Subsistence, Transformational and Something in Between

By Elisa Gamberoni, Rachel Heath, and Emily Nix*

December 17, 2012

1 Introduction

Workers in the developing world, particularly in Sub-Saharan Africa, are frequently self-employed. In fact, of the adults in Sub-Saharan Africa whose primary employment is not family farming, more workers (51.4 percent) work in household enterprises than for wages in public or private farmers or in agriculture. [6].[1] In Sub-Saharan Africa, self-employment often constitutes the only alternative outside the agriculture sector. For example in Uganda, despite impressive growth rates, the majority of the nonagricultural employment was created in household enterprises and microenterprises[1]. Women are well-represented among these smaller size businesses: in Tanzania, for instance, 80 percent of micro entrepreneurs are female[7]. Given that such a large proportion of the labor force is employed in microenterprise, supporting them is key to poverty alleviation strategies. Futher, since women generally own

*Gamberoni:World Bank, egamberoni@worldbank.org; Heath: The University of Washington, rmheath@uw.edu; Nix: Yale University, emily.nix@yale.edu. This work was made possible with support of the UN Foundation – ExxonMobil Foundation research collaboration on women's economic empowerment (http://www.unfoundation.org/features/building-a-roadmap.html) and with support from the Bank Netherlands Partnership Program. We also thank Elena Bardasi, Louise Fox, and participants at the United Nations Greentree Convening on Women's Economic Empowerment for helpful comments. We are grateful to Paolo Falco, Thomas Pave Sohneson and Jorge Huerto Munoz for help with the data used in the report. . The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent.

[1]Overall, 18 percent of respondents worked in household enterprise.

smaller businesses, understanding gender traits that affect profitability is key for identifying the appropriate strategy.

A key challenge to designing effective policy efforts is the substantial heterogeneity of microenterprises, and how this heterogeneity among entrepreneurs results in potentially heterogeneous outcomes for a given policy. Schoar argues in "The Divide between Subsistence and Tranformational Entrepreneurship" that one important distinction, perhaps the most crucial distinction, is between subsistence entrepreneurs and transformational entreprenuers[8]. Schoar defines subsistence entrepreneurs as those who turn to self-employment solely to provide subsistence income for themselves and their families. These entrepreneurs choose to be self-employed as an alternative employment opportunity, perhaps driven to self-employment due to lack of wage jobs or lack of skills necessary for wage jobs. In contrast, the transformational entrepreneurs establish businesses with the ability and goal to grow and create jobs for others[8].

In such a setting, subsistence entrepeneurship may not be a first step towards tranformational entrepreneurship for the vast majority engaged in subsistence enterprises. The implications of this could be pivotal for policy. For example, if the goal of policy is growth then identifying which of the subsistence entrepreneurs is able to transition to transformational entrepreneurship and focusing a policy on these individuals while supporting the existing transformational entrepreneurs may be the most efficient policy. As another example, improving access to capital may increases the profits of transformative entrepreneurs but not subsistence entrepreneurs, who may have low entrepreneurial ability or high opportunity cost of their own time and may not choose to make profitable investments that require additional time or attention to supervise. Similarly, if one's goal is poverty alleviation, it may not be possible to dramatically increase profits of purely subsistence entrepreneurs by providing money into their business. Instead, more impact may be possible by supporting growth policies that make jobs with better income paths available.

The goal then is twofold. First, identify subsistence and transformational entrepreneurs. However, the distinction is surely not quite so severe, as it seems possible that some subsistence entrepreneurs could transition to transformational entrepreneurship. Identifying how this transition can be achieved provides the second goal. The aim of this report is to provide descriptive evidence on these two questions.

How should one distinguish between subsistence and transformational entrepreneurs? Ex post, the distinction is clear, as transformational entrepreneurs are by definition those businesses that achieve growth. Such enterprises will thus consist of businesses that hire employees and yield positive profits after

subtracting expenses, including basic subsistence income.

This implies that a distinction based on the size of the business might be a useful first approximation: the self-eemployed without employees will include the subsistence entrepreneurs who face different challenges compared to transformational micro or small enterprises. However, these classifications may mask substantial heterogeneity within these groups. Specifically, businesses with no employees surely include some transformational entrepreneurs in the making. This heterogeneity may have an important gender dimension as well, if the differences within and between groups differ by men and women. For example, Fafchamps et al (2011) focus on a sample of Ghana based enterprises without employee and find that only those women running firms with above median earnings at baseline benefit from receiving in-kind grants[5]. This is not the case for men. This would suggest that in this setting, men across the distribution of firm profits are capable of some level of transformational entrepreneurship while women who are poised to be transformational entrepreneurs are primarily found at the top of the distribution of business profits.

This report provides descriptive evidence on traits that could be used to identify transformational versus subsistence entrepreneurs, given ex post income results. The report uses nationally representative data from Ghana, Rwanda, and Tanzania to identify gender specific traits that affect earnings. The analysis starts by looking at difference in the mean and variance of earnings between wage and self-employed workers conditional on key traits such as gender, age, and education. We then categorize the diversity of personal traits and input usage within broad microenterprise categories such as whether the enterprise has employees. We use nationally representative data for Tanzania, Ghana, and Rwanda to suggest that aspects of our findings persist in a variety of settings in sub-Saharan Africa. Given that each country displays substantial heterogeneity within each of the broad classifications, we then explore which of these traits lead to differences in microenterprise profits.

We do this first by displaying cumulative distribution functions of microenterprise earnings by gender, and then comparing these distributions to earnings distributions conditional on observable characteristics of the microenterprise. Given that these distribution functions display gender differences in earnings, we then conduct an Oaxaca decomposition that explains what proportion of the difference in earnings by gender is due to differences in traits versus in returns to these traits. Given that we find considerable variation in microenterprise success based on the personal characteristics we examine, we then investigate where the particular populations targeted by interventions fit in the overall distribution of income in each country. Specifically, we use the

results from a randomized trial to predict the impact of a specific intervention across the groups. We present this evidence to examine the potential usefulness of interventions that have been demonstrated to have specific effects in one country in other contexts. Such results also provide a bigger picture of the costs and benefits of an intervention, by placing the results in a country wide distribution of income.

2 Data

To examine the self-employed, we rely on two data sets. We focus mostly on a rich cross country data set that includes Ghana, Rwanda and Tanzania but also conduct limited analysis using a panel data set from Ghana.

For the cross-sectional analysis we have generated a rich cross country data set from a series of labor force surveys. The analysis is done on three countries for this draft: Ghana, Rwanda and Tanzania. In future drafts additional countries will be added to the analysis. In our computed profits, we follow de Mel, Mckenzie and Woodruff's paper "Measuring Microenterprise Profits: Must We Ask How the Sausage is Made?"[3]. Unfortunately, the surveys do not ask profits, which is their recommended measure. Instead, we calculate profits as revenues-expenses, which are collected, and are able to make many of the adjustments they suggest in their paper in calculating Ghana and Rwanda's profits. Unfortunately, there is minimal information available for Tanzania, so we are unable to compute profits as rigorously in Tanzania..

Table 1 provides descriptive statistics for the cross country data set for the three main occupation types: agricultural employees, wage employees, and the self-employed.

Table 1: Descriptive Data on Employment Types

	(1) mean	sd	min	max	count
Agriculture					
Monthly Income in USD	.	.	.	.	0
Education (Categorical, 1-8)	3.11988	1.322324	1	8	22731
Age in Years	36.37232	13.32529	15	65	28537
Male Dummy	.5003679	.5000086	0	1	28537
Married Dummy	.652276	.4762562	0	1	28537
Paid Employment					
Monthly Income in USD	78.8756	126.4605	0	3131.868	10113
Education (Categorical, 1-8)	3.922381	2.111999	1	8	9946
Age in Years	33.51334	11.78493	15	65	10381
Male Dummy	.6710336	.4698604	0	1	10381
Married Dummy	.4721125	.4992457	0	1	10381
Self-Employed					
Monthly Income in USD	94.61385	237.6206	-383.5302	3127.436	8458
Education (Categorical, 1-8)	3.764384	1.564015	1	8	9316
Age in Years	35.14116	10.926	15	65	9939
Male Dummy	.4564846	.4981279	0	1	9939
Married Dummy	.605896	.488682	0	1	9939
Total					
Monthly Income in USD	86.04345	185.6985	-383.5302	3131.868	18571
Education (Categorical, 1-8)	3.452933	1.636843	1	8	41993
Age in Years	35.5144	12.60205	15	65	48857
Male Dummy	.5277033	.499237	0	1	48857
Married Dummy	.6045602	.4889499	0	1	48857
Observations	48857				

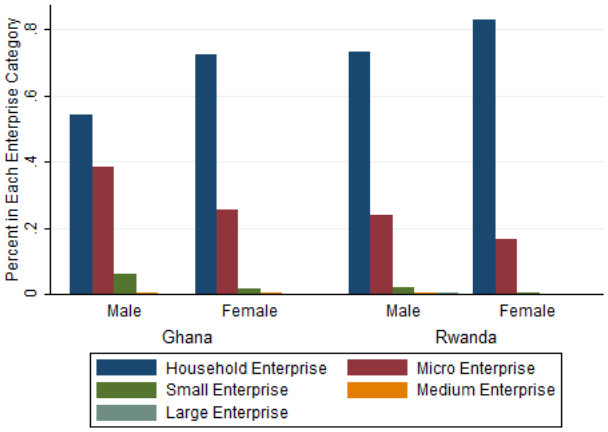


Figure 1: Enterprise Size by Country and Gender

The Ghana panel data set allows us to observe self-employed earnings over time. Such data is pivotal in order to identify a transition from subsistence to transformational entrepreneurship. In this data set we have 214 self-employed individuals who are observed for 4 consecutive years., from 2005-2008. We use this sample in the assessing the savings intervention in Part 6.

Concerning the type of businesses under analysis, Figure 1 illustrates the percentage of male and female owners of household (defined as a business without employee), micro (1 to 4 employees), small (5 to 19 employees), medium (20-99 employees), and large (100 + employees) businesses. As the figure shows, the sample, and thus the analysis, mainly focuses on household, micro, and to a lesser extent small businesess. Additionally, women appear more likely to own smaller size businesses compared to men, in line with the literature that generally shows a negative correlation between female ownership and business size (WDR, 2012).

