Pre analysis plan for “Empowering the next generation? Manufacturing jobs and maternal aspirations”

Andreas Kotsadam (The Ragnar Frisch Centre for Economic Research)
Janneke Pieters (Wageningen University)
Espen Villanger (Chr. Michelsen Institute)

Registered on June 6, 2019.

1. Summary of the project
Export manufacturing work provides an increasingly important earnings opportunity in industrializing developing countries, especially for women. Manufacturing firms, especially in the clothing industries, hire women on low-skilled line operation jobs. In low-income settings, these jobs might be the only opportunity to earn off-farm income for women who lack assets and skills to start their own business. At the same time, factory jobs are typically associated with social stigma. The work can be physically demanding, wages are low, and working conditions are typically poor (Blattman and Dercon 2018).

In this study we analyze intergenerational effects of these jobs. Employment is considered an important driver of women’s empowerment, and several studies show that female employment can improve child human capital outcomes by increasing women’s intra-household bargaining power (e.g. Majlesi 2016). Our aim is to understand the impacts of the jobs on women’s aspirations for the children, especially their daughters. We ask whether working in manufacturing factories affects women’s aspirations for their daughters, in terms of education, occupational choice, and overall empowerment. As possible channels, we will investigate how women perceive these jobs in terms of appropriateness/social stigma, if women update their beliefs about the content of the job in terms of skills and physical strength required and health impacts, and if the work affects their income and social networks.

2. Study design, intervention, and prior studies
We have been collaborating with 25 large shoes and garment manufacturing companies across five industrial parks in Ethiopia (see Figure 1) that were hiring new workers. The factories are
located in Tigray, Amhara, Oromia, SNNP, and Dire Dawa. The companies first assessed all job applicants and determined whether each applicant was eligible for the job or not, based on verbal and physical tests. Then, from the pool of eligible candidates at each company, a list of all women with a partner\textsuperscript{1} was compiled (this is our study sample), from which the company randomly selected half to receive a job offer. The other half was assigned to the control group and received no job offer. All applicants were informed about the procedure before the randomization was conducted.

![Industrial parks where study factories are located, with 50 km buffer zones](image)

**Figure 1: Industrial parks where study factories are located, with 50 km buffer zones**

### 3. Data sources
Pre-randomization and follow-up data have already been collected using a detailed structured survey instrument. All women in the study sample were interviewed before they started working and the first follow up data collection was around 6 months after the first interview. The dates for the data collection vary and depend in particular on when the firms hired, which happened between March 2016 and March 2018.

At baseline, 1871 women were interviewed. Of these, 374 were not randomly allocated to either the control group of a job offer due to a misunderstanding in one place and due to internet problems during the state of emergency in another. We do not include these women in our study.

\textsuperscript{1} The data collection initially focused on impacts on intimate partner violence, which is why the study population consists of partnered women.
analysis, so that our final study sample consists of 1498 randomly assigned women. Out of these, 1262 were interviewed for the first follow-up.

For the present study, we aim to interview all 1498 women in the final study sample. The survey instrument includes modules gathering demographic and background information, including poverty measures and other socio economic variables, and a module on domestic violence and female empowerment. A previous study focused on the impact of the job offer on domestic violence (Kotsadam and Villanger, 2019). The study found that the job offer increased women’s formal sector employment and income, but did not affect women’s gender attitudes, decision-making power in the household or their self-reported domestic violence.

In our analysis we will use baseline and follow-up data previously collected, as well as new follow-up data. Below, we describe all variables to be used in the analysis and we indicate in case the relevant data was collected in previous waves of the survey.

4. Variable construction

Outcome variables

Aspirations for daughter; all women

- Daughter_schooling: response to question AB1
- Daughter_factory: dummy equal to 1 if response to AB2=23 and 0 otherwise
- Daughter_occupation: based on the response to question AB2, equal to 0 if response equals 50 or 60 (no job or student); 1 if response is between 01-04 (agriculture); 2 if response equals 7 or 9 (own business); 3 if response equals 5, 6, 8, or 10-16 (informal or unskilled job); 4 if response is between 17-23 (factory job or other semi-skilled job); or 5 if response is between 24-32 (professional job).
- Daughter_empowerment: an index constructed from AB3, AB4 and AB5. First the response to AB5 will be reverse-coded so that a higher value indicates less tolerance of beating. Then each variable will be standardized, after which a weighted average is generated where the weights are proportional to the sums of the rows of the inverted variance-covariance matrix (following Anderson, 2008).
Aspirations for sons; all women
- **Son_schooling**: response to question AB6
- **Son_factory**: dummy equal to 1 if response to AB7=23 and 0 otherwise
- **Son_occupation**: based on the response to question AB7, equal to 0 if response equals 50 or 60 (no job or student); 1 if response is between 01-04 (agriculture); 2 if response equals 7 or 9 (own business); 3 if response equals 5, 6, 8, or 10-16 (informal or unskilled job); 4 if response is between 17-23 (factory job or other semi-skilled job); or 5 if response is between 24-32 (professional job).

