Supplementary Documentation: Information, Intermediaries, and International Migration*

Samuel Bazzi
UC San Diego, NBER, and CEPR

Lisa Cameron
Univ. of Melbourne

Simone Schaner
Univ. of Southern California, NBER

Firman Witoelar
Australian National University

May 12, 2022

1 Overview

This document provides additional detail on the various data sets we use and how we constructed outcomes variables. We also provide additional detail on how we trained random forests to predict migrant experience and on how we constructed measures of agency quality.

2 Data Sets and Variable Construction

2.1 Original Survey Data

Our sample consists of 400 of the largest female-migrant-sending villages in eight of the large female-migrant-sending districts on the island of Java. We dropped very large villages with a population density of more than 4,000 persons per square kilometer or a total population of more than 15,000 to limit logistical complexity during intervention implementation.

We collected four rounds of survey data, beginning with a baseline survey, which took place between April and June 2015. The "tracking sample" (4,805 women interested in migrating in the future; half with prior migration experience and half without) took a long-form survey, which collected information on personal characteristics, including education, cognitive ability, risk attitudes, beliefs about migration, and (for former migrants) details of the woman's most recent migration experience. An additional sample of 5,607 women who had migrated before received a short survey, which focused on past migration experience. We use baseline data for the the following purposes:

*Emails: Bazzi (sbazzi@ucsd.edu), Cameron (lisa.cameron@unimelb.edu.au), Schaner (schaner@usc.edu), Witoelar (FirmanWitoelar.Kartaadipoetra@anu.edu.au).

- Constructing agency quality rankings, which populate the report cards
- Calculating average migration outcomes by agency quality quintile, which informs the infographic
- Verifying randomization balance
- Constructing measures of predicted migration experience

We targeted the tracking sample for three more survey rounds. A first short-run follow up (midline 1) took place between August and December 2016. The second follow-up (midline 2) spanned October to December 2017, and the final endline survey took place between May and October of 2019. During the endline we also interviewed an additional, 2,418 women who were not part of the tracking sample but had migrated during the post-intervention period. The midlines were very short, collecting data on basic migration outcomes (e.g. whether migrated, what destination, which agency), plans for future migrations, and beliefs about migration outcomes. The endline was longer, collecting additional detail on the migration process, costs and payments associated with migration, and women's experiences at the agency and abroad. The endline also collected data on beliefs about migration outcomes, current economic activity, and household income and assets. Since we observe differential direct interview rates during both midlines but not the endline, we use the endline to construct primary outcomes whenever possible.

The remainder of this section provides additional detail on how key variables used in the analysis were constructed. Unless otherwise noted, we do not use informant responses when constructing variables.

- **Exposure to interventions.** Community meeting attendance is based on women's self-reports of whether they attended a community meeting about migration in the village in 2015. We did not specify the content of the meeting beyond this. For women interviewed in person, enumerators displayed samples of the infographic, report card, and comic and asked women if they recognized them. Women interviewed over the phone were asked to report exposure based on a verbal description of the intervention material.
- **Migration.** *Migrated post intervention* is based on a woman's report of whether she migrated abroad in October 2015 or later. *Migrated with agency* indicates a woman migrated abroad in October 2015 or later *and* used an agency to migrate. *Migrated without agency* indicates a woman migrated abroad in October 2015 or later and *did not* use an agency to migrate. We use informant reports whenever a woman was not directly interviewed.
- **Migration destination and occupation.** We record the destination country for each migration, and classify countries according to whether they are in Asia, the Middle East and North Africa (MENA), or other parts of the world. To assess occupation, we ask migrants about the type of work they perform, allowing them to specify multiple options. Formal sector work includes non-domestic helper and non-caregiver jobs typically done outside the home for a wage (factory worker, cruise ship attendant, cashier, etc.).
- **Migration plans.** To capture general interest in migration, we ask women whether they plan to migrate (again) in the future. To measure firmness of plans we also construct a dummy variable

indicating that the woman plans to migrate within the next year. We also asked women if they had completed the following pre-migration tasks: secured family permission, got permission from the village head, chosen a sponsor, and chosen an agency.