3 Self-Employment in Ghana, Rwanda and Tanzania

To motivate the analysis below, we first present a description of self-employment and its place in the larger economy. The data show that a very large portion of individuals in countries in Sub-Saharan Africa are self-employed - far more than in developed countries. The data further show that self-employment

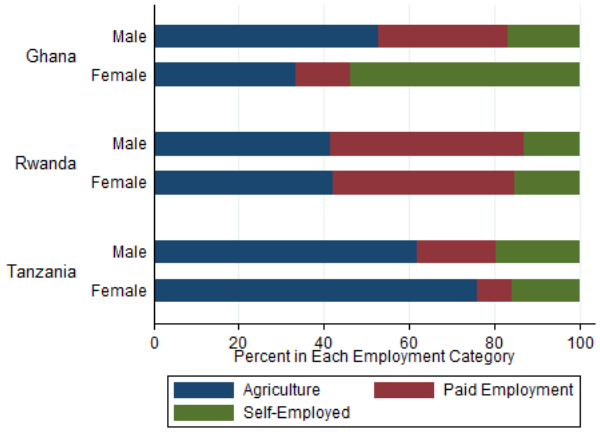


Figure 2: Percent in Agriculture, Paid Employment, and Wage Work by Country

tends to provide marginally higher average income but much higher variability in income. For each of these statistical facts, we will also point out the gender differences that exist.

3.1 Self-Employment Levels in Sub-Saharan Africa

The breakdown of primary employment by country and gender is shown in Figure 2 below. Clearly, agriculture is still the primary occupation in Ghana, Rwanda and Tanzania. However, for all but Rwanda, the second most likely employment category is self-employment. In the paper "Household Enterprises in Sub-Saharan Africa - Why They Matter for Growth, Jobs and Poverty Reduction", Louise Fox points out that self-employment is not only a large, but also a growing sector in countries across subsaharan Africa[6].

There are distinct gender and country differences in the type of primary employment held. In Ghana and Tanzania very few women report paid employment as their primary form of employment. Most women in Tanzania work in agriculture, but close to 20% report self-employment as their primary form of employment. In Ghana, the majority of women report self-employment as their primary form of employment. In Rwanda, close to 80% of women report agriculture and wage employment as their primary employment, with a much smaller percentage reporting self-employment as their primary form of employment. Men are more likely than women to report wage employment as their primary form of employment in all three countries.

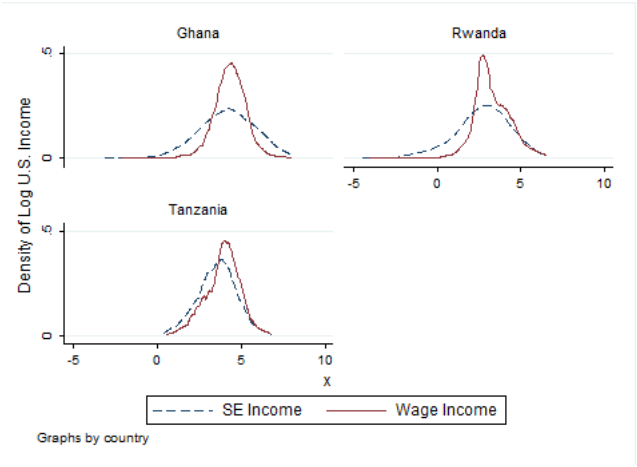


Figure 3: Distribution of Log U.S. Monthly Income by Country and Employment Type

3.2 Self Employment: Higher Mean, Greater Variance in Income

In Figure 3 below we present graphs of the distribution of primary incomes across countries, pulling from a rich cross country data set. Self-employment, relative to wage employment, offers marginally higher average wages for men. This is more easily seen in Table 2. However, self-employment is characterized by far greater variability in earnings. For women, self-employment offers, relative to wage employment, higher average earnings only in the case of Ghana. This is again accompanied by higher variability. The results thus indicate a positive relation between higher earnings and volatility, which can be considered a proxy for risk.

Men have substantially higher average earnings, both in wage employment and in self-employment, though men also have greater variability in their earnings. This can be seen in the graphs of the income distributions by countries below as well as in Table 2. This seems in line with the idea that women are more risk averse than men: Croson and Gneezy (2009), reviewing recent studies, suggest that women are more risk averse than men but gender difference in financial risks, for example, disappear among high level professionals (such as mutual fund managers) [2]. In our sample, higher risk aversion (proxied by difference in the variance of earnings) seems to also imply lower earnings for women compared to men.

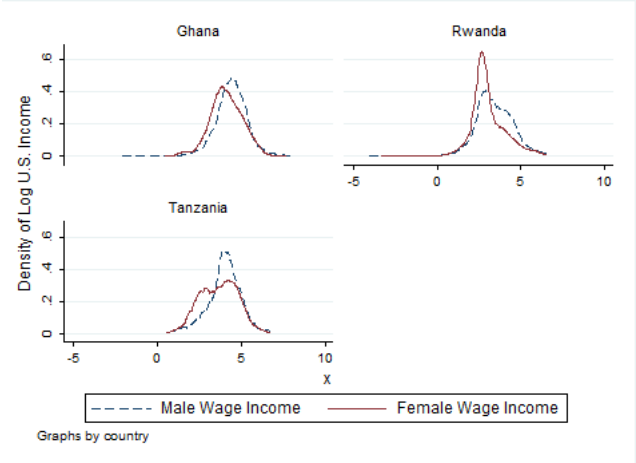


Figure 4: Distribution of Wage Workers' Log U.S. Monthly Income by Country and Gender

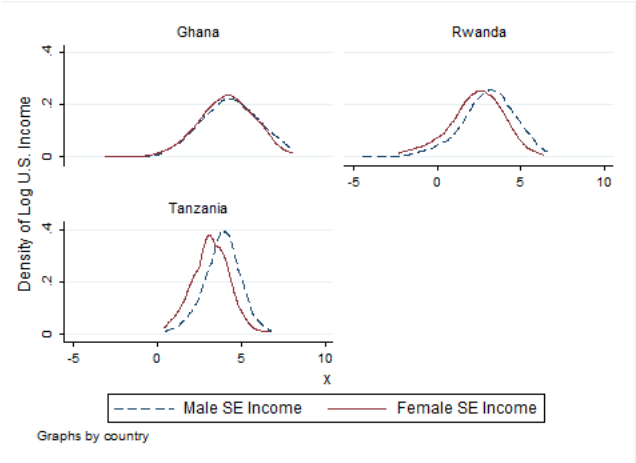


Figure 5: Distribution of Self-Employed Log U.S. Monthly Income by Country and Gender

Table 2: Mean and Standard Deviation of Monthly Income in U.S. Dollars

		Mean			Standard Deviation		
		Ghana	Rwanda	Tanzania	Ghana	Rwanda	Tanzania
Wage Employment	Male and Female	127	50	81	200	81	97
	Male	136	56	87	214	83	101
	Female	102	41	70	152	76	87
Self Employment	Male and Female	152	46	66	371	84	93
	Male	168	58	83	431	97	105
	Female	146	30	45	350	57	69

3.3 Unconditional and Conditional CDF of Incomes

We first look at the unconditional cumulative distribution function (cdf) of earnings for each gender in wage employment and self-employment in Ghana, Tanzania, and Rwanda in order to see directly how earnings of wage versus self employment workers compare at each point in the earnings distribution. If the cdf for the self-employed is below the cdf for wage workers, then at these points the self-employed make less on average than wage workers. On the contrary, if the wage workers' cdf is above that of the self-employed, this indicates that wage workers earn higher wages than the self-employed at this point in the distribution of earnings.

For Tanzanian and Rwandan women, the cdf for self-employment income is strictly dominated by the cdf for wage income. For men in these two countries, only a tiny minority do better in self-employment at the right hand tail of the distribution. For these two countries in particular it seems unlikely, based on the unconditional cdfs, that the majority of individual who are self-employed would optimally choose self-employment if wage employment were available. This implies that some number of individuals may be seeking self-employment as an alternative to unavailable wage employment, and are more likely to be subsistence entrepreneurs. However, the selection could be warranted based on observable differences in endowments or skills, which we will discuss in more detail below. In Ghana, the unconditional cdfs cross with wage work providing higher average income on the left hand side of the distribution but self-employment providing higher average income on the right hand side of the distribution. In Ghana, this is the case for both men and women.

On the left hand side of the graph for each of the three countries, the cdf of wage employees is below that of the self-employed. This means that there are more workers with very low earnings in self-employment across countries and genders. However, the extent of the gap between the earnings of the self-employed and wage workers varies by country and gender. For example, among male workers in Tanzania, self-employed workers do worse, but only by a small margin: 25 percent of self-employed workers earn below $25 a month, whereas 17 percent of wage workers earn below $25 a month. In contrast, among female workers in Rwanda, self-employed do substantially worse: 25 percent of self-employed workers earn below $5 a month , whereas 5 percent of wage workers earn below $5 a month. This gap does not appear to be driven by greater degree of part-time work in self-employment versus wage employment in Rwanda. There do not appear to be major differences in hours worked between the self-employed and wage workers for men and women for all three countries. However, the percent self-employed is much lower in Rwanda, so

perhaps the underlying sectoral selection mechanism explains part or all of the difference. This can be seen in Figures 9 and 10 which show the hours worked distributions for wage workers and the self-employed for men and women for all three countries. Similar to developed countries, women have greater variability in their hours. Further, self-employed women have even greater variability in hours worked than wage employed women, whereas the hours worked of self-employed and wage employed men looks very similar. Women who are self employed work 7 hours less per month on average in Ghana and on average 18 hours less per month in Tanzania. However, unlike developed countries, there do not appear to be any distinctive peaks at part time hours, though there is a peak at what is presumably full time hours in Ghana and Tanzania.

There are also differences among the high earners. In Ghana, both male and female self-employed workers out-earn wage workers at the upper end of the distribution. In Tanzania and Rwanda, by contrast, the upper end of the earnings distribution is very similar among wage and self-employed workers, for both men and women.

