Effects of migration political attitudes and social trust: evidence from green card lottery PS 531: Pre-analysis plan

Ikromjon Tuhtasunov

May 15, 2021

1. Describe a substantive question in social science. What theory are you assessing? Why should anyone care?

Politicians in Uzbekistan have long delayed institutional reforms citing people's unreadiness for democracy. Some even say that democratic institutions are not suitable to the Uzbek culture and people. Following this logic would suggest that Uzbeks as a nation are not compatible with democratic views. I want to disprove such claims by comparing political attitudes of two groups—Uzbek immigrants in the US and their non-migrant compatriots in Uzbekistan. The green card lottery that randomly allocates an opportunity to migrate to the US makes these two groups closely comparable. I expect that the green-card holders in the US demonstrate higher levels of support for democracy thanks to their exposure to democratic institutions in the US. Their experience of living in a democracy should also allow them to have more sophisticated understanding of democracy in comparison to unsuccessful lottery participants in Uzbekistan. In addition, I expect green card holders to demonstrate higher trust towards different ethnicities due to their interactions with diverse communities.

Multiple factors at the individual and societal levels affect how people view democracy (Canache 2012). Experience of living within a well-functioning democracy leads many migrants to acquire democratic values and pro-democratic views (Diamond, 1994). They participate equally in electoral activities (Bilodeau et. al 2010). Many migrants who return to their home countries serve as agents of democracy in spreading liberal ideas. Whether immigrants live and work among ethnically diverse communities and how much they interact with diverse groups might explain their social trust levels (Kokkonen et al. 2014). I want to test all these four theories in case of Uzbek migrants in the US. My expectation from this project is that it not people, but institutions in their counties matter for their views on democracy.

2. The study you propose involves learning about a theory by observing certain of its implications. What one or two hypotheses that arise from the theory are you planning to assess? Why or how does the theory justify your expectations about these hypotheses?

Green card winners who have immigrated from a non-democratic country such as Uzbekistan are exposed to democracy in the US as they engage with democratic institutions, norms, and culture in their daily life. I expect this increases their support for democracy since they experience direct benefits of democracy. Their conceptualization of democracy among migrants should also be consistent with the general understanding of liberal democracy whereas people in Uzbekistan are expected to have simplistic understanding of democracy. More formally, my two hypotheses are as following:

Hypothesis 1 : Green card lottery winners from Uzbekistan have higher support for

democracy than unsuccessful green card lottery participants in Uzbekistan Hypothesis 2 : Green card lottery winners from Uzbekistan have more complex understanding of democracy than unsuccessful green card lottery participants

Coming from a relatively homogeneous society like Uzbekistan, green card winners interact with diverse groups of people than do their fellows residing in Uzbekistan. Interaction with people in heterogeneous communities should increase their social trust and respect for people from different religions, ethnicity, and language. There is one hypothesis regarding social trust levels of green card participants. Hypothesis 3 : Green card lottery winners from Uzbekistan have higher level of trust for people of other nationality and religion than unsuccessful green card participants in Uzbekistan.

3. What data and research design will help you answer this question? Why are you making these choices?

In this project, I am conducting an original survey to gather information about political attitudes and social trust levels of green card lottery participants. The primary respondents in my survey are those who participated green card lottery in Uzbekistan. If they participated outside Uzbekistan or are already green holders through channels other the lottery, they are not included in my data. In my survey, I ask respondents questions about their views on different political regimes (1), characteristics of democracy (2), suitability of democracy to Uzbekistan and the role of Islam (3), and social trust (4). The primary respondents in my survey are those who participated in green card lottery in Uzbekistan, both successful and unsuccessful participants.