- **Beliefs.** In order to elicit beliefs about migration outcomes, we first prompted the respondent with a migration outcome, such as "gets at least one day off per week". We then asked the respondent to report how many female migrants in their village (out of 10) experience this outcome, using – when interviewing in person – a card with a line of 10 women to assist with visualization. For a subset of scenarios, we also asked the respondent to report her likelihood of experiencing the same, again on a 0 to 10 scale. When constructing endline beliefs indices we limit attention to beliefs questions included in both the phone and in person survey instrument. We follow Kling et al. (2007) when constructing standardized indices, first imputing missing index components with the treatment group mean, then standardizing each component relative to the control group mean and standard deviation, and finally averaging all components into an overall index. The *agency index* includes questions about the following: will the agency take identity documents*, will the agency give information on migrant worker rights, will the agency provide clean food and water, will the agency staff treat migrant workers with respect, will the agency staff let the migrant workers leave the dorms/training center, will the agency follow legal procedures, will the agency give accurate information. The *job quality index* includes information about the following: will the migrant have to work more than 12 hours per day, will the job match the contract*, will the migrant get a day off, will the migrant's salary be retained*, will the migrant be paid less than their contract, will the employer hold the migrant's identity documents*, will the employer provide presents, will the migrant have to return early, will the migrant experience physical abuse. The *infographic index* includes the first two components of the agency index and the first 6 components of the job quality index. We also generate baseline analogs of these indices to use in some analyses – variables starred in the preceding text were not available at baseline and therefore excluded from these indices. We also generate a baseline sponsor beliefs index, which includes the following components: will the sponsor give good pocket money, will the sponsor refrain from extortion, will the sponsor take the migrant to a good agency. Index components are coded so that a higher value indicates a better outcome. The baseline analogs of the indices exclude beliefs about the agency/employer taking identity documents, the job matching the contract, and salary retention because these questions were not included in the baseline survey.

- **Migration experience.** Our analysis focuses on the subset of experience measures collected as part of both the phone and in-person survey instruments. We follow Kling et al. (2007) when constructing standardized indices, first imputing missing index components with the treatment group mean, then standardizing each component relative to the control group mean and standard deviation, and finally averaging all components into an overall index. The *pre-departure preparation index* includes: use of an agency, whether the agency provided training, time spent on training, the share of government-mandated training topics covered by the agency (equipment/tools required for job, job skills, destination information, how to remit money, migration insurance policy, how to behave on the job, destination country culture, how to get help when abroad, the repatriation process, migrant worker rights, the migration contract), the migrant's subjective grade (0-10) of the agency training, whether the migrant signed a contract (in Indonesian, that she understood) while at the agency, whether the agency allowed the migrant to leave the training facility and residence, whether the agency held the migrant's identity documents, whether the agency followed legal procedures (per the migrant's assessment), and the migrant's subjective overall rating of the agency on a 0-10 scale. All outcomes in this index are coded to zero if the woman did not use an agency to migrate. The *job quality index* includes: whether the migrant was given a weekly day off, the job matched the contract, the employer allowed the migrant to retain her identity documents, the migrant had her own private living quarters, the migrant received proof of payment, the migrant was allowed to leave the employer's residence, and the migrant's overall subjective rating of the migration experience. The *pay index* includes: total wages net of salary deductions, total earnings (wages plus other income from the agency, sponsor, and employer) net of costs (salary deductions plus other migration costs paid to the agency, sponsor, employer, or other entities), whether the migrant received the full contracted salary, whether the migrant received salary payments on time, and whether the migrant received additional pay for overtime work.

- **Agency choices.** We use migrants' reports of their placement agency's name to construct agency quality classifications. During each survey round we first mapped the migrant reports to the list of placement agencies sanctioned by the Indonesian government at the time of the survey. (For post-baseline survey rounds we updated our original baseline list, keeping the names of any agencies that lost certification over the ensuing years.) If a woman did not know the name of her agency, or the agency was not on the list of sanctioned agencies, we classify this migration as one with an "unknown agency". Migrations with sanctioned agencies that were not listed on the report card are classified as "ungraded agency" migrations. Finally, we split migrations with graded agencies into "high grade" (top third of grade distribution among realized endline migrations), "average grade" (middle third), and "low-grade" (bottom third) migrations.

- **Household income and expenditure.** The endline survey asked either the tracking sample woman (in the case of in-person interviews) or the informant (when the focal woman was either unavailable or interviewed over the phone) to report on household economic wellbeing. To calculate *household income*, we aggregate reports of wage earnings and business profits earned by residents in the past month, average monthly remittances from non-resident household members, and average monthly agricultural and "other" income (including rents, social protection payments etc.). Agricultural income is calculated as the market value of output (consumed and sold) less the cost of inputs, including labor. We value family labor at market rates when calculating costs. Total expenditures reflect the respondent's report on all spending in the past month. We separately ask respondents to report food expenditure in the past month.