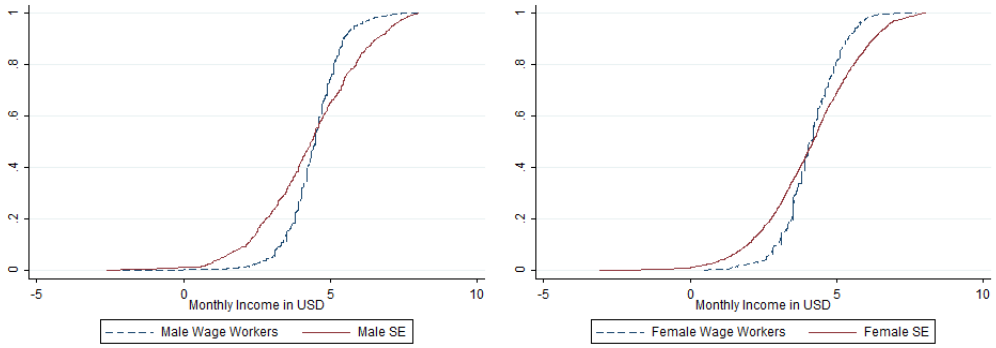


Figure 6: Ghana Male and Female Unconditional CDF of Monthly Income by Employment Type

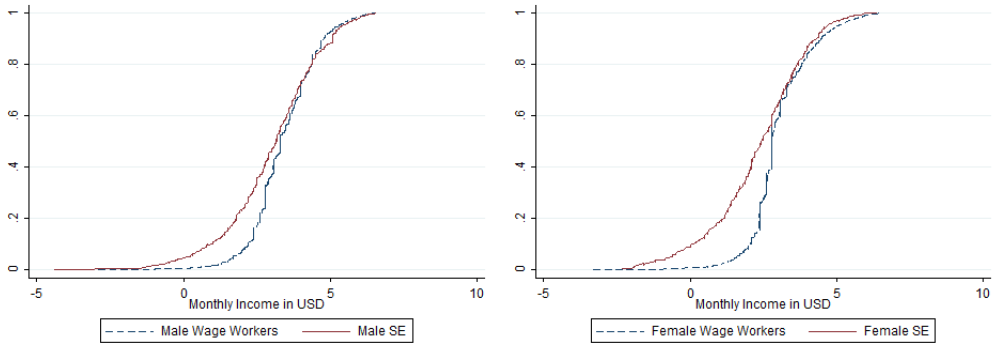


Figure 7: Rwanda Male and Female Unconditional CDF of Monthly Income by Employment Type

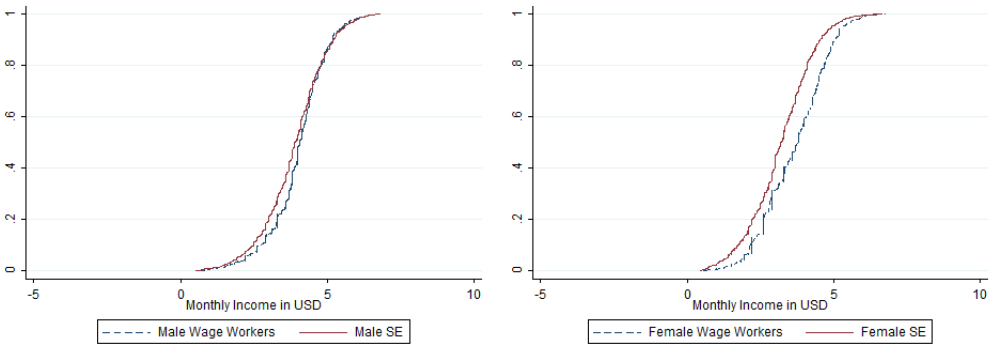


Figure 8: Tanzania Male and Female Unconditional CDF of Monthly Income by Employment Type

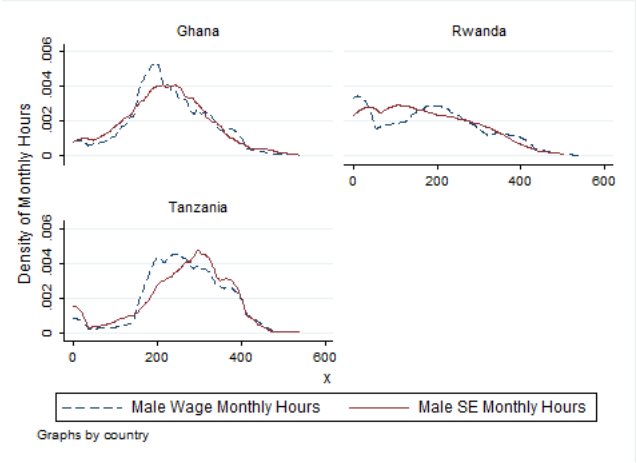


Figure 9: Distribution of Male Monthly Hours Worked by Country and Occupation

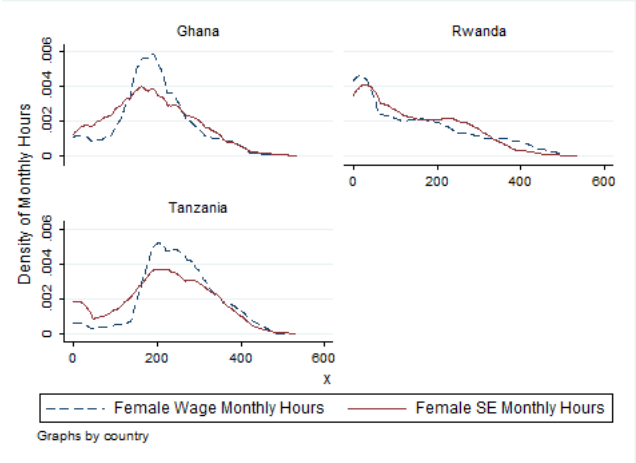


Figure 10: Distribution of Female Monthly Hours Worked by Country and Occupation

If the left tails of the unconditional cdfs demonstrate, as they appear to do, that wage work is far more profitable than self-employment, why do individuals at the lower end not choose to do wage work instead? Further, in Tanzania and Rwanda the unconditional cdfs for self-employment income are dominated by the unconditional cdfs for wage employment. The selection into self-employment, even when it appears to pay far lower than wage work, could be due to constraints on the number of wage jobs. However, the shapes of the unconditional cdfs could also be due entirely to selection. If individuals in Rwanda and Tanzania lack the education necessary to succeed in wage work, for example, the selection into self-employment could be optimal without the existence of any constraints on available wage jobs. To address this possibility we also look at the conditional cdfs of wage income and self-employment income. As with the unconditional cdfs, we present the conditional cdfs separately by country and gender.

The conditional cdfs are constructed as follows. First, we do a regression with log income in U.S. dollars as the dependent variable. We include education, gender, age, dummies for marital status, and whether the individual has a savings account as independent variables. We do the regressions separately for self-employment and wage employment. This allows the returns to education, for example, to be different for self-employment and wage employment. We then use the coefficients generated by the regressions to predict wage employment income for the self-employed and hypothetical self-employment income for wage workers. This means that for a self-employed individual, we use

the regressions to obtain both a predicted value of his/her self-employment earnings given the observables as well as a predicted value for his/her wage income, given his/her observables. We then plot the graph with the predicted self-employed and wage employment earnings for wage and self-employed individuals separately, y gender and country. Mathematically, the predicted values are produced as follows, with $\hat{y}$ equal to monthly log income in U.S. dollars and the coefficients, $\hat{\beta}^i$, estimated in the regressions.

$$\hat{y}_{wage} = \hat{\beta}^1_{wage} education + \hat{\beta}^3_{wage} age + \hat{\beta}^4_{wage} married$$

$$\hat{y}_{se} = \hat{\beta}^1_{se} education + \hat{\beta}^2_{se} savings + \hat{\beta}^3_{se} age + \hat{\beta}^4_{se} married$$

An F-test of the homogeneity of the variances shows that the distributions of the predicted values have not only different means but also variances. We thus proceed to graph the conditional cdfs below.

If the choice was driven entirely by the combination of differences in endowments or returns to endowments between the two occupations, this should be reflected in the conditional cdfs. For example, for wage workers, we would expect the conditional cdf for wage work income to be entirely below that of the conditional cdf for self-employment income. Similarly, for the self-employed, if selection into an occupation is entirely explained by observables, the conditional cdf of self-employment income should always be below that of projected wage work income.

However, this is never the case. For example, for Ghanaian men, the conditional cdf for wage work is below that of self-employment at the lower end of the distribution, which implies that wage work provides higher average income for Ghanaian men in the left tail. This is true both for men currently in wage work as well as for self-employed men. In other words, the self-employed at the lower tails would, conditional on observables, do better in wage work and those in wage work do better than they would do if they were self-employed. However, the conditional cdfs then cross, which implies that at the right tail, Ghanaian men currently in wage work would appear to be better off selecting into self-employment. The same is true for Ghanaian men currently self-employed, namely those with lower incomes would appear to do better in wage employment while those with higher incomes appear to be choosing correctly, and should be in self-employment. For Rwandan men and women, the results from the graphs of the conditional cdfs are similar, though the point where the conditional cdfs cross is different. For Tanzanian men the exact opposite is true, as demonstrated in the graphs below.