To examine the impact of migration, I will compare green card holders and unsuccessful lottery participants on their political attitudes and social trust. As the opportunity to migrate is randomly assigned, this serves as a natural experiment to compare those who migrated as a result of the lottery selection and those who stayed in the country because the lottery did not select them. Randomization of the lottery makes this observational setting approximate to true experiments. This obviates confounding factors and provides credible evidence for causal effects as Dunning (2012) suggests. By design, this project is similar to Angrist's (1990) paper on the effects of military induction as a result of Vietnam-era draft lottery on later earnings). I use a lottery instrument for subsequent migration similarly as Agrist uses it for military draft, and this design is different from another Vietnam lottery paper by Erikson and Stoker (2012) who studied direct effects of the lottery on political attitudes.

4. What are the advantages and disadvantages of this research design to addressing your question?

The advantage of this research design is that the green card lottery provides randomiza-

tion. This allows me to examine the impact of migration by comparing average differences in attitudes of those who migrated and those who stayed as a result of a lottery draw. In other words, the advantage of this randomization is that there is a mechanical way to infer the causal effect (Rosenbaum 2017). Without randomization, I would have to make certain assumptions about whether unobservable factors are affecting both the decision to migrate and political attitudes. Without randomization, I might be comparing those who choose to come to the US because of their preference for democracy with non-immigrants whose primary motivation to migrate would be to live by the ocean. In this regard, higher support for democracy among migrants in comparison to their fellow non-migrants would not be a result of a migration, but due to differing pre-existing political views. Randomization solves this issue by randomly allocating an opportunity to migrate to those who want to migrate to the US. As long as migration is through random opportunity to migrate, I can be confident in my estimation of migration effects on political attitudes and trust.

Although the lottery system is very advantageous to my research design, it creates some problems. Because the lottery randomizes participants into two groups of successful and unsuccessful applicants, the lottery might have short-term effects on the participants. Feeling unhappy about the lottery results, unsuccessful participants may underreport their support for democracy. Similarly, lottery winners may overreport their support for democracy due to their excitement. Additionally, there might be other disadvantages. For instance, the green card lottery provides an opportunity to migrate but does not guarantee that those participants who are selected will migrate to the US. Selected applicants have to go through a visa and interview process and there are chances that they might fail. Rejected applicants may form negative feelings about the US bureaucracy and translate these feelings into their attitudes towards democracy. I expect these effects to be short-term and I will take some approaches to tackle, which I will explain below.

5. Describe your measures and any indices you construct

Recalling description of my survey in Part 3, I have four groups of questions. The first group includes 3 "feeling thermometer questions" about political systems. Because a military rule is not conceptually an opposite of democracy, I should not assume that positive feelings towards a military rule directly translates into negative feelings towards democracy or vice versa. Therefore, each political system receives a separate index. The second group includes multiple-choice questions about main characteristics of democracy. This group takes a single index ranging from 1 to 4, with each score representing correct marking by a respondent. Group 3 is about Uzbekistan; questions differ, so each question receives a separate index. There are 4 questions about social trust in the group 4. Two questions about general trust receive one index and two questions about trust in people of other nationality and religion

together receive one index.

6. Use data to make the case that your research design allows you to interpret observed quantities (like observed data comparisons or parameters of models to data) as theoretically relevant and clear

6.3 Although randomization clears the problem of confounding factors, using the green card lottery as a proxy randomization for migration might bring some problems. The Direct effects of the lottery might drive the estimates of migration effect in one direction or another. Although I expect such effects to be short-term, I cannot rule out these effects. I make several comparisons to address this problem. First, I will compare recent green card winners who have not yet travelled to the US with unsuccessful lottery participants. Due to randomization, these two groups should be similar in their political attitudes unless there is "lottery effect". The difference in political attitudes of these groups does not identify whether the lottery effect is due to winning the lottery or failing to win it or both. To examine it, I will compare green card holders in the US with recent lottery winners (1). I will also compare recent unsuccessful lottery participants with those unsuccessful applicants from the previous year or earlier, but not this year (2). I will look at the difference between these two comparisons. If the lottery effect for recent winners is higher, then I might be overestimating the migration effect. If the lottery effect for unsuccessful applicants is higher, it means my estimates are understated.