- **Household assets and dependence on social protection.** The endline survey asked either the tracking sample woman (in the case of in-person interviews) or the informant (when the focal woman was either unavailable or interviewed over the phone) to provide details on household assets and receipt of social protection. Enumerators also recorded the quality of the home based on observation. The *housing quality index* is constructed using factor analysis. We first create a series of dummy variables identifying the home's type of roof, wall, floor, source of drinking water, and toilet facility. The first factor assigned opposite-signed weights to higher vs lower quality construction (e.g. a tile versus a palm roof), so we extract that factor, rotate it so higher values indicate better outcomes, and standardize it relative to the control group mean and standard deviation. To construct the *asset index* we run factor analysis on a series of dummy variables indicating ownership of a bicycle, motorcycle, boat, television, air conditioner, heater, gas stove, refrigerator, motor boat, car, house, and land. We extract the first factor, which weights assets positively, with larger weights on "higher status" assets like a motorcycle, television, and house, and standardize it relative to the control group. Finally, we create a *dependence on social protection index*, which uses factor analysis to aggregate dummy variables for receipt of seven of Indonesia's biggest social assistance programs/social assistance cards (Program Keluarga Sejahtera, Bantuan Siswa Miskin, Kartu Keluarga Sejahtera, Kartu Indonesia Pintar, Kartu Indonesia Sehat, Jaminan Kesehatan Nasional, and Raskin/Bantuan Pangan Non Tunai) – again we extract the first factor, rotate it so higher values correspond to better outcomes, and standardize it relative to the control mean.

- **Occupational status.** We asked all tracking sample women (or informants, in cases where a direct interview was not possible) to report on the woman's current occupation, regardless of migration status. Women currently abroad are classified as "on migration" regardless of what they do abroad. Women in Indonesia as classified as either unemployed (this includes women who are not actively seeking work), an unpaid family worker, a casual worker (low skill, irregular work often paid in daily or weekly installments), a wage worker, or self employed.

- **Monthly earnings.** Women in Indonesia were asked to report their total earnings in the past month. Non-workers are coded as earning zero income. For women on migration, we calculate total earnings less deductions to date, divided by the number of months abroad.

2.2 Administrative Data

We use administrative data on migrant departures and returns in our analysis. Here we describe the different datasets we received and how they were processed to facilitate analysis.

- **Placements Data: 2011-2013.** Before starting the project, we obtained administrative placements data for all of 2011, 2012, and part of 2013, which we used to create a list of sample villages and create some of our randomization strata. These data include information on the migrant's gender, agency, destination, home district, and address, the latter of which was written in a non-standard format. We limited the sample to our eight study districts and extracted sub-district and village names from the full addresses using a combination of code and hand inspection. We were able to match 94 percent of 2012-2013 female placements in our study districts to village codes. We use these data to calculate the number of migrants per village, as well as the Herfindahl index based on 2012 and 2013 placements.

- **Repatriation Data: 2010-2013.** We received data on migrant returns processed by a dedicated terminal for return migrants at Soekarno-Hatta airport in Jakarta between 2010-2013. While not all return migrants passed through this (now defunct) terminal, a large number did, with numbers processed ranging from 135,289 in 2013 to 357,854 in 2010. These data include information on the migrant's gender, home address, country of work, and agency. The data also record the reason why the migrant returned, including end of contract, visit to Indonesia, or "troubles", a catch-all term meant to identify migrants returning early due to issues abroad. We combine agency-specific counts of troubled workers in 2013 with the total number of agency departures in 2011 (since most contracts are two years long) to construct a proxy of agency quality, which we used to validate our own agency grades.

- **Placements Data: 2015-2019.** After the end of field activities we received a final transfer of placements data spanning 2015-2019. These data include information on the month of departure, destination, placement agency, migrant gender, and migrant address. The government coded province, district, sub-district, and village names in these data and did not provide string addresses. The percent of female placements in our eight study districts with a valid village name is 50.4, 68.6, 83.8, 91,8, and 91.3 percent in 2015, 2016, 2017, 2018, and 2019 respectively. Overall, 83 percent of records in the post-intervention period have a valid village name. We use these data to construct measures of total departures by village, departures by destination, and to construct measures of agency market share.