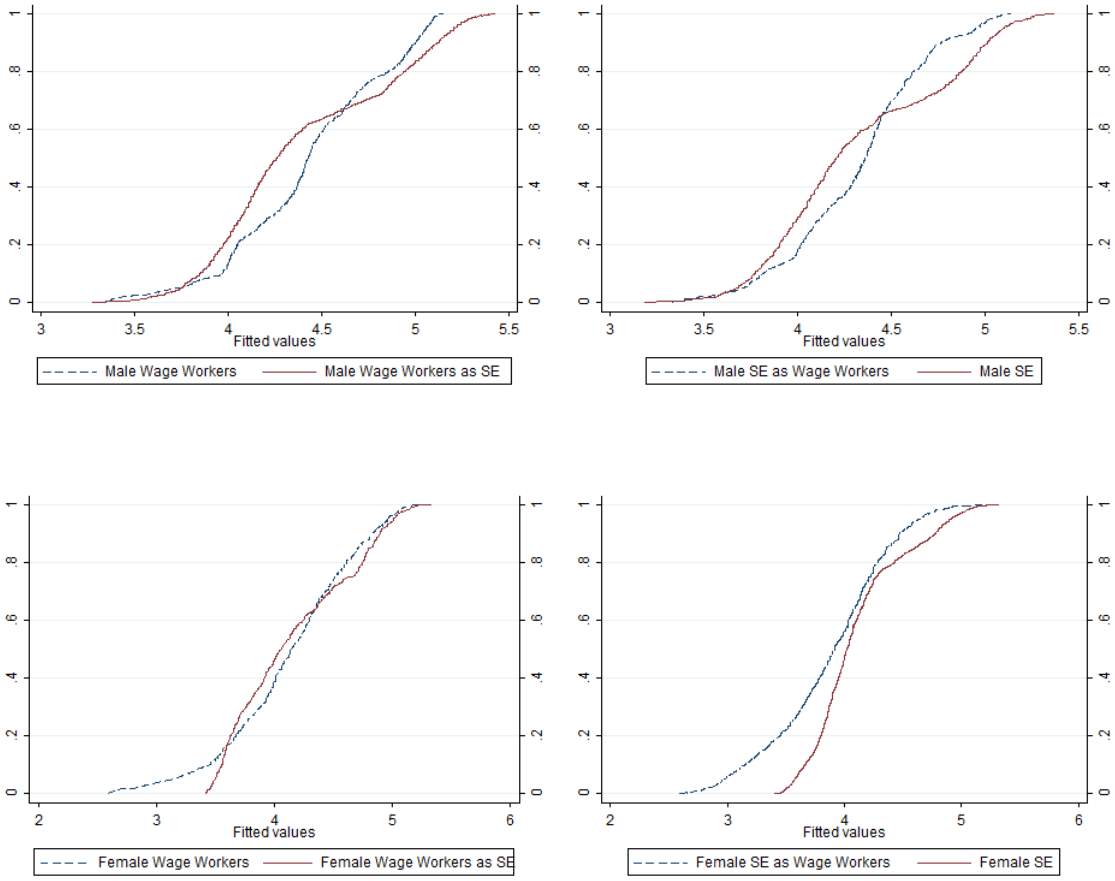
For Ghanaian women, the story is different. What is particularly interesting is that the predicted optimal occupation across the distribution differs for women currently in wage work and self-employed women. For female wage workers, the conditional cdf for self employment is below that of wage work in the left and right hand tail, which implies that in the tails, women are better off self-employed. However, in the middle of the distribution, conditional on observables, women in wage work appear to be making the most profitable decision. In contrast, conditional on observables, self-employed women in Ghana have a conditional cdf for self-employment income that is strictly dominated by that for wage work, which has the surprising implication that conditional on observables in this data set, self-employed women would be better off in wage work in Ghana, even though it is not the case that all women currently in wage work should be in wage work.

For Tanzanian women, the results are somewhat similar to those for Ghanaian women. Some differences do exist, though, as for self-employed women, self employment appears to be preferable in the extreme left tail and for women in wage work the conditional cdfs cross only once implying that self-employment is preferable in the lower tail, but wage work is preferable beyond that point.

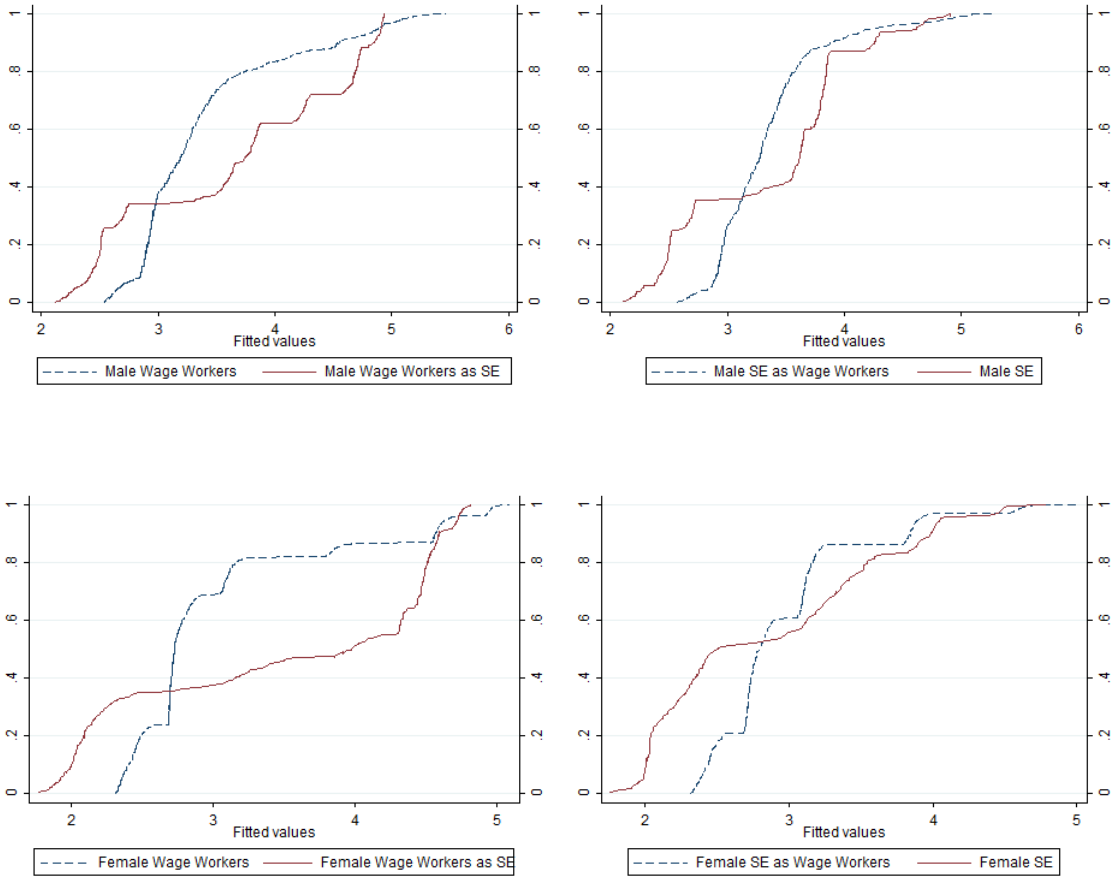
However, it is important to point out that the explanatory power of our regressions is quite low. This means that some important explanatory variables, such as risk aversion or entrepreneurial ability, are unobserved. As discussed above, risk aversion could drive occupational choice given that self-employment income is far more variable than wage income. This would imply that those who are more risk averse would be more willing to engage in self-employment. Entrepreneurial ability could obviously drive individuals into self-employment, and cause these individuals to have higher earnings in self-employment than in wage employment. Thus, while conditional on our observables, it may appear that certain individuals would be better off in either wage work or self-employment, their choice could very well have been driven by these other factors which are not observed in our data.

In summary, while our basic observables clearly cannot entirely account for the occupational choices we see in the data, differences in observable traits play a role in the selection into an occupation.

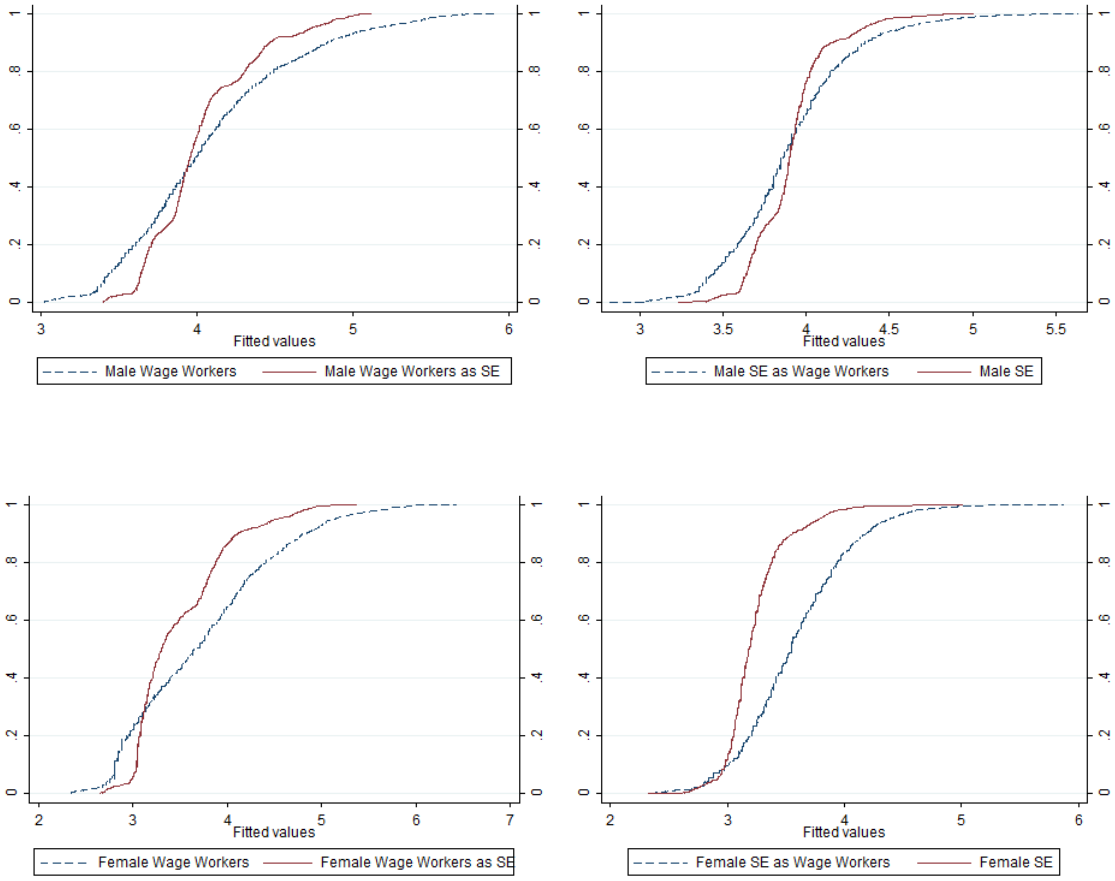
Ghana


Fitted values
Male Wage Workers
Male Wage Workers as SE
Fitted values
Male SE as Wage Workers
Male SE
Fitted values
Female Wage Workers
Female Wage Workers as SE
Fitted values
Female SE as Wage Workers
Female SE

Rwanda


Fitted values
Male Wage Workers
Male Wage Workers as SE
Fitted values
Male SE as Wage Workers
Male SE
Fitted values
Female Wage Workers
Female Wage Workers as SE
Fitted values
Female SE as Wage Workers
Female SE