6.4 The computer algorithm for the green card lottery administers randomization. If is truly random, successful and unsuccessful participants should be similar in demographic characteristics. If they are not similar, then the assignment of an opportunity to migrate is not random. To analyze it, I will create a table that shows the association between success in the lottery and demographic characteristics. I will use chi-squared tests to examine the relationship. The null hypothesis states that there is no relationship. I will use p-values to make judgements. P-values express the probability that the observed results are under the null-hypothesis. When the p-value is less than widely accepted value of 0.05, I reject the null hypothesis, meaning there might be association between demographic characteristics and lottery success. If the p-value is bigger than 0.05, I fail to reject the hypothesis, suggesting that there might be association between these variables. In this regard, it is important to check the association of demographics with migration and success in the lottery. If lottery winners are systematically being rejected visas by demographic factors, I may not be able to assume that migrants and non-migrants(including rejected winners) are comparable. I will use fake data to calculate chi-squared tests.

7. Explain your plans for any missing data or extreme outcome or covariate values you may encounter when you get the real data (or perhaps you have the

background data but not the real outcomes, so you can explain your plans for such data issues in that case here too).

My survey sets a value range for responses so I will not have extreme outcomes. The survey, however, allows respondents to skip questions so there might be missing data. It could be in the outcome variables or assignment variables. Either way, a lot of missing values might be problematic. From my experience of conducting a similar survey, I expect I will not have systematic missing data. I will run R commands to check for missing values. In case a lot of entries are missing, I will use Amelia package in R to impute multiple entries for missing values. This package uses a bootstrapping algorithm to impute entries without changing the relationship between observed data. I will create five datasets and use each of them for my analysis. I will then combine results.

8. What statistical tests do you plan to use? Explain why you chose these tests and any decision making criteria you will use upon seeing the results of the tests. You should also engage with the problem of multiple testing here if you are going to show the results of more than one test. (Recall that confidence intervals and hypothesis tests convey more or less the same information. So a confidence interval is a form of testing.)

Recalling earlier discussion, green card lottery success does not guarantee migration to the US. It is unclear why the US Consular service might reject visas for some GC lottery winners. If visa decisions are based on factors that might be associated with political attitudes of lottery winners, treatment group of migrants is not due to a true randomization. There is no data on how many lottery winners are rejected visas each year. Most sources cite administrative reasons for visa rejections. The number might be so small that my survey data may even not capture it. If the proportion of rejected winners make up around 5-10 % of total winners in my dataset, I will use the lottery success as an instrumental variable to estimate the effect of migration. At 5-10% visa rejection rate, the lottery success has a strong first-stage. Randomization of the lottery makes it an exogenous variable, satisfying the exclusion restriction condition. The estimates produce local average treatment effects (LATE). In case of only a few rejected winners in my dataset, I will use differences in means as my t-statistic to compare Uzbek green card holders in the US with unsuccessful lottery participants. In this case, my treatment group is Uzbek immigrants who migrated to the US as a result of winning green card and the control group is unsuccessful lottery participants. To compare test statistics, I will use power. Note that I am not employing multiple tests. I am using either instrumental variable or differences in means. If I were to use both of them, I might have to choose the one with higher power value when in fact two tests are capturing different relationship and the other test might be better despite its lower power value.

9. Explain how you will judge the performance of those tests. Will you only use simple false positive rate and power? Or do you need to add family-wise error rate? false discovery rate? Or something else? Explain why you made this choice.

The power of a hypothesis test is the probability of a "true positive". In other words, It is the probability that the test correctly rejects the null hypothesis when the alternative hypothesis is true. It is the probability of avoiding a "false negative", i.e. type II error. I use the power and false positive rate to judge the performance of my tests. Because I am comparing differences in means between the treatment and control group, power and false positive rate help test the null hypothesis of no difference in means scores. Since I am not using multiple tests, I do not need to add family-wise error rate. I use *lm* and *lmrobust* commands to estimate the differences in means.