2.3 Secondary Data

In order to calculate the number of female migrants per capita, which we used to stratify our randomization, we require village-level population estimates. We use the Village Potential (*Podes*) triennial administrative census from 2011, available from the Indonesian Central Statistical Agency (*Badan Pusat Statistik*, BPS) . This survey provides information on key characteristics of all Indonesian villages typically reported by the village head.

2.4 Random Forest Predictions of Migration Experience

We train a random forest to predict migration experience associated with migrations recorded at baseline. To do this, we first construct baseline analogues of the pre-departure preparation, job quality, and pay indices. The components of the indices are the same, except we eliminate a dummy variable indicating that the migrant used an agency from the pre-departure preparation index, since – by design – all migrants interviewed at baseline had used an agency. To simplify the analysis we then average the three indices to create a single overall measure of experience.

We train the algorithm on tracking sample women with past migration experience at baseline. When training the forest we enable honest splitting, meaning the sub-samples used to determine a tree's splits differs from that used to populate the leaf nodes (Wager and Athey, 2018); all available parameters (the fraction of the sample used to build each tree, the number of variables tried for each split, the minimum number of observations in each tree leaf, the fraction of data used for determining splits, whether to prune estimation sample trees so no leaves are empty, the maximum imbalance of a split, and the imbalance penalty) are tuned via cross validation.

When preparing variables for the forest, we convert all categorical variables into dummy variables, with missing values assigned their own category. We construct separate dummy variables to identify missing values of continuous variables, and then recode missing continuous variables to the mean. The following list details all variables included in the forest.

- Age
- Marital status
- Education
- Randomization strata
- Ethnic group
- Can read
- Can write
- Dummies identifying preferred gamble in an incentivized risk task
- Discount factor implied by a series of hypothetical monetary choices
- Beliefs for self, and others in village (out of 10): will receive good pocket money from sponsor, sponsor will take person to a good agency, sponsor will not extort, agency will provide clean food and water, agency staff will treat migrant with respect, migrant will be allowed outside agency, agency will follow legal procedures, agency will provide accurate information, agency will provide information about migrant worker rights, migrant will get weekly day off, migrant will receive presents, migrant will not be paid less than contract, migrant will not have to work more than 12 hours/day, migrant will not return early, migrant will not suffer physical abuse
- Expected number days off: more than once a month, approximately once a month, at least twice a year, approximately once a year, never
- Expected salary net of fees
- Expected hours of work per day
- Year (as of baseline) hopes to migrate
- Plans to migrate to Asia
- Plans to migrate to MENA
- Plans to skip sponsor, go directly to agency
- Dummies for planned occupation abroad (elder care, babysitter, nurse, domestic worker, driver, store/restaurant/hotel staff, agricultural worker, mining worker, construction worker, factory worker, cruise ship crew, other)
- Dummies for who received advice on sponsors from (family, friends/neighbors, village head, agency, another sponsor, other)
- Dummies for who received advice on agencies from (family, friends/neighbors, village head, agency, another sponsor, other)
- Has chosen a sponsor
- Number of sponsors known from village
- Number of sponsors known from outside village
- Number of sponsors talked to from village
- Number of sponsors talked to from outside village

- Dummies identifying woman's reported qualities of a good, bad sponsor
- Dummies identifying woman's reported qualities of a good, bad agency
- Trust questions: most people can be trusted/need to be careful; would a lost purse with IDR 200,000 and an ID card be returned by someone outside your village: very likely, somewhat likely, somewhat unlikely, very unlikely. For the Heckman decomposition we generate a dummy variable identifying individuals who report returning the card is somewhat or very unlikely.
- Mental health inventory, dummies identifying often/sometimes/never to the following scenarios in the past 4 weeks: had trouble sleeping, been bothered by things that don't usually bother, felt lonely, experienced sadness, experienced anxiety or fear, had difficulty concentrating, normal tasks felt like an effort, had difficulty remembering/recalling something. for the Heckman decomposition we create an aggregate score, summing across scenarios, where "often" is coded to 2, "sometimes" to 1, and "never" to zero.
- Fraction correct: Raven's Matrix questions
- Fraction correct: Math questions
- Locus of control score
- Big 5 scores: extraversion, agreeableness, conscientiousness, neuroticism, openness

Figure 2 demonstrates that predicted experience based on baseline characteristics is robustly positively correlated with actual experience on subsequent migrations recorded at endline.