Tanzania


Fitted values
Male Wage Workers
Male Wage Workers as SE
Fitted values
Male SE as Wage Workers
Male SE
Fitted values
Female Wage Workers
Female Wage Workers as SE
Fitted values
Female SE as Wage Workers
Female SE

3.4 How much heterogeneity can be captured by standard policy-relevant splits?

One question that the heterogeneity documented in the previous section prompts is whether policymakers who are interested in targeting certain parts of the earnings distribution can do so by clustering entrepreneurs by rough categories, either outcomes (such as whether the business has employees), reported reasons for beginning business (whether the respondent self-reported beginning the business because he/she "could not find a good job" or for "pull" factors such as good business opportunity), or by personal traits related to business success. The following graphs provide the earnings distributions within each of these cluster for males and females in Tanzania (for whom we have their reported reason for beginning a business). The first panel groups entrepreneurs based on whether their business has any employees. The distribution of businesses with employees is shifted to the right for both genders, as expected, although there is substantial overlap, particularly among women. The second panel groups entrepreneurs based on whether they report a push or a pull reason for starting their business. Interestingly, there is no difference in average earnings across the groups for either men or women. The "push" entrepreneurs have higher variance in their earnings in women but lower variance for men. The final panel groups entrepreneurs based on traits. Specifically, we display the profits for workers above and below the median age (32) and with and without primary education. While there is a lot of overlap in the distributions, they are also different from each other, especially among women. This suggests that traits such as age and education might be more appropriate for constructing homogeneous groups of entrepreneurs in terms of earnings (and thus for constructing target groups). In the next section, we thus formally analyze the impact of entrepreneurial traits on profitability, analyzing whether certain traits might matter more for female entrepreneurs compared to men.

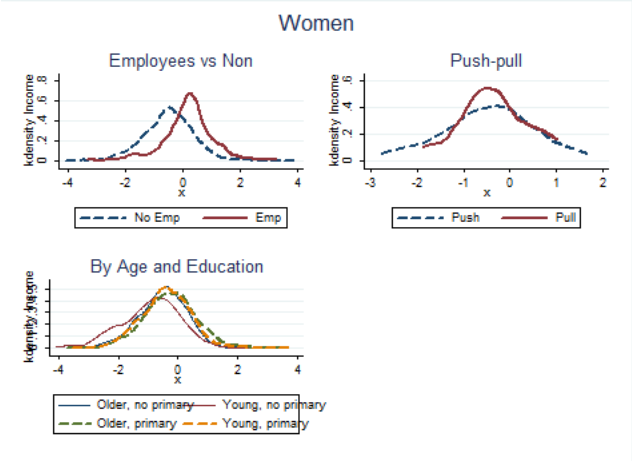


Figure 11: Distribution of Female Income for Selected Clusters by Country

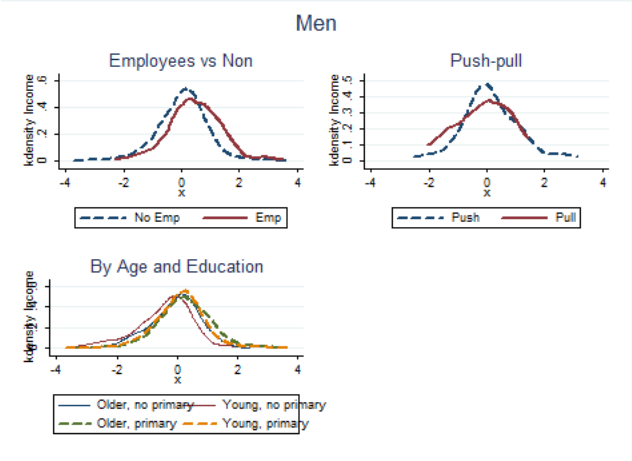


Figure 12: Distribution of Male Income for Selected Clusters by Country

Gender Decomposition of Profits

As presented in Section 3, employment choices and outcomes vary a great deal across men and women. Men earn more on average than women in all categories of employment. From Table 2 we see that on average, male entrepreneurs in Ghana make $168 a month compared to $146 for female entrepreneurs for a difference of $22. In Rwanda, male entrepreneurs on average make $58 a month compared to $30 for female entrepreneurs for a difference of $28. In Tanzania, male entrepreneurs on average make $83 a month compared to $45 for female entrepreneurs for a difference of $38. These are not insignificant differences, and yield fairly large differences in annual incomes. The Figures 13, 14 and 15 below demonstrate the differences in male and female self-employed income for Ghana, Rwanda and Tanzania.

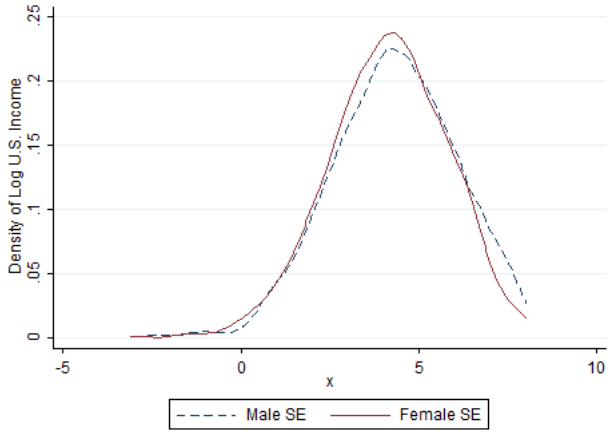


Figure 13: Distribution of Income by Gender in Ghana

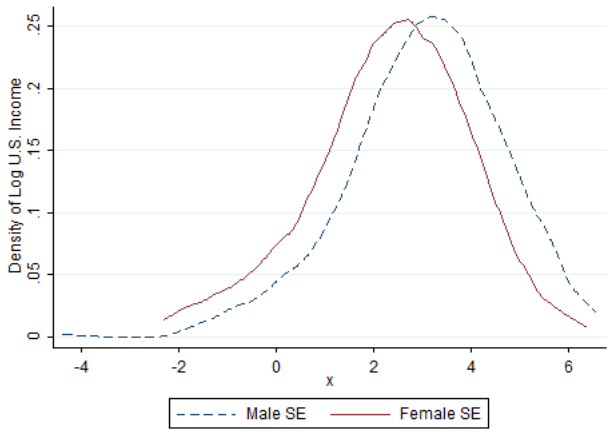


Figure 14: Distribution of Income by Gender in Rwanda

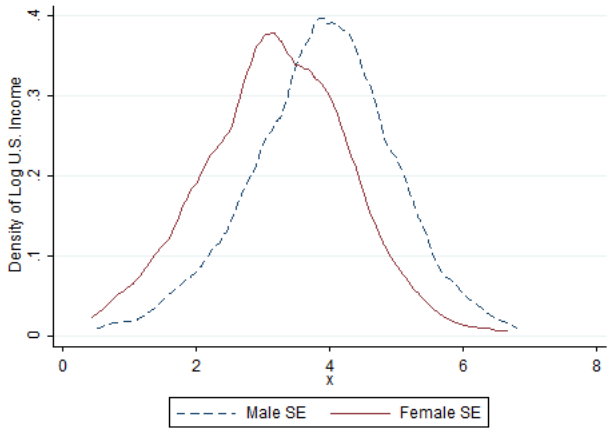


Figure 15: Distribution of Income by Gender in Tanzania

What drives the difference in earnings across men and women? To address this question, we first present results from regression models with profits as a dependent variable. We include education, a dummy for a savings account, a dummy indicating if the individual is married, an asset index, and the number of children. The asset index is variable we generate using principal component analysis on a series of indicators of wealth, including whether the individual owns a home, bike, watch, cell phone, landline, iron, television, gas or electric stove, refrigerator, car, the individual's house quality and number of rooms in the home, and whether the individual has electricity at home.[2] We also include a gender dummy as an independent variable and interact it with each of the other independent variables. The regression results can be found in the appendix. We then test if the coefficients on the interaction of gender with the other variables are all equal to zero. For Ghana we obtain an F-statistic of 5.4 with 7 and 2,272 degrees of freedom and a p-value of 0 so that we can reject the null. This implies that an approach that allows for different regression coefficients is warranted. For Tanzania we obtain an F-statistic of 80.4 with 6 and 3,957 degrees of freedom and a p-value of 0 so we can also reject the null for Tanzania quite handily. For Rwanda we obtain an F-statistic of 5.25 with 6 and 515 degrees of freedom and a p-value of 0 so we also reject the null for Rwanda.

This leads us to use the Blinder-Oaxaca decomposition. The goal of this decomposition is to look at differences in an outcome (here it is profits) across two groups and to separately identify the contribution that differences in endowments make to the overall difference separately from differences in the coefficients and differences due to the interaction. Put in other words, we use the Blinder-Oaxaca decomposition to assess what part of the gender differences we see in profits is due to observable differences in endowments (for example men just have more schooling and resources than women) and what part of the gender difference is due to different returns. Different returns to men and women with the same characteristics could be indicative of discrimination in the labor market, but could also be due to omitted variable bias if we do not include as regressors all determinants of income. Given that it is impossible with the given data set to guarantee that we have included all determinants

[2]For household quality, we include three measures: whether or not the household has good walls, floors, and roof. Good walls are walls made of stone or bricks. Good floors are floors constructed of concrete or tiles. A good roof is constructed from concrete, metal or tiles. We are not able to include all the assets in the asset index, as not all of the variables are available for all countries. For Ghana we are missing variables for radio ownership, bicycle ownership, watch ownership, car ownership, refrigerator ownership, and iron ownership. For Tanzania, we are missing a variable for household ownership.

of income in our equations, differences in returns to the same endowments for men and women should not be taken as evidence of discrimination. Instead, it shows that discrimination may be happening, or there may be an omitted variable that is driving the results. We will present the broad results for each country in table below. This table shows the raw differential between men and women for each country as well as the differential due to endowments and the differential due to differences in returns to endowment. The extensive tables with all results for the multiple specifications can be found in the appendix. However, the results can be seen in the bar graph for all three countries. The bar graph depicts the overall amount of the difference that is due to differences in endowments between men and women and differences in the returns to these endowments. This overall contribution to the difference in income between men and women is depicted with the triangle on each bar. The bar graph further depicts the contribution to the difference of the specific variables we use in the analysis.