10. Show and explain how your test performs in regards those properties (at least you will show false positive rate and power

I need to compare how my tests perform when different parameters and models are assigned. I will use DeclareDesign for these purposes. In these designs, I am examining the effects of migration on how people view democracy in the feeling thermometer. I will explain steps I have taken to construct simulated models and how I tested them.

I declared my population to be the sample. There are 121 respondents in my treatment group. For potential outcomes, I put 0.25 conditionally in the expectation that there is going to be 25 percent difference between two groups in their feelings about democracy. I use the mean of the differences between groups—average treatment effect as my estimand. For my models, I used differences in means and lm. I want to see how these modeling functions differ. I simulated the data 1000 times. The power came out to be 0.25 and 0.25 for two tests. Because power is not close to 1, so the hypothesis test is not very good at detecting a false null hypothesis. This is probably because my data is fake, and randomization is not truly random.

11. What statistical estimators do you plan to use? Explain why you chose these estimators. Especially explain what is your target of estimation | what is the estimand?

Because I am comparing two groups on political attitudes and social trust, the differences between these groups are my estimands. To estimate them, I use average treatment effect as my estimator. In cases where the compliance to lottery success is not perfect, I will use local average treatment effect as my estimator. This allows me to calculate the estimand for subgroup of lottery winners who migrated to the US. When there is no effect of migration on political attitudes and social trust, the difference between these two groups is close to zero.

12. Explain how you will judge the performance of those estimators (especially bias and MSE)?

Because my data is fake and the assignment is not truly random, there is high level of bias in my estimates. However, I expect the bias to be 0 when true randomization is properly done through the lottery. In DeclareDesign, my estimand is differences in means. I am using lm and differences in means as my models. Bias and RMSE are provided by DeclareDesing. Bias tell me how my results are off from the true results. In random experiments, bias should be close to zero. RMSE is the variation in different results simulations produce. We can think of it as distance between predicted values and observed values. RMSE can capture noise in the estimates and if the noise is small, then it means the model is good estimating the effect of the treatment.

13. Show and explain how your estimator performs in regards those properties (at least bias and MSE).

I used DeclareDesign to simulate the data using the same strategy before. I set my estimand as differences in means and my estimator is ATE. I used lm and differences in means as my models. The biases are high for both models and RMSE square is also high at 0.22 for lm model, though it is slightly low for differences in means. The nature of the data is such that there is high level of bias in the estimates. With random and more accurate data, bias should be smaller or close to 0. RMSE should also small given that the values in my data are going to small. High RMSE would be problematic as we want the errors to be small. Since I am interested in the difference between the outcome and treatment groups rather than precise estimates of these models, some errors are acceptable.

14. Make one mock gure or table of the kind you plan to make when you use the actual outcome. Interpret the results of the mock analysis as if it were the real analysis. Saying something like, "If the real outcome were as I have simulated it, then the following table/figure would mean such and so about the theory."

Table 1: Views on political systems

	democracy	dictatorships	military	expert rule
t-test	–1.88	1.24	0.45	-1.70
p-value	0.03	0.89	0.67	0.04
df	270.15	223.42	182.07	217.21

My results on the DeclareDesign did not produce meaning figures, So I decided to present some findings from t-test. They demonstrate the difference on views between migrants and non-migrants. Note that because of data parameters, the differences in means are not as theories predict. The differences in means for views on democracy and rule by the experts seem to have statistically significant differences at 0.05. if these were the results in the actual findings, I would be very surprised because the differences in means are negative. I expect them to be positive.