3 Constructing Measures of Agency Quality

3.1 Indicators of Quality of Migration Experience

An inherent challenge to measuring agency quality is that there are many potential indicators spanning both inputs (e.g. time and quality of pre-departure training, effort made to ensure the migrant is well informed of her contract and rights, quality of partner agency in the destination country) and outputs (compensation, working conditions, employer quality). We estimated quality by first creating an aggregate "migration experience" measure, which combines measures across the following domains:

- **Experience with the Indonesian Placement Agency:** This domain includes indicators for whether the woman received at least 10 days/2 weeks training (the lowest legal minimum at the time of the study), the number (out of 11) of key topics a migrant was trained on, an indicator for whether the woman signed a contract while at the PT, and an indicator for whether the woman's job was in accordance with her contract.

- **Experience with the Destination Placement Agency:** This domain included indicators for whether a woman worked with an agency in the destination country, whether the agency picked her up at the airport (versus the employer), whether the agency allowed the migrant to retain her own personal documents, whether the agency gave the woman information on migrant worker rights and whether the agency gave the migrant information on how to seek help in the destination country.

- **Compensation:** This domain consists of average monthly pay (both regular salary and bonuses) net of migration costs.

- **Experience with the Employer:** This domain included indicators for whether or not the migrant experienced salary cuts, received at least one day off per week, was not required to work more than 12 hours per day, was paid for overtime, was given proof of payment, did not sustain injuries on the job, was allowed to contact her family, was allowed to pray, was not paid late, experienced retained salary payments, was fired, was allowed to retain her personal documents, experienced physical abuse, experienced sexual abuse, experienced verbal abuse and experienced other forms of abuse. All the indicators were constructed so that a value of 1 signaled a positive experience (e.g. no salary cuts, allowed to pray) and 0 signaled a negative experience.

We created an aggregate index by standardizing each of the above index components, calculating the average within each domain, and then taking a simple average across the four domains. We chose these inputs for two reasons. First, qualitative research with former and potential migrants and interviews with NGO and government stakeholders suggested that these inputs are important determinants of migrants' experiences. Second, we found that this combination of inputs was particularly successful at identifying high-performing agencies in our out-of-sample validation tests.

3.2 Estimating Agency Quality

We now use our index of migration experience quality to obtain estimates of agency quality. Migrant experience is assumed to be determined by three main factors: First is the migrant's qualifications and skill as a worker – for example, more qualified migrants will likely be paid higher salaries and secure jobs with better amenities. Second is the input of her agency – the agency can improve a woman's experience in several ways, e.g. by providing comprehensive training, providing information on migrant workers' rights and by partnering with more reputable counterpart agencies in the destination country. Third are all other factors – such as whether a migrant is "lucky" and gets a kind, considerate employer. We can represent this in equation form:

$$experience_{ia} = \mathbf{qual}'_{ia}\boldsymbol{\eta} + \gamma_a + \varepsilon_{ia} \tag{1}$$

where $experience_{ia}$ is the experience of migrant i migrating with agency a, $\mathbf{qual}_{ia}$ is a vector of characteristics capturing migrant quality/experience, γ_a is an agency fixed effect, and ε_{ia} is an error term capturing idiosyncratic factors. Given this structure, the agency effect will also capture systematic variation across agencies in terms of job characteristics including destination country and job type. We decided to include this in the agency effect because, conditional on migrant qualifications, this can be an important driver of migrant welfare.

We employ empirical Bayes techniques, commonly used in the economics literature measuring teacher quality (Chetty et al., 2014; Herrmann et al., 2016; Kane and Staiger, 2008; McCaffrey et al., 2004), to obtain our measure of agency quality. A key advantage of this technique (as compared to estimating agency fixed effects via OLS) is that it "shrinks" agency effects towards the mean in a way that is proportional to sample size. Thus, when fewer migrant ratings are available, there is more mean reversion.

To operationalize this, we first use OLS to residualize out measures of migrant quality, including dummy variables for highest education, age, age squared, migration year dummies, and district of residence dummies. We then construct empirical Bayes estimates for the agency effects and discard estimates for agencies with fewer than 30 migrants in our survey. We re-scale the remaining estimates to run from a minimum of 50 to a maximum of 95, in order to mimic the distribution of grades in the Indonesian school system.