In each of the three countries, between 32 and 36 percent of the male-female gender gap in wages can be explained by differences in mean characteristics. For Ghana and Rwanda, we observe whether an individual has access to a savings account, and mean gender differences in savings account ownership contribute the most to the portion of the gender wage gap that can be attributed to differences in explanatory variables. In Tanzania, we do not observe savings account ownership; for Tanzania, lower average ages and education level of female entrepreneurs are the biggest contributors to the explained portion of the gender gap.

There are several important things to note about the remaining unexplained percentage of the gender wage gap. The first is that in each country, returns to age contributes negatively to the gender wage gap: the age profile for self-employed females is higher than for self-employed males. While we cannot tell whether this difference is due to increased returns to experience for female entrepreneurs (perhaps women lack beneficial business networks upon entry into self-employment but build them up slowly) or a selection effect (older women whose businesses are not profitable tend to leave the labor force), it is still notable that the difference in profitability of businesses run by older versus younger workers is bigger for females than males. The remaining difference is almost entirely in the constant. So while unfortunately this result does not tell us a variable whose differential returns by gender can explain the gender wage gap, the lack of explanatory power of variables such as education or number of children is an important result itself. Women do not seem to be penalized for having children (in Tanzania, children actually contribute negatively to the gap), and they are able to translate their education into earnings at the same

rate (or even more in the case of Tanzania) as men . Instead, we posit that a large part of the gender wage gap appears due to variables more difficult to measure, such as access to business networks or specific choices when running a business (such as men taking greater risks) that result in greater average profits for men.

Table 3: Decomposition by Gender: Traits

	Ghana	Rwanda	Tanzania
Raw Differential (Men-Women)	0.193	0.550	0.671
-due to endowments	0.077	0.218	0.057
-due to coefficients	0.064	0.303	0.605
-due to interaction	0.052	0.029	0.008

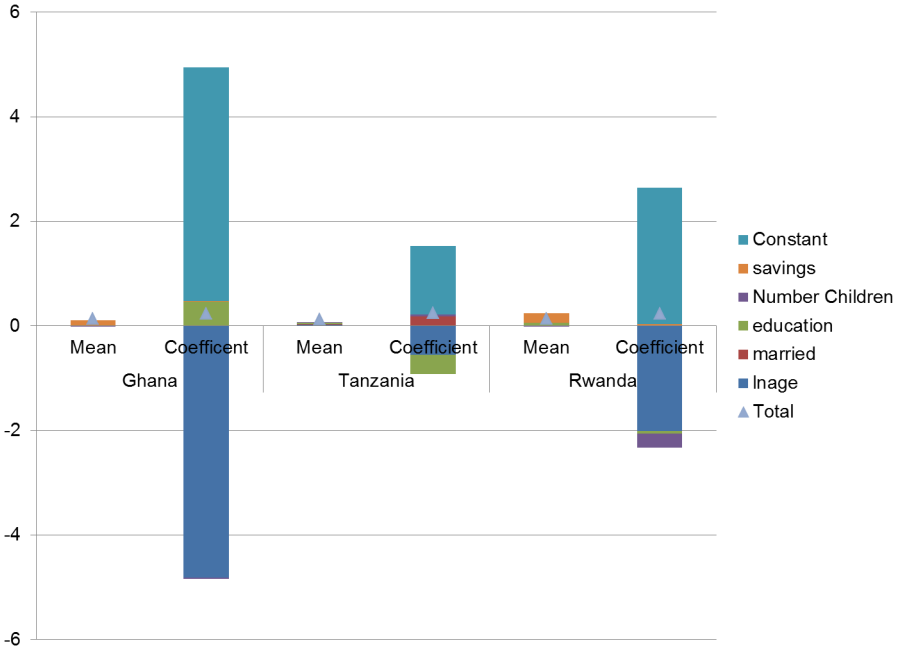


Figure 16: Oaxaca Decomposition of Male-Female Income Gap by Country with Traits

Table 4: Decomposition by Gender: Traits and Sectors

	Ghana	Rwanda	Tanzania
Raw Differential (Men-Women)	0.193	0.550	0.671
-due to endowments	0.022	0.218	0.119
-due to coefficients	0.135	0.303	0.591
-due to interaction	0.036	0.029	-0.039

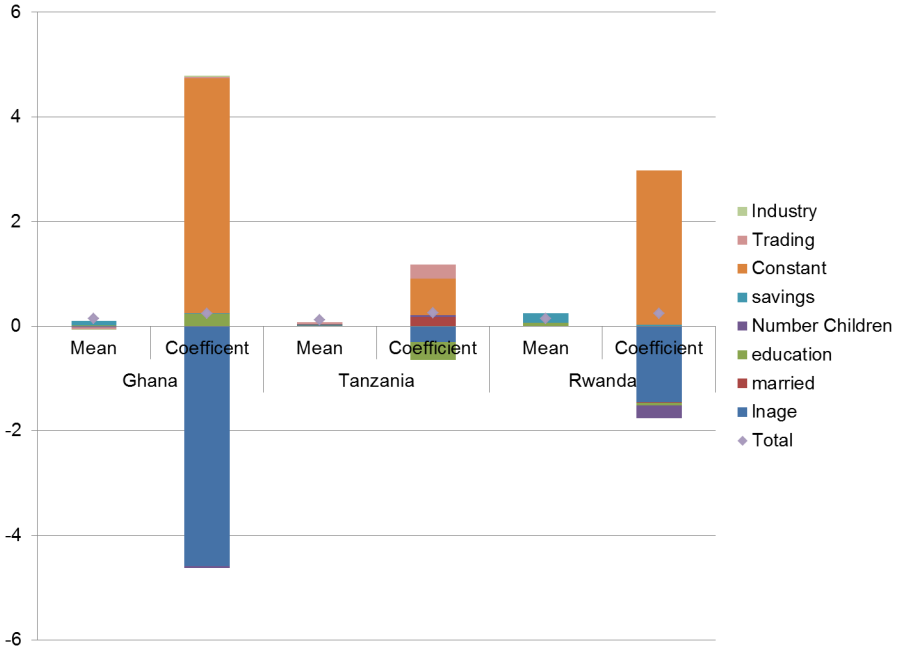


Figure 17: Oaxaca Decomposition of Male-Female Income Gap by Country with Traits and Sectors

Table 5: Decomposition by Gender: Traits, Sectors, and Asset Index

	Ghana	Rwanda	Tanzania
Raw Differential (Men-Women)	0.194	0.364	0.666
-due to endowments	0.046	0.004	0.038
-due to coefficients	0.131	0.121	0.664
-due to interaction	0.017	0.239	-0.036

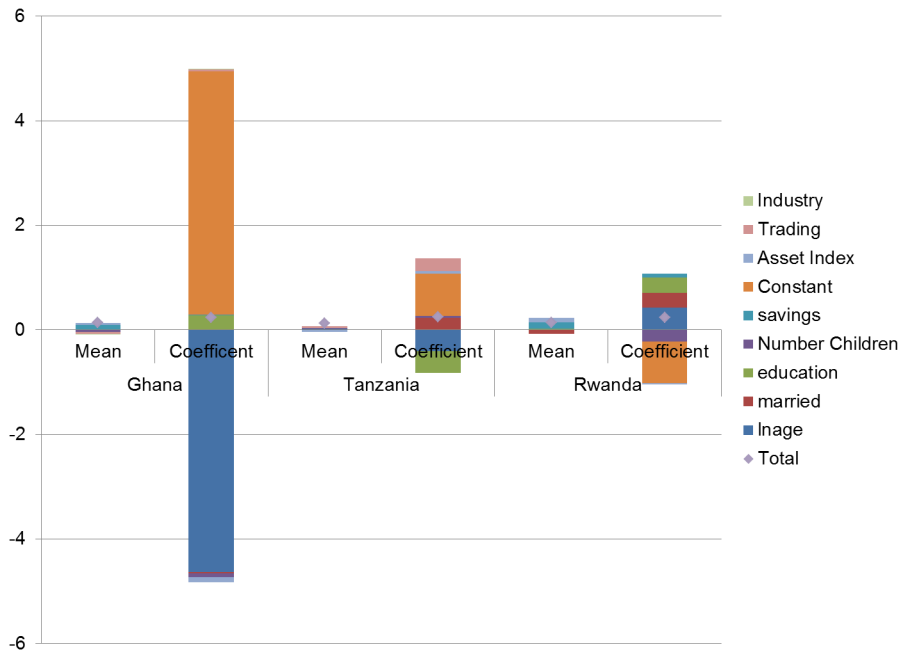


Figure 18: Oaxaca Decomposition of Male-Female Income Gap by Country with Traits, Sectors, and Asset Index

4 Policy Implications

This paper analyzes the role of individual traits on occupational choice (defined as the choice between self-entrepreneurship versus wage workers) and on profitability of smaller size businesses. The analysis based on Ghana, Rwanda, and Tanzania shows that a classification of entrepreneurs based on broad categories (such as the dichotomy of necessity versus opportunity or the size of business) provides a useful categorization. However the relevance of individual specific traits demonstrated so far in this paper shows that additionally considering the traits of individual entrepreneurs helps researchers and policy makers to identify groups that are homogenous in terms of earnings.

The analysis shows first that traditional traits such as education, experience, marriage (which can be seen as a proxy for time availability) do not fully explain occupation choice. Based on the observable traits, certain individuals would be better off in either wage work or self-employment. Unexplained reasons behind occupational choice might include risk aversion. For example, average self-employment earnings for men are larger but more volatile than earnings in wage work. Along the same line, we find that men have on average higher but more volatile unconditional earnings in self-employment.

Looking at business profitability, we find that across all countries under analysis, female owners enjoy a premium on their age compared to men. While we are not able to fully explore the reason behind this results, age might reflect the experience of entrepreneur in running a business. Women might enjoy a premium if their experience allows them to reduce the disparities they face in other fields. For example, men might find it easier to build a network and to obtain information compared to women. However, this disparity might be reduced via experience. In the case of Tanzania, other variables also affect male and female entrepreneurs differently: compared to men, women apper to enjoy a premium on education but to be penalized if they are married. The former might still reflect the role of education in allowing women to reduce gaps in other dimensions. The latter might reflect the fact that married women have less time to devote to their businesses due to household duties.

Differences in average endowments explain only part of the gap. However, among the observable variables, access to a saving account appears particularly important in the countries for which we have this information (Ghana and Rwanda) . Compared to returns to traits, policy makers can more readily affect the difference in the average savings endowment. We thus propose in the next session an exercise that builds on these findings.

