Job Search Assistance for Refugees in Jordan: An Adaptive Field Experiment

Pre-Analysis Plan

Stefano Caria, Grant Gordon, Maximilian Kasy, Simon Quinn & Alex Teytelboym

February 9, 2019

1 Introduction

This document outlines our pre-analysis plan for using Thompson sampling to test the effects of job search assistance in Jordan. We will work with a population that is composed of Syrian refugees and Jordanian nationals residing in urban areas of Jordan. Jordan offers a unique opportunity to study refugees' integration in the labour market. While in many countries refugees do not have the right to work, or have to clear long bureaucratic hurdles before being granted such right, Jordan has recently passed a series of reforms – known as the Jordan Compact – which give refugees the right to work in low-skilled formal jobs. However, since the introduction of the reform, formal employment rates among refugees have failed to grow as rapidly as expected. In this project, we will evaluate interventions that are tailored to removing the main obstacles refugees may encounter in the formal labour market.

2 Summary of intervention

2.1 Description of treatments

We have designed three interventions to support job search in this context. These interventions will be denoted by $D \in \{0, 1, 2, 3\}$, where 0 is the control group. These interventions provide financial, information, and psychological support. Here is a brief description of each intervention:

(i) **Financial support**. We will offer an unconditional transfer of the value of 65 JOD (91.5 USD at the time of writing). This transfer is designed to support the recipient to pay for the cost of job search, which includes transport, grooming, time costs and, for at least some study participants, child care.

(ii) **Information support**. We will provide information on (i) the formal job search and application process, and (ii) what to expect from a formal job. This information will be delivered through flyers and a testimonial video describing the job search process from the eyes of a job-seeker. Syrian refugees have had little exposure to the formal labour market in Jordan and this intervention is designed to fill this information gap.

(iii) **Psychological support**. We will provide a job-search planning session and SMS reminders to help respondents overcome self-control problems related to job search. This intervention is motivated by recent evidence indicating substantial self-control problems and intention-behaviour gaps in job search (DellaVigna and Paserman, 2005; Caliendo et al., 2015; Abel et al., 2017).

Parallel experiment, described in separate pre-analysis plan In parallel, we will also evaluate a **direct placement intervention**. The direct placement intervention will be the subject of a separate pre-analysis plan, and we anticipate that it will be reported in a different academic paper to the paper that discusses these job search interventions.

2.2 Institutional background and sample selection

Jordan is a middle-income country that has experienced a very large inflow of refugees since the beginning of the Syrian crisis. In February 2016, Jordan signed the Jordan Compact, a policy that created a system of work permits for its population of refugees. In this project, we partner with an international NGO that works on refugee policy to evaluate a number of interventions designed to help refugees find work in Jordan.

We plan to interview a sample of individuals aged 18-45 that comprises both registered Syrian refugees and Jordanian nationals. To be eligible for this study, individuals have to be willing to take up low-skilled wage work in the immediate future. Individuals will be reached through a number of methods that have been piloted by the NGO. These include sampling from lists of refugees and unemployed people collected by community organizations and through door-to-door sampling in selected neighborhoods.

We plan to sample study participants continuously from February 2019 to September 2019. During these interviews, we will check eligibility and further collect information on jobseeker characteristics which will be used by our treatment assignment algorithm (see below).

Study participants will be randomized into a treatment group at the end of the interview.

Randomization will be carried out automatically by the handheld machine operated by the enumerator. The probability of being assigned to a given treatment group will be determined adaptively using the algorithm described in detail below.

After the initial interview, we will follow up with study participants at four points in time. Individuals will be administered full interviews one, three, and six months after the initial interview. We describe how we will analyze data from these interviews in Section 4.3 below. Further, we will administer a short phone interview focused on employment outcomes six weeks after the initial interview. This short interview will provide us with a binary measure of current wage employment – which we denote Y_{it} – that we will use for the treatment assignment algorithm.

2.3 Conditioning variables for treatment assignment

We will allow treatment probabilities to vary with exogenous covariates X. We will in particular condition on the following binary covariates, resulting in 16 strata:

(i) nationality (a dummy for whether the respondent is Jordanian, defined. as having a Jordanian national ID),

(ii) gender (a dummy for being female),

(iii) education (a dummy for having completed high school or more), and

(iv) work experience (a dummy for having experience in wage employment).

3 Treatment assignment

Our experiment has two **objectives**. Our primary objective is to get as many experimental participants into formal employment as possible. Our secondary objective is to test the effectiveness of alternative interventions. These two objectives suggest different experimental designs. Our design, described in this section, trades off these two objectives.