The link for my Github repository click

Code Appendix

```
data4 <- read.csv("dem_survey1.csv")
names(data4)[1:20] <- c("age", "education", "gender", "lottery", "migrant", "dictator",

# Chi-square tests
data4$age <- as.numeric(data4$age)
age_treat <- table(data4$migrant, data4$age)
chisq.test(age_treat)     # p-value is 0.43

gender_treat <- table(data4$migrant, data4$gender)
chisq.test(gender_treat)  # p-value is 0.07

edu_treat <- table(data4$migrant, data4$education)
chisq.test(edu_treat)  # p-value is 0.08

age_success <- table(data4$lottery, data4$age)
chisq.test(age_success)  # p-value 040

gender_success <- table(data4$lottery, data4$gender)
chisq.test(gender_success)    # p-value 0.11

edu_success <- table(data4$lottery, data4$edu)
chisq.test(edu_success)   # 0.03

# given my data is fake and treatment is not random, there are some associations

# Declare desing
```

```
iv1 <- ivreg(dem_system ~ migrant | lottery , data = data4)
summary(iv1)

m1 <- lm(dem_system ~ migrant, data = data4)
summary(m1)

simple_random_assignment_step <-
  declare_assignment(Z = simple_ra(N = N, prob = 0.6),
                     legacy = FALSE)
simple_random_assignment_step(data4)

study_diagnosands <- declare_diagnosands(
  bias = mean(estimate - estimand),
  rmse = sqrt(mean((estimate - estimand)^2)),
  power = mean(p.value <= 0.05)
)

m_population <- declare_population(data = data4)
m_outcomes <- declare_potential_outcomes(Y_Z_0 = as.numeric(dem_system), Y_Z_1 = Y_Z_0 +
m_inquiry <- declare_estimand(PATE = mean(Y_Z_1 - Y_Z_0))
m_assignment <- declare_assignment(m = 121, label = "Z")
m_reveal <- declare_reveal(Y,Z)
m_est1 <- declare_estimator(Y ~ Z, model = difference_in_means, label = "PATE")
m_est2 <- declare_estimator(Y ~ Z, model = lm, label = "lm")

test_design1 <- (m_population + m_outcomes + m_assignment + m_inquiry + m_reveal + m_est
sim_data1 <- draw_data(test_design1)
m_est1(sim_data1)
m_est2(sim_data1)
set.seed(34444)
simulate_design(test_design1)
des_diagnose1 <- diagnose_design(test_design1, bootstrap_sims = 0, sims = 1000)
des_diagnose1

with(sim_data1,sd(Y[Z==1], na.rm = TRUE))
with(sim_data1,sd(Y[Z==0], na.rm = TRUE))
```

```
with(data4,sd(dem_system[migrant==1], na.rm = TRUE))
with(data4,sd(dem_system[migrant==0], na.rm = TRUE))

study_diagnosands <- declare_diagnosands(
  bias = mean(estimate - estimand),
  rmse = sqrt(mean((estimate - estimand)^2)),
  power = mean(p.value <= 0.05)
)

t.test(dem_system ~ migrant, data = data4, paired = FALSE,
        alternative = "less")
t.test(dictator ~ migrant, data = data4, paired = FALSE,
        alternative = "less")
k2 <- t.test(expert ~ migrant, data = data4, paired = FALSE,
        alternative = "less")
k1 <- t.test(military ~ migrant, data = data4, paired = FALSE,
        alternative = "less")
```

1. Bibliography

Angrist, J. D. (1990). Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records. The American Economic Review, 313-336.

Bilodeau, A., McAllister, I., Kanji, M. (2010). Adaptation to democracy among immigrants in Australia. International Political Science Review, 31(2),141-165.

Canache, D. (2012). Citizens' conceptualizations of democracy: Structural complexity, substantive content, and political significance. Comparative Political Studies, 45(9), 1132-1158.

Diamond, L. (1994). Rethinking civil society: Toward democratic consolidation. Journal of democracy, 5(3), 4-17.

Dunning, T. (2012). Natural experiments in the social sciences: a design-based approach. Cambridge University Press.

Erikson, R. S., Stoker, L. (2011). Caught in the draft: The effects of Vietnam draft lottery status on political attitudes. American Political Science Review, 221-237.

Kokkonen, A., Esaiasson, P., Gilljam, M. (2014). Migration-based ethnic diversity and social trust: A multilevel analysis of how country, neighbourhood and workplace diversity affects social trust in 22 countries. Scandinavian Political Studies, 37(3), 263-300.

Rosenbaum, P. (2018). Observation and experiment. Harvard University Press.