4.1 Incorporating Randomized Interventions Results with the Typology of Entrepreneurs: The Effect of Making Savings Accounts Accessible

In Dupas and Robinson's paper "Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya" the authors find that providing access to an interest-free bank account to the self-employed increases productive investment[4]. Further, the authors argue that these investments result in increased consumption expenditures of the self-employed. The authors find heterogeneous results, with access to savings accounts increasing women's investments and expenditures while it does not affect investments and expenditures for men. Even more interesting, the bank accounts are interest free and charge substantial withdrawal fees, which means that they actually charge negative interest rates and yet the effect is still there. This suggests that women face strong savings barriers and high returns to capital.

We use the measured effects of providing savings accounts and demonstrate the impact on women entrepreneurs' income of expanding the intervention to the entire population in the three countries in our cross section: Ghana, Rwanda and Tanzania. We do this by first matching the random sample the authors ran the trial on to a sample of individuals in our data. Once we have a comparable sample, we then use their estimates of the effects of the intervention to predict the effect on the broader population. We present the results by graphing a histogram of incomes before and after the intervention, with those affected shown in red and those who are unaffected shown in blue..

In order to generate a sample in each country that roughly matches the sample on which the randomized trial was run, we use the means and standard deviations from the authors' sample characteristics to pull a comparable sample from each country in our data. We do not have data on all of the characteristics they provide summary statistics for in the paper, but we do match on age, marital status, number of children, education, and received loan in past year. Dupas and Robinson restrict the sample to women and men who are self-employed as market vendors or bicycle taxis. In their sample practically all men are bicycle taxi drivers and all women are market vendors. We thus restrict our sample to women traders. We deem this to be a close enough fit to their sample. Since they find little effect for men but relatively large effects for women, we do not include men in the sample.

The authors do not measure changes in incomes directly. Instead, they measure three outcomes: percent that use the savings accounts, changes in business investments, and changes in expenditures. We use changes in expenditures as a proxy for change in incomes. Changes in expenditures are

heterogeneous, and accrue primarily to women who use the account. In generating a treatment effect, we use the authors' intent to treat effect. This is the average effect across the treated sample. The authors find that expenditures range from a 12.43% increase on previous expenditures to a 21.02% increase on previous expenditures. The recorded past expenditures almost exactly matches reported income, so we use the change in expenditures as a proxy for change in income. Thus, to estimate the effect of this intervention in Ghana, Rwanda and Tanzania, we increase incomes of the applicable sample by 12.43%.

It is important to point out two caveats to this sort of analysis. First, we assume that the results of the randomized trial hold as is when expanded to the entire economy, so that there are no general equilibrium effects. General equilibrium effects could mitigate the changes demonstrated in the trial. For example, the authors present suggestive evidence that the women in the sample may be able to save at the cost of transfers to social networks outside of the household. If these transfers are going to malaria treatment for friends and neighbor's children, for example, the savings accounts could lower social welfare. However, if the transfers are going to social spending that is not welfare enhancing then the effects of the savings accounts would seem to be welfare enhancing. It seems very unlikely that in Dupas and Robinson's paper on savings constraints the general equilibrium effects would wipe out the results, but it is equally unlikely that they would remain exactly the same.

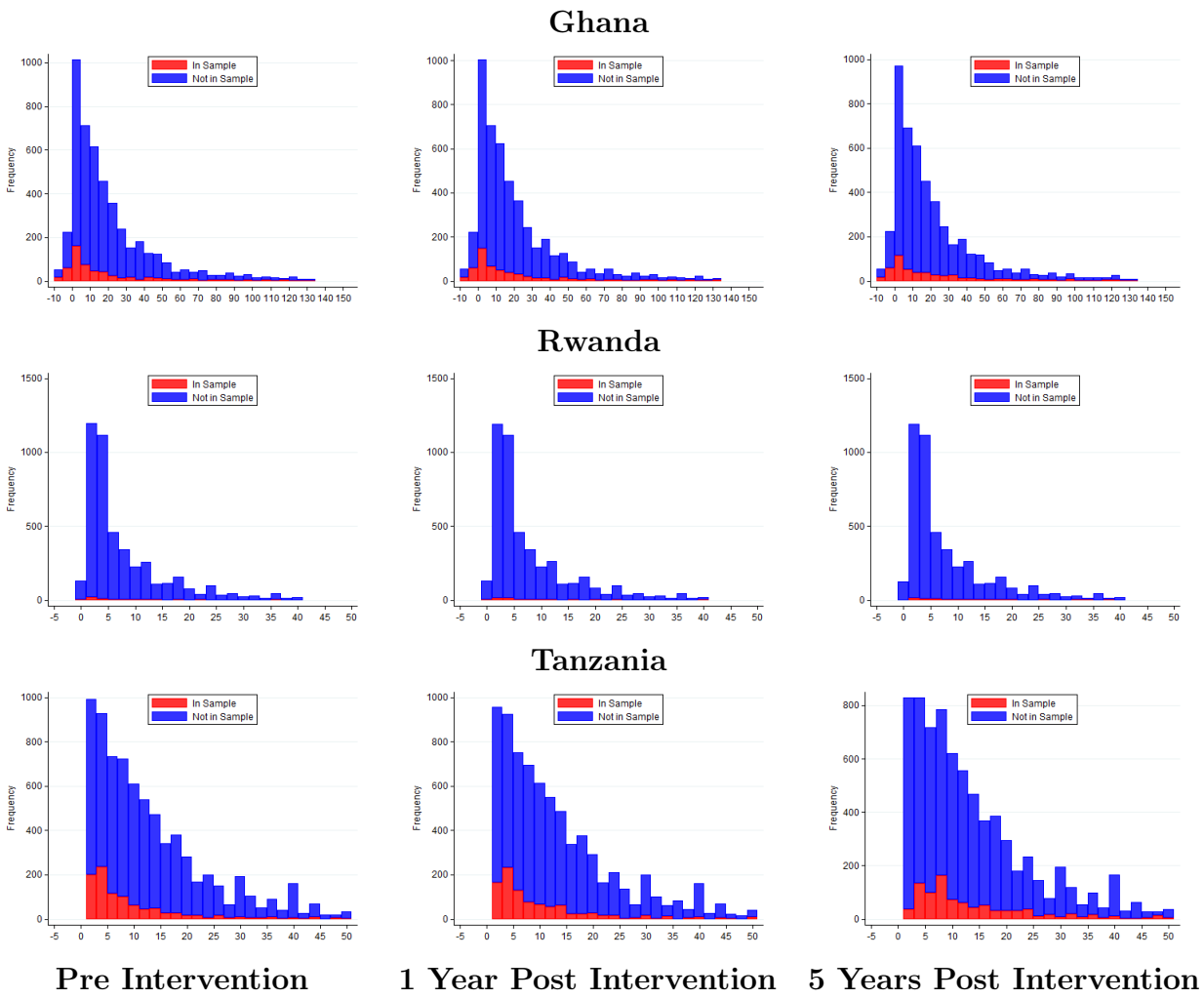
Second, there is the question of external validity. Cultural and other unobserved differences in Kenya, that do not exist in Ghana, Rwanda or Tanzania, might be driving the results. For example, the authors look carefully at the results and find evidence that the effects are likely due to women's inability to save at home due to transfers demanded by family and friends. If this constraint on savings does not exist in other countries, then we would not expect the same impact of increasing access to savings accounts.

Due to these caveats, it is best to read the graphs that show the economy wide effects on the distribution of incomes as an upper bound on the results. Below, we first present the summary statistics of our pulled, matched sample to compare with the summary statistics from the original sample the intervention tested. These summary statistics can be seen in 6. Next , we present graphs with the intent to treat estimates of effects. For each country we present three graphs. First, we present the pre-intervention histograms of incomes. The histograms make clear where along the income distribution those in the matching samples in Ghana, Rwanda and Tanzania are, by dividing the bars giving the total frequency in each income category into two colors. Red indicates individuals who would qualify for the intervention (women who are

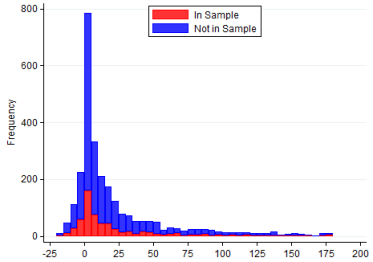
self-employed market vendors) while blue indicates individuals who would not qualify for the intervention. Pre-treatment (the column on the left), the earnings distributions of the women in the sample who match the treatment group appear similar to the earnings distribution overall, although the earnings of women who qualify are slightly more lumped near zero. The middle and right columns show the earnings distribution after 1 and 5 years respectively, assuming the "treated" women had similar growth rates to Dupas and Robinson. While the experimental effects are small relative to the variance in the distribution of earnings, so that the earnings distribution doesn't shift dramatically, we do see the "treated" individuals disproportionally moving away from the bunch at zero earnings and into the right tail of the distribution relative to the non-treated. This latter group could perhaps benefit from the intervention. While the intervention found no effect for men in the original sample, other self-employed women could theoretically benefit from the intervention. However, we exclude individuals who do not match Dupas and Robinson's original treatment group from our "treatment" group (and instead leave them in the "control" group) as we lack measures of the results of such a treatment on other groups, and wish to extrapolate as little as possible.