3.1 Prior

Let t denote the day of intervention, and let i index individuals within days. Note that we do *not* have a panel, so that individual i on day t is different from individual i on day t' when $t \neq t'$. Let x index strata, and let d denote treatment values. Let θ^{dx} be the corresponding average potential outcome. Assume

$$
\begin{aligned}
Y_{it}^{d}|(X_{it}=x,\theta^{dx},\alpha_0^d,\beta_0^d) &\sim Ber(\theta^{dx})\\
\theta^{dx}|(\alpha_0^d,\beta_0^d) &\sim Beta(\alpha_0^d,\beta_0^d)\\
(\alpha_0^d,\beta_0^d) &\sim \pi,
\end{aligned}
$$

where π is some prior distribution for (α^d,β^d) (specified in Section 3.4 below), and parameters are independent across treatment arms. It follows immediately that s_t^{dx}, the number of successes for treatment d and stratum x, follows a Beta-Binomial distribution given (α^d,β^d) and n_t^{dx}. Hierarchical Bayesian models of this form are discussed in Gelman et al. (2014), chapter 5.

3.2 Sampling from the posterior

Denote by $\boldsymbol{\theta},\boldsymbol{m}_t,\boldsymbol{r}_t$ the vectors of parameters, cumulative trials, and cumulative successes indexed by both d and x, and by $\boldsymbol{\alpha},\boldsymbol{\beta}$ the vectors of hyperparameters indexed by d. Let ρ index replication draws, with ρ ranging from 1 to R. We sample from the posterior distribution of $(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta})$ given $\boldsymbol{m}_{t-1},\boldsymbol{r}_{t-1}$ using the following Markov Chain Monte Carlo algorithm.

(i) Gibbs step:
Given $\boldsymbol{\alpha}_{\rho-1}$ and $\boldsymbol{\beta}_{\rho-1}$, draw θ^{dx} from the $Beta(\alpha_\rho^d+s^{dx},\beta_\rho^d+m^{dx}-s^{dx})$ distribution.

(ii) Metropolis steps:

- Given $\boldsymbol{\beta}_{\rho-1}$ and $\boldsymbol{\theta}_\rho$, draw α_ρ^d by sampling from a normal proposal distribution (truncated below), and accept this draw if an independent uniform draw is less than the ratio of the posterior for the new draw, relative to the posterior for $\alpha_{\rho-1}^d$. Otherwise set $\alpha_\rho^d=\alpha_{\rho-1}^d$.
- Similarly for $\boldsymbol{\beta}_{\rho-1}$ given $\boldsymbol{\theta}_\rho$ and $\boldsymbol{\alpha}_{\rho-1}$.

Markov Chain Monte Carlo methods are reviewed in Gelman et al. (2014), chapter 11. This algorithm converges to a stationary distribution that equals the joint posterior of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ given $\boldsymbol{m}_t,\boldsymbol{r}_t$. In particular, we have that the posterior probability of a treatment d being optimal given x, in the sense of maximizing the probability of employment, is given by

$$
P\left(d=\arg\max_{d'}\theta^{d'x}|\boldsymbol{m}_t,\boldsymbol{r}_t\right)=\operatorname*{plim}_{R\to\infty}\frac{1}{R}\sum_{\rho=1}^{R}\mathbf{1}\left(d=\arg\max_{d'}\theta_\rho^{d'x}\right),
$$

where θ^{dx} is the probability for a participant of type x to find employment when receiving treatment d.

3.3 Assignment algorithm

Denote

$$\hat{p}^{dx} = \frac{1}{R} \sum_{\rho=B}^{B+R} \mathbf{1}\left(d = \arg\max_{d'} \theta_{\rho}^{d'x} \right),$$

the estimated posterior probability that d is the optimal treatment given x. B is the so-called "warm-up" period for the MCMC algorithm.

Two popular algorithms for assigning treatment are fully random assignment and Thompson sampling. Our experiment will use a combination of these. **Fully randomized sampling** assigns treatment d with probability 0.25 to units in every stratum. This maximizes power for tests of non-zero treatment effects.

Thompson sampling assigns treatment d with probability $\hat{p}^{dx}$ to units in stratum x. Thompson sampling minimizes expected regret (cf. Bubeck and Cesa-Bianchi 2012), or equivalently maximizes average outcomes, in the large sample limit. In particular expected regret only grows at a logarithmic rate with the number of experimental units.

Our primary goal is to maximize the labor market outcomes of experimental participants, but power for all treatment comparisons is a secondary objective. In **our sampling procedure**, we therefore assign treatment d to units in stratum x with probability

$$(1 - \gamma) \cdot \hat{p}^{dx} + \gamma \cdot 0.25.$$

where γ is the share of observations that are randomized between treatment arms with equal probability.