After constructing the ratings, we ran out-of-sample validation tests to ensure that our measures of quality could predict the experiences of migrants outside our estimation sample. Here, we correlated our rankings with rates of problems as recorded in government repatriation records, as well as with the International Organization for Migration's (IOM's) records on verified victims of trafficking. We drop all women in the government data who live in our study districts to ensure that there is no overlap between the women in the government repatriation records and our survey data. Figure 3 shows the average rate of problems in government data, per 2013 arrivals records, and the rate of victims of trafficking for three categories of agencies: those with grades less than 65 (roughly the bottom third of agencies), those with grades 65-85 (roughly the middle third of agencies) and those with grades above 85 (roughly the top third of agencies). Note that there is no bar in Panel B for the "grade above 85" group, because no agencies in this group had recorded victims of trafficking.

It is clear that migrants who choose agencies with higher grades experience fewer problems. We have also verified that these correlations hold up when we limit our analysis to agencies that are legally certified to place women outside the Middle East. This is particularly important given the government moratorium on sending informal sector workers to this region of the world.

We also verify that agency grades calculated at baseline and endline are correlated, per Figure 4.

References

Chetty, R., J. N. Friedman, and J. E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, September 2014, *104* (9), 2593–2632.

Herrmann, M., E. Walsh, and E. Isenberg, "Shrinkage of Value-Added Estimates and Characteristics of Students with Hard-to-Predict Achievement Levels," *Statistics and Public Policy*, 2016, *3* (1), 1–10.

Kane, T. J. and D. O. Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," 2008.

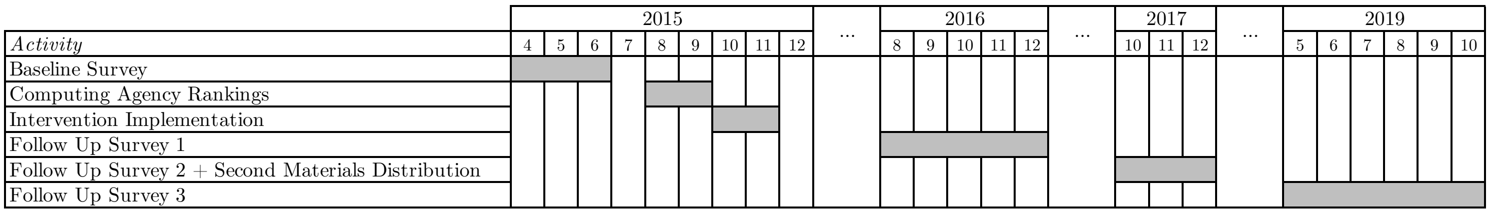
Kling, J. R., J. B. Liebman, and L. F. Katz, "Experimental analysis of neighborhood effects," *Econometrica*, 2007, *75* (1), 83–119.

McCaffrey, D. F., J. R. Lockwood, D. Koretz, T. A. Louis, and L. Hamilton, "Models for Value-Added Modeling of Teacher Effects," *Journal of Educational and Behavioral Statistics*, 2004, *29* (1), 67–101.