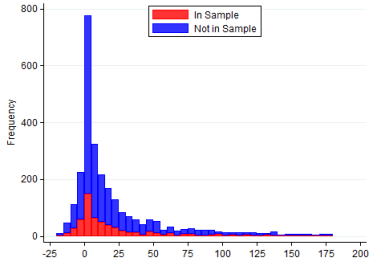
Table 6: Dupas and Robinson Sample Summary Statistics

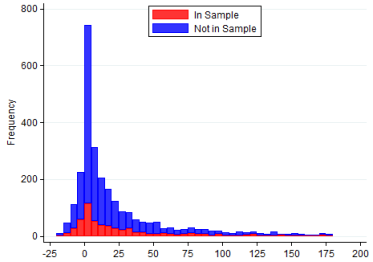
	(1) mean	sd	min	max	count
Ghana					
Age in Years	36.88392	10.49405	16	65	715
Married Dummy	.648951	.4776323	0	1	715
Household Size	4.699301	2.606915	1	21	715
Education (Categorical, 1-8)	4.876923	1.565189	1	8	715
Literate Dummy	.9608392	.1941133	0	1	715
Contribution to Community Savings	.	.	.	.	0
Weekly Income in USD	19.41978	28.37274	-6.634107	132.9263	555
Rwanda					
Age in Years	34.04348	11.25715	17	58	92
Married Dummy	.4021739	.4930235	0	1	92
Household Size	5.576087	2.750559	1	14	92
Education (Categorical, 1-8)	2.177778	.8423853	1	5	90
Literate Dummy	.6304348	.4853319	0	1	92
Contribution to Community Savings	37.24716	69.04187	0	448.0287	92
Weekly Income in USD	5.808705	6.569186	.7168459	35.8423	52
Tanzania					
Age in Years	34.00634	10.61293	15	65	1262
Married Dummy	.6259905	.4840579	0	1	1262
Household Size	4.889857	2.524376	1	18	1262
Education (Categorical, 1-8)	3.082166	.7443051	1	8	1071
Literate Dummy	.8541997	.3530458	0	1	1262
Contribution to Community Savings	.	.	.	.	0
Weekly Income in USD	9.06034	8.803477	1.328609	49.20775	962
Total					
Age in Years	35.00242	10.68407	15	65	2069
Married Dummy	.6239729	.484504	0	1	2069
Household Size	4.854519	2.56852	1	21	2069
Education (Categorical, 1-8)	3.722814	1.462758	1	8	1876
Literate Dummy	.881102	.3237467	0	1	2069
Contribution to Community Savings	37.24716	69.04187	0	448.0287	92
Weekly Income in USD	12.617	18.94743	-6.634107	132.9263	1569
Observations	2069				


Ghana
Rwanda
Tanzania
In Sample
Not in Sample
Frequency
Pre Intervention
1 Year Post Intervention
5 Years Post Intervention

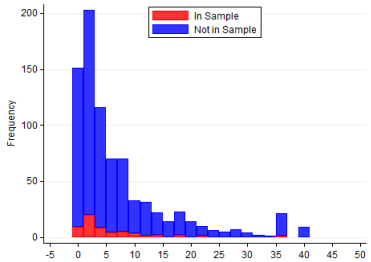
Ghana Self-Employed

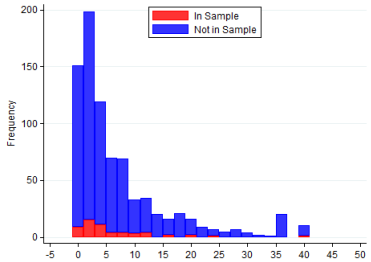


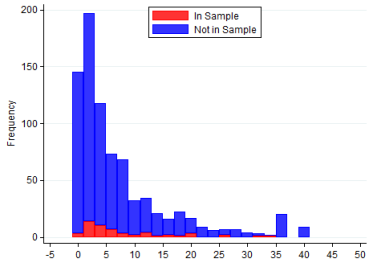




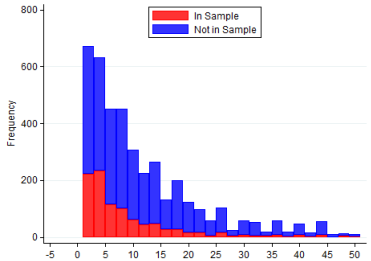
Rwanda Self-Employed

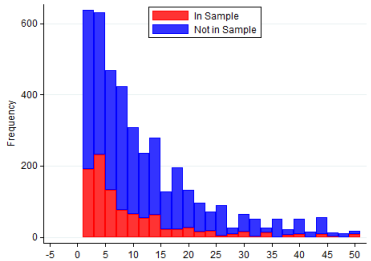


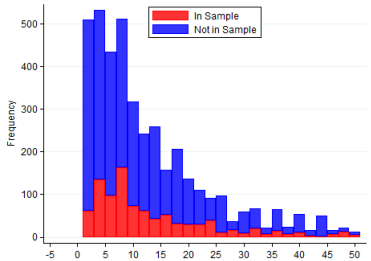




Tanzania Self Employed





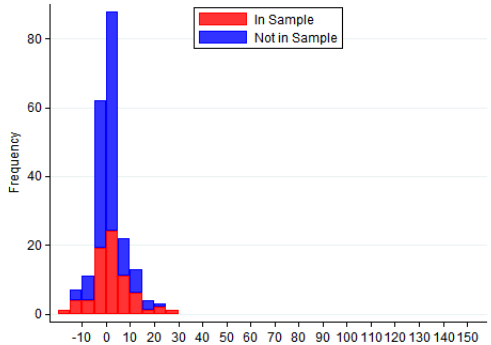


Pre Intervention **1 Year Post Intervention** **5 Years Post Intervention**

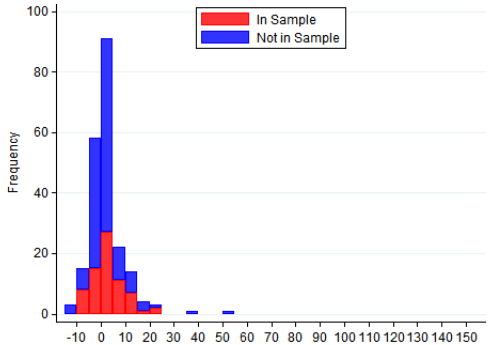
We can also look at the change in the distribution of the growth of income. To do this, we utilize panel data from Ghana. We examine the distribution in the change in income from 2005-2006 and from 2006-2007. We assume the intervention occurs in 2005, and compound the results to get the change in 2006-2007. The graphs that show the change in the distribution of the growth of income are presented below. Overall, we see that the biggest effects of the intervention were concentrated around zero growth rates. That is, the intervention shifted the earnings of some entrepreneurs who would be making slightly negative profits toward slightly positive profits. While these effects may not prompt the type of explosive growth that generates substantial job creation, these small effects may be particularly important for entrepreneurs near the subsistence level.

Ghana Panel

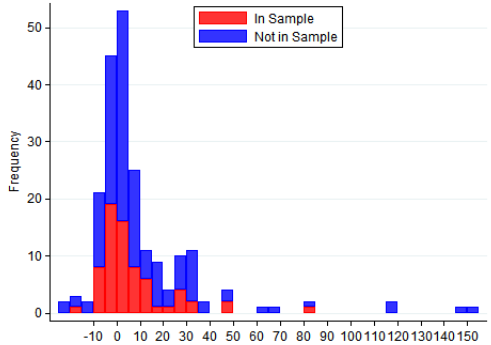
2005-2006 Income Growth Distribution Pre Intervention



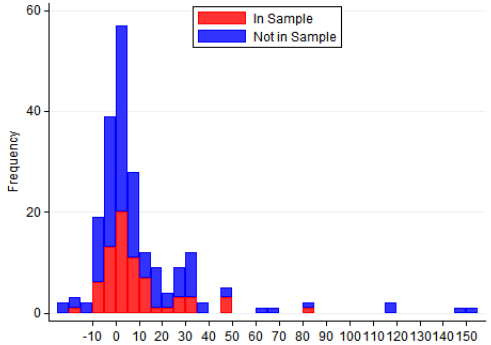
2005-2006 Income Growth Distribution Post Intervention



2006-2007 Income Growth Distribution Pre Intervention



2006-2007 Income Growth Distribution Post Intervention



References

[1] P. Chuhan-Pole, V. Korman, M. Angwafo, and M. Buitano. Africa's pulse an analysis of issues shaping africa's economic future. *World Bank Report*, 2011.

[2] Rachel Croson and Uri Gneezy. Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474, 2009.

[3] Suresh de Mel, David J. McKenzie, and Christopher Woodruff. Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics*, 88(1):19–31, January 2009.

[4] Pascaline Dupas and Jonathan Robinson. Savings constraints and microenterprise development: Evidence from a field experiment in kenya. Working Paper 14693, National Bureau of Economic Research, January 2009.

[5] Marcel Fafchamps, David McKenzie, Simon R. Quinn, and Christopher Woodruff. When is capital enough to get female microenterprises growing? evidence from a randomized experiment in ghana. Working Paper 17207, National Bureau of Economic Research, July 2011.

[6] Louise Fox and Thomas Pave Sohnesen. Household enterprises in sub-saharan africa: Why they matter for growth, jobs, and livelihoods. *SSRN eLibrary*, August 2012.

[7] J. Kweka and L. Fox. The household enterprise sector in tanzania: why it matters and who cares. *World Bank Policy Research Working Paper Series, Vol*, 2011.

[8] A. Schoar. The divide between subsistence and transformational entrepreneurship. In *Innovation Policy and the Economy, Volume 10*, pages 57–81. University of Chicago Press, 2010.

Appendix

	Ghana	Rwanda	Tanzania
VARIABLES	Log U.S. Monthly Income	Log U.S. Monthly Income	Log U.S. Monthly Income
Male Dummy	1.194***	1.757***	1.016***
	(0.393)	(0.566)	(0.186)
Age	0.0180***	0.0181*	0.0119***
	(0.00359)	(0.00975)	(0.00262)
Male*Age	-0.0373***	-0.0308**	-0.00715**
	(0.00773)	(0.0134)	(0.00344)
Savings	0.725***	0.554	
	(0.0962)	(0.336)	
Male*Savings	0.0750	0.212	
	(0.181)	(0.422)	
Education	-0.00131	0.196**	0.175***
	(0.0257)	(0.0787)	(0.0349)
Male*Education	0.0942*	-0.0715	-0.139***
	(0.0543)	(0.0988)	(0.0435)
Married	0.221***	-0.466**	-0.218***
	(0.0826)	(0.205)	(0.0504)
Male*Married	-0.0148	0.550*	0.412***
	(0.181)	(0.281)	(0.0718)
Asset Index	0.337***	0.589***	0.313***
	(0.0397)	(0.114)	(0.0258)
Male*Asset Index	-0.239***	-0.0318	0.0586*
	(0.0822)	(0.140)	(0.0345)
Number Children	0.0707***	0.0816*	-0.0406***
	(0.0228)	(0.0449)	(0.0141)
Male* Number Children	-0.0480	-0.0558	0.0167
	(0.0511)	(0.0599)	(0.0184)
Constant	2.877***	1.334***	2.265***
	(0.190)	(0.467)	(0.153)
Observations	2,286	527	3,969
R-squared	0.093	0.259	0.220

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1