heterogeneity analysis), we can use our data to accurately estimate power, to gain more confidence that the coefficient is truly zero.

3.8. Multiple hypothesis testing

While we are not testing many hypotheses, we are still evaluating more than one in a context in which incorrectly rejecting the null has considerable implications. Given this, we will estimate the false discovery rate (FDR) following Benjamini et al (2006). Anderson (2008) offers a method to implement their approach that yields q-values more easily comparable to p-values, and we will implement the Anderson approach.

3.9. Index construction

As mentioned, our outcome variables are composed of indexes produced using self-reported prejudice indicators. We will use the listed variables to construct indices following Anderson (2008). Specifically, we will use the ICW_INDEX package in Stata to generate our indices (Bouguen & Varejkova 2020).

References

- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (No. w24003). National Bureau of Economic Research.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484), 1481-1495.
- Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3), 491-507.
- Blattman, C. (2015): Clusterjerk, the much anticipated sequel. Blogpost available at: <https://chrisblattman.com/blog/2015/12/11/clusterjerk-the-much-anticipated-sequel/> [accessed: May, 25 2022]
- Bouguen, A., & Varejkova, T. (2020). ICW_INDEX: Stata module to aggregate the variables included in the varlist into an index.
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources*, 50(2), 317-372.
- Casey, K., Glennerster, R. and Miguel, E., 2012. Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *The Quarterly Journal of Economics*, 127(4), pp.1755-1812.
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3), 1071-1102.