Aspirations for daughters age 5-15; only measured for women with daughters age 5-15
- **Daughter5to15_schooling**: 0 if response to A12 equals 1 and response to A15.1 equals 0; response to A10 if response to A12 equals 3 and response to A15.1 equals 0; response to A15.2 if response to A12 equals 2 or response to A15.1 equals 1.
- **Daughter5to15_factory**: dummy equal to 1 if response to A25=23 and 0 otherwise
- **Daughter5to15_occupation**: based on the response to question A25, equal to 0 if response equals 50 or 60 (no job or student); 1 if response is between 01-04 (agriculture); 2 if response equals 7 or 9 (own business); 3 if response equals 5, 6, 8, or 10-16 (informal or unskilled job); 4 if response is between 17-23 (factory job or other semi-skilled job); or 5 if response is between 24-32 (professional job).

If the respondent has more than 1 daughter in the age group 5 to 15, we will use her answers for her youngest daughter.

Aspirations for sons age 5-15; only measured for women with sons age 5-15
- **Son5to15_schooling**
- **Son5to15_factory**
- **Son5to15_occupation**

All measured the same as for daughters age 5-15, as defined above. If the respondent has more than 1 son in the age group 5 to 15, we will use her answers for her youngest son.

Income
- **Household_income**: the sum of Respondent, Husband, and Other’s income from questions D4.1 to D4.6 (in Birr). These variables were also collected at baseline and in previous follow-up surveys.
Beliefs about factory job requirements

- **Factory_schooling**: response to KC8
- **Factory_skills**: index constructed from reading (equal to 0 if response to KC1 equals 0 or response to KC2 equals 99; otherwise equal to the response to KC2); writing (equal to 0 if response to KC4 equals 0 and otherwise equal to the response to KC5); bills (response to KC3); math (the sum of responses to KC6.A-KC6.D). These variables that will be part of the index are constructed first. Then each variable will be standardized, after which a weighted average is generated where the weights are proportional to the sums of the rows of the inverted variance-covariance matrix (following Anderson, 2008). For all variables, response 99 (do not know) is recoded to 0 and we create a separate indicator for respondents answering 99 to any of the variables.
- **Factory_physical**: response to KC7 (continuous 0-10)
- **Factory_health**: the sum of responses to KC9.A to KC9.E (continuous 0-5)

Acceptability of factory work for women

- **Factory_stigma**: Response to KA0.1 (continuous 0-10)
- **Factory_stigma_community**: Response to KA0.2 (continuous 0-10)

Beliefs about returns to education

- **Returns_vocational_job**: Difference between response to KD3.A and KD1.A
- **Returns_university_job**: Difference between response to KD4.A and KD1.A
- **Returns_grade10_factory**: Difference between response to KD5.B and KD5.A
- **Returns_vocational_factory**: Difference between response to KD5.C and KD5.A
- **Returns_university_factory**: Difference between response to KD5.D and KD5.A

Work-related social networks

- **Has_workfriends**: response to N4 (binary). We also ask about topics discussed with friends, to be used descriptively.
**Explanatory variables**

- *Treated_job*
- *Any_factory_job*: Dummy variable equal to one if the response to KA0.4 is greater than zero, and zero otherwise.
- *Months_factory_job*: Response to question KA0.4 if the response to KA0.4 is greater than zero, and zero otherwise.

**Control variables**

Our main control variables, \(X_1\), will be the same baseline variables as used by Kotsadam and Villanger (2017) and they will be coded in the same way. These are: Age, Muslim, Protestant, Any formal wage job, Medium education, and High education.

In addition to the variables in \(X_1\) we add the following variables to a full set of controls, \(X_2\).