In our experiment, there will be **delayed observability** of the employment outcomes Y_{it} – we measure employment 6 weeks after the intervention for each participant. As a consequence, treatment assignment is conditioned only on the outcomes of participants from 6 weeks before, or earlier, and we assign participants in the first 6 weeks randomly to each treatment arm with probability 0.25.

3.4 Tuning parameters

The algorithm as described above depends on several tuning parameters. We will use the following parameters for our prior and treatment assignment algorithm:

- We choose a **prior** for the hyper-parameters (α, β) with density equal to $(\alpha + \beta)^{-2.5}$, up to a multiplicative constant. In doing so, we follow the recommendation of Gelman et al. (2014), p.110, for picking a "non-informative" hyper-prior.

- The **share of observations fully randomized** is set to $\gamma = 0.2$, which implies that the probability of being assigned to each treatment is bounded below by 0.05.
- For Markov Chain Monte Carlo draws from the posterior, we use a **warm-up** period of $B = 1,000$, and then draw $R = 10,000$ **replications**; averaging over these gives our estimated posterior distribution. These values are generously chosen relative to standard recommendations (cf. Gelman et al. (2014) chapter 11), making convergence likely. In our simulations these values yield stable posterior probabilities.

4 Outcomes

This section describes the estimates, credible sets, and randomization tests that we will report in our paper.

4.1 Estimates

Let p^x be the sample share of participants in stratum x and, as before, let $\hat{p}^{dx}$ denote the posterior probability that some treatment d is optimal, estimated using MCMC replication draws; similarly for $\hat{E}$ which denotes the estimated posterior expectation.

Policy comparisons The main goal of this experiment is to find treatments that maximize the probability of employment, for each demographic subgroup. We want to do so both within the experiment (for the experimental participants themselves), and after the experiment (in a possible roll-out of the best policy for a wider population).

To evaluate our success, we will report the following welfare contrasts. First, within the experiment, we compare the average potential outcomes for the **actually chosen treatment** assignment to the average that would have obtained under **random assignment**,

$$\Delta_1 = \tfrac{1}{N} \sum_{i,t} \left(\hat{E}\left[\theta^{D_{it} X_{it}}\right] - \tfrac{1}{4} \sum_{d} \hat{E}\left[\theta^{d X_{it}}\right] \right)$$

This estimate measures how much better we did for our experimental participants, compared to a conventional design with fully random assignment.

Second, after the experiment, we wish to compare the **optimal targeted policy**, and the

optimal non-targeted policy, to the default of **no intervention** (treatment 0),

$$\Delta_2 = \sum_x \left(\max_d \hat{E}\left[\theta^{dx}\right] - \hat{E}\left[\theta^{0x}\right] \right) p^x,$$
$$\Delta_3 = \max_d \sum_x \left(\hat{E}\left[\theta^{dx}\right] - \hat{E}\left[\theta^{0x}\right] \right) p^x.$$

The definition of Δ_2 allows the optimized d to depend on x, while the definition of Δ_3 requires the same d to be implemented for all x. In particular $\Delta_2 \geq \Delta_3 \geq 0$ by construction.

Average potential outcomes In addition to these policy comparisons, we will also report more conventional estimates, of average potential outcomes, both conditional on covariates and unconditional, and average treatment effects. That is, we will report $\hat{E}\left[\theta^{dx}\right]$ for all x and d, as well as $\sum_x \hat{E}\left[\theta^{dx}\right] p^x$ for all d, and, for $d = 1, 2, 3$,

$$\delta^d = \sum_x \left(\hat{E}\left[\theta^{dx}\right] - \hat{E}\left[\theta^{0x}\right] \right) p^x.$$

Probability optimal Lastly, for each treatment we will report again the posterior probability that it is optimal, both conditionally and unconditionally. That is, we will report $\hat{p}^{dx}$ for all d and x, as well as $\sum_x \hat{p}^{dx} p^x$ for all d.

4.2 Tests and Inference

Primary: Bayesian inference Our primary form of inference will be Bayesian, based on the hierarchical default prior described in Section 3.1 above. To construct credible sets (i.e., sets that have a given posterior probability of containing the true parameters), we will report 0.025 and 0.975 quantiles, based on MCMC draws. We will do so for all our estimates listed in the previous section. This yields sets that have a posterior probability of 95% to contain the true parameters, conditional on the data of the experiment.

We would like to emphasize that standard Bayesian inference, in contrast to standard frequentist inference, remains valid for adaptive designs such as ours, since the likelihood function is not affected by adaptivity. In large samples, however, our credible sets also have 95% frequentist coverage probability, i.e., they are confidence sets in the usual sense; cf. van der Vaart (2000), chapter 10.

As secondary information, we will also report 0.05 and 0.95 quantiles of the posterior distribution, as well as posterior standard deviations.