Wager, S. and S. Athey, "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

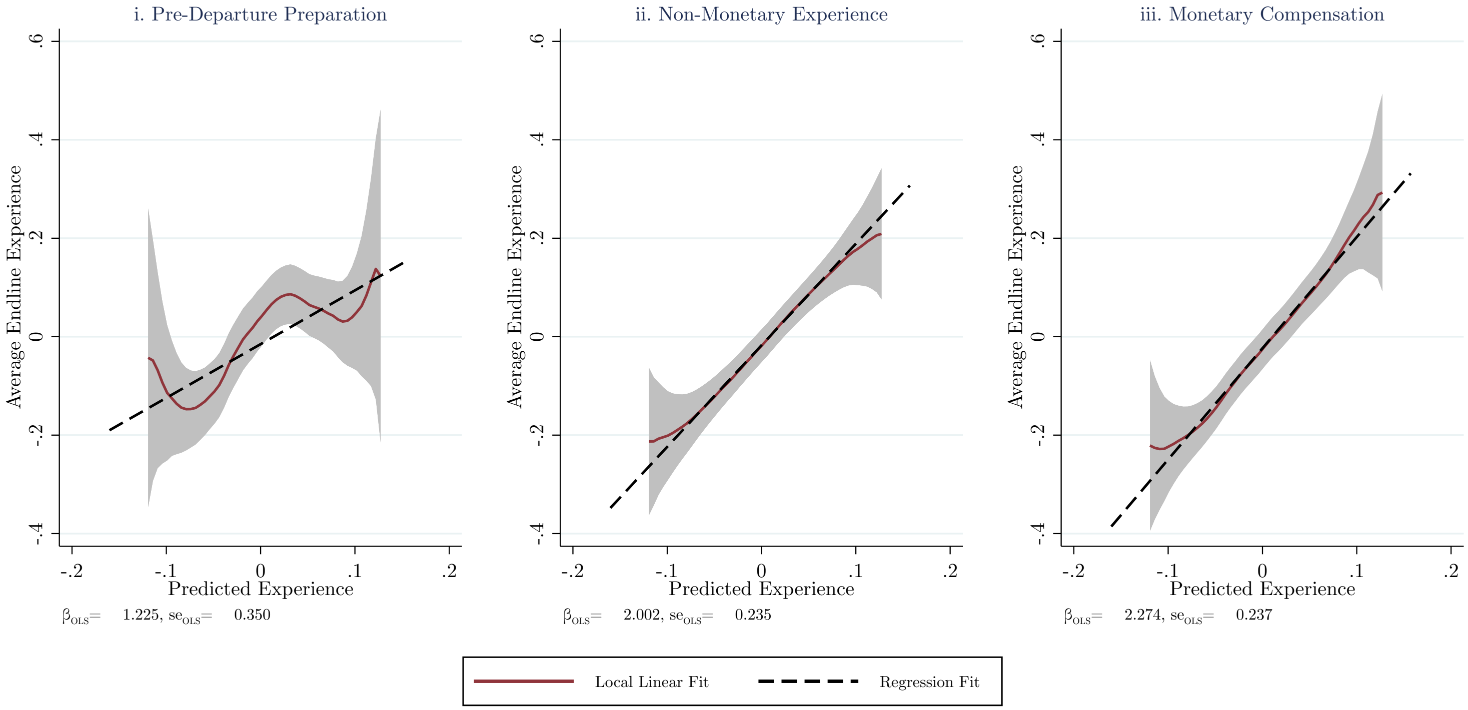
4 Supplementary Figures

Figure 1: Study Timeline



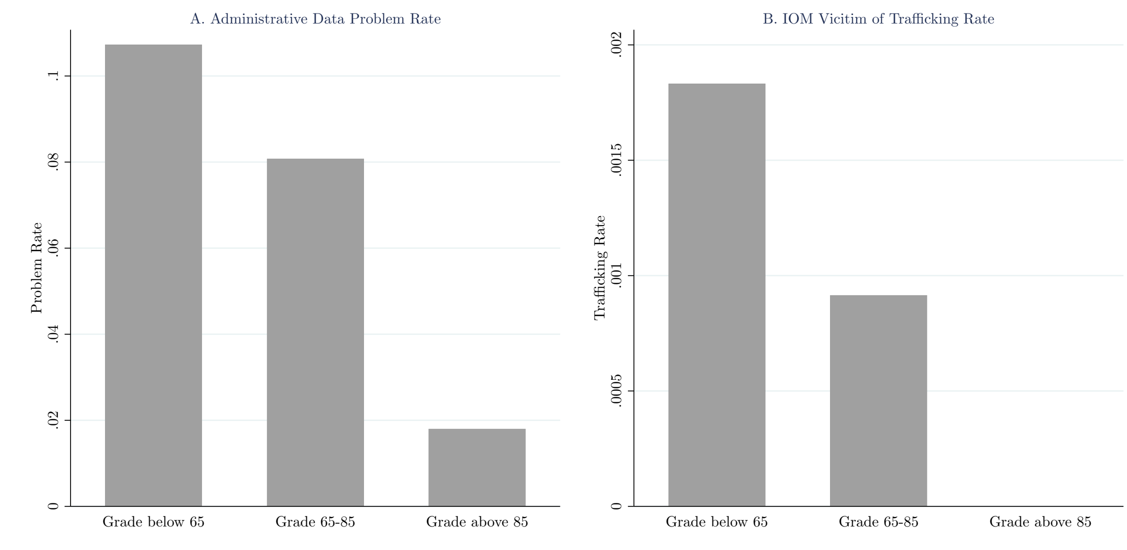
Notes: Numbers in boxes below years indicate months. The second materials distribution involved distributing report card, comic, and infographic materials to women in the relevant treatment groups participating in an in-person survey for follow up 2.