- Whether currently has any sons and whether currently has any daughters: we construct two dummy variables from questions A3 and A4.
- Baseline gender role attitude regarding education and employment: index constructed from section GA, questions 2-8 (we do not include question GA1, as this is only answered by women with children). The index will be constructed as in Kotsadam and Villanger (2017).
- Baseline decision-making related to work and spending: index using section J, questions 6, 8, 9, 12, 14, 15 (questions on children’s schooling and sickness are not used, as they are missing for women with no children), constructed as in Kotsadam and Villanger (2017).
- Occupation of respondent’s father, constructed from baseline response to KA13. Answers will be coded as 0=did not work/Do not know/not applicable; 1=agriculture and fishing (codes 01-04); 2=retail and commercial (codes 05-10); 3=unskilled trades (codes 11-16); 4=skilled and semi-skilled trades (codes 17-23); 5= professional (24-27).
- Occupation of respondent’s mother, constructed from baseline response to KA14. Same coding as father’s occupation
5. **Estimation and main hypothesis**

Our main hypothesis is that offering women a factory job affects their educational, occupational, and empowerment aspirations for daughters. The main estimating equation to estimate the effect on educational and empowerment aspirations is the following intention to treat model:

\[ y_{i,t} = aTreated\_Job_i + c_t + e_{i,t} \] (1)

Where \( y_{i,t} \) corresponds to aspirations for daughters’ school attainment \((Daughter\_schooling)\), factory work \((Daughter\_factory)\), or empowerment \((Daughter\_empowerment)\) of woman \( i \) in time period \( t \) (e.g. \( t4 \) is the fourth follow up). This is regressed on a dummy variable that equals one if the woman was randomized to get the job and zero if not. This captures the intention to treat effect and it gives us an estimate of the effect of being randomly assigned to the group receiving a job offer. We control for list fixed effects \((c_l)\), which are blocking variables as the women are randomly assigned within lists of equally qualified partnered women.

We will test for balance across treatment and control group on all variables in the set \( X1 \) (Age, Muslim, Protestant, Any formal wage job, Medium education, and High education), and also include \( X1 \) as control variables, in order to see if we can increase precision of our estimate of \( a \). In addition, we will experiment with other sets of control variables to see if we can get more power and improve precision in the estimates by picking optimal controls from the total list of controls \((X2)\) using LASSO (Belloni et al. 2014; Ahrens et al. 2018). The main specification will, however, be the one with only the list fixed effects as controls.

For occupational aspirations \((Daughter\_occupation)\) we will use a multinomial model (logit or probit) with the same explanatory and control variables described above.

To account for imperfect compliance we also estimate the effect of treatment on the treated, using an IV model of the following form for \( Daughter\_schooling, Daughter\_factory, \) and \( Daughter\_empowerment: \)

\( Any\_factory\_job_{i,t4} = aTreated\_Job_i + c_t + e_{i,t} \) (2)

\[ y_{i,t4} = Predicted(Any\_factory\_job)_{i,t4} + c_t + e_{i,t} \] (3)
We will also test the IV regressions with the variable, *Months_factory_job*, to explore whether the total time spent in factory work is a relevant channel. We will add control variables as described above. For the outcome variable *Daughter_occupation* we will estimate the effect of treatment on the treated using a limited information maximum likelihood estimator, following Roodman (2011).

6. **Mechanisms and other hypotheses**

We will also test a number of secondary hypotheses and potential mechanisms. In particular we will run equation 1 and the IV regressions with the outcome variables listed in section 4: Income, Beliefs about factory job attributes, Appropriateness of factory work for women (own and perceived community), Gender attitudes, and Work related social networks.

For the aspirations related to schooling, we will run the estimations described in section 5 above for *Daughter5to15_schooling*, *Daughter5to15_factory*, and *Daughter5to15_occupation* on the subsample of respondents with at least one daughter in the age group 5 to 15.

The aspiration regressions for schooling and occupation will also be run with sons instead of daughters.

7. **Power calculations**

We expect to have between 1000-1300 women still included in the study with around 50 percent assigned to treatment. Assuming we have 1000 women and do not control for anything, we still have a relatively small minimum detectable effect of 0.17 s.d with 80 percent power and a significance level of 5 percent.

8. **Other issues**

Attrition: Attrition from the sample will be investigated. In particular, we will check whether attrition is related to treatment status by the following regression:

\[ \text{Attrition}_i = a \text{Treated}_i + c_t + e_{i,t} \]
Where *Attrition* is dummy equal to 1 if individual i is not answering the fourth (or later) follow-up. Unless the difference between treatment and control is significantly different from zero at the 5 percent significance level, all estimations will proceed without any adjustment for attrition. If there is a statistically significant difference we will employ Lee bounds (Lee 2009).

Limited variation: In order to limit noise caused by variables with limited variation, questions for which 95 percent of the observations have the same value within the relevant sample (for both treatment and control together) will be omitted from the analysis.

Missing values on independent variables: If we have missing values on variables we will code the variables as zero and include dummy variables controlling for missing status so that we do not lose observations.

**References**