Secondary: Randomization inference Additionally, we will provide randomization-based p-values that are valid under the sharp null hypothesis that there are no treatment effects, i.e., under the null that $\theta^{dx} = \theta^{d'x}$ for all d, d', x. Under this null, we can generate counterfactual data by re-running our assignment algorithm repeatedly, leaving outcomes as they are in our data, but generating new treatment assignments. The distribution of test-statistics over this re-randomization distribution can be used to construct critical values and p-values that are exact in finite samples, under the sharp null.

We will provide randomization-based p-values for **one-sided** tests using the welfare contrasts Δ_1, Δ_2, and Δ_3. These provide tests with power, respectively, for the questions "Did our sampling procedure improve in-sample welfare relative to full randomization?" and "Does the optimal treatment (conditional or unconditional) improve welfare in the population relative to the control treatment?"

We will additionally provide randomization-based p-values for **two-sided** tests using the estimated treatment effects δ^1, δ^2 and δ^3. These provide tests with power for the questions "Are average outcomes different under treatment d, relative to the control treatment?"

Re-optimization and inference A key question in defining both credible sets and randomization inference is whether d has to be "re-optimized" across different draws from the posterior or from the randomization distribution, in the definition of Δ_2 and Δ_3 (cf. Andrews et al. 2018). As it turns out, the answer is "no" for credible sets, and "yes" for randomization inference. The reason is that credible sets are defined in terms of the distribution of θ conditional on the observed data. The optimized d is a function of these data, and thus non-random from the perspective of Bayesian inference. By contrast, randomization-based p-values are defined in terms of the distribution of treatment assignments, conditional on the sample vector of covariates and potential outcomes. The optimized d does depend on the realized treatment assignment, and thus varies across draws.

Relatedly, note that there is no reason that the conclusions of Bayesian inference and randomization inference should coincide, since they provide answers to different questions.

4.3 Secondary outcomes – follow-up interviews

We will carry out follow-up interviews with study participants one, three and six months after treatment. In these interviews, we will collect data about employment and earnings. Further, we will measure a number of outcomes that we expect to change in case employment and earnings are affected: well-being, social integration, and migration decisions. We will consider the following outcomes:

1 *Wage employment*: A dummy for whether the respondent currently has a wage-paying job.

2 *Earnings*: Earnings from main job (0 if not in wage employment).

3 *Well-being*: A well-being index that comprises (i) monthly expenditure, (ii) life satisfaction (0-10 scale), (iii) an indicator of negative affect (feeling anxious on previous day on a 0-10 scale), (iv) an indicator of positive affect (feeling happy on previous day on a 0-10 scale). (We will construct this index using the method outlined in Anderson (2008).)

4 *Social integration*: An index of seven social integration questions (each question asks the respondent to report on a scale from 1 to 5 how much he or she agrees with a given statement, for example, 'I feel connected to Jordan'). (We will construct this index using the method outlined in Anderson (2008).)

5 *Migration*: A dummy for whether the respondent intends to migrate to a different country in next 12 months (this does not include return migration to Syria).

For each of these outcomes W_{it}^j, $j = 1, \ldots, 5$, we will report **weighted averages** of the form

$$\beta_j^d = \tfrac{1}{N} \sum_{it} \frac{\mathbf{1}(D_{it} = d)}{p^{dx}} \cdot W_{it}^j,$$

where

$$p^{dx} = \frac{\sum_{it} \mathbf{1}(D_{it} = d, X_{it} = x)}{\sum_{it} \mathbf{1}(X_{it} = x)}.$$

and N is the total number of experimental participants.

For each of these β_j^d, we will report **two-sided p-values** based on **randomization inference** under the sharp null, as described for our primary outcomes above.

4.4 Heterogeneity by nationality

We will repeat this analysis (in particular, the analysis of section 4.3) separately for those who hold a Jordanian national ID and those who do not.

4.5 Exploratory analysis

We anticipate we will run additional exploratory analysis, motivated by the results, to investigate mechanisms.

References

Abel, M., Burger, R., Carranza, E., and Piraino, P. (2017). Bridging the intention-behavior gap? The effect of plan-making prompts on job search and employment. *Working Paper*.

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.

Andrews, I., Kitagawa, T., and McCloskey, A. (2018). Inference on winners. Technical report, cemmap working paper.

Bubeck, S. and Cesa-Bianchi, N. (2012). Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Foundations and Trends in Machine Learning*, 5(1):1–122.

Caliendo, M., Cobb-Clark, D. A., and Uhlendorff, A. (2015). Locus of control and job search strategies. *Review of Economics and Statistics*, 97(1):88–103.

DellaVigna, S. and Paserman, M. D. (2005). Job search and impatience. *Journal of Labor Economics*, 23(3):527–588.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Taylor & Francis.

van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press.