Figure 2: Validating Predicted Migration Experience



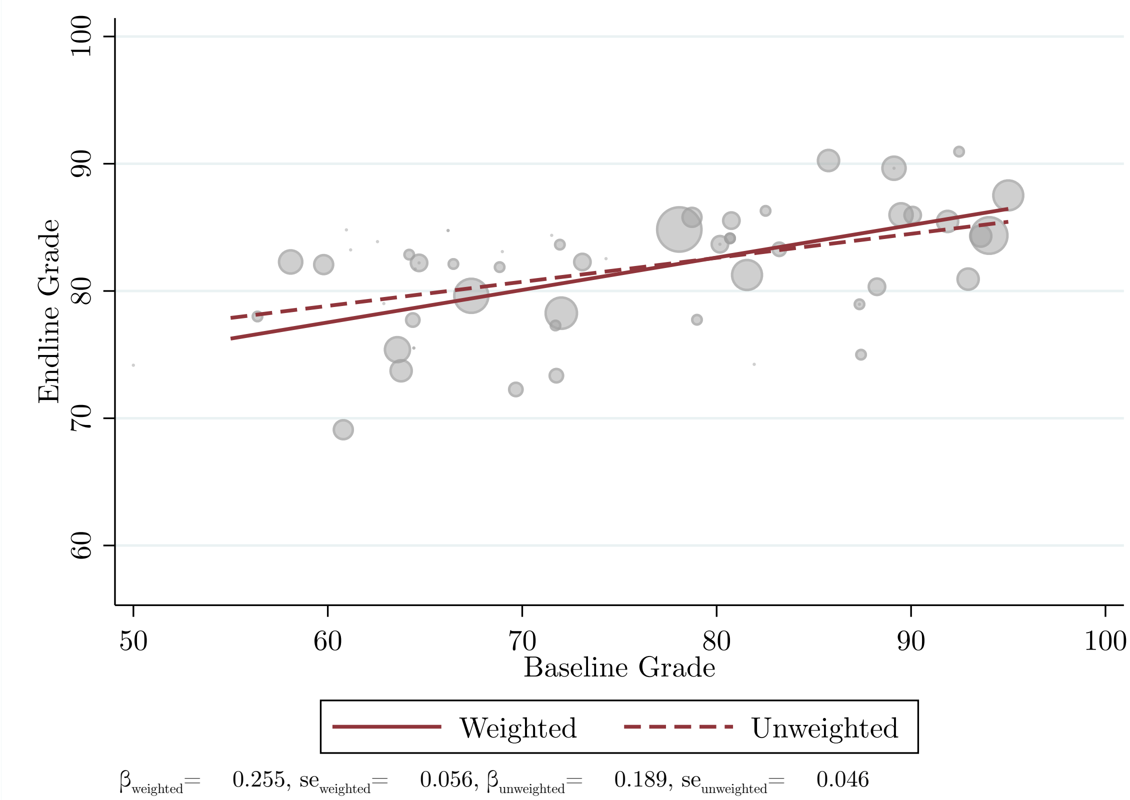
Notes: The x-axis is predicted migration experience (an average of the pre-departure preparation, job quality, and pay indices) based on a random forest model fit on pre-intervention migration experiences of tracking sample women who had already migrated at baseline. Predicted experience is trimmed at the top and bottom 1 percent. The y-axis is endline migration experience (an average of the pre-departure preparation, job quality, and pay indices) among all tracking sample women. Graphs report the results of local linear fit with 95 percent confidence intervals in grey, as well as linear regression fit. Standard errors for linear regressions are clustered at the village level.

Figure 3: Out-of-Sample Validation of Agency Quality Index



Notes: Panel A reports the mean rate of problematic repatriations in 2013 for agencies that we score below 65, 65-85 and above 85. Panel B reports the corresponding means in trafficking reported by the International Organization for Migration.

Figure 4: Correlation Between Endline Agency Grades and Baseline Agency Grades



Notes: This graph reports the correlation between agency grades calculated at endline using the extra experience sample and agency grades calculated at baseline. The size of each circle indicates the number of tracking sample women using each agency at endline. Best fit lines are from OLS regressions where endline grade is the outcome and baseline grade is the independent variable. Weighted estimates weight the regression by the number of women using the agency at endline, clustering standard errors at the agency level. Unweighted estimates are run at the agency level and include heteroskedasticity robust standard